PEPTIDE MODELS OF PARALLEL β-SHEET AND AMYLOID (MIS)FOLDING

By

Vanessa M. Kung

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN – MADISON

2016

Date of final oral examination: November 30, 2015

The dissertation is approved by the following members of the Final Oral Committee:

Samuel H. Gellman, Professor, Chemistry
M. Thomas Record, Professor, Chemistry
Robert J. McMahon, Professor, Chemistry
Ronald T. Raines, Professor, Chemistry
Tehshik P. Yoon, Professor, Chemistry
Deane F. Mosher, Professor, Biomolecular Chemistry

Acknowledgements

Thank you to Professor Samuel H. Gellman and my labmates, for stimulating my development as an independent researcher. Thank you to my undergraduate coworkers, Xiangyang Liu, Stefan P. Kostelyna, N. Shirley Nwangwa, and C. Mei Yon, for their assistance with syntheses. I am very grateful to my committee, Professors M. Thomas Record, Robert J. McMahon, Ronald T. Raines, Hans J. Reich, Charles P. Casey, Deane F. Mosher, Anna Huttenlocher, and Tehshik P. Yoon, for their advice and support. It was a pleasure to learn from numerous experts during grad school, including those listed at the bottom of this page. I am particularly indebted to Tom, Gabriel, Charles, Darrell, Gail, Marco, Milo, Charlie, Martha, Monika, and Randy, for educating and invaluable discussions. I treasured interactions with Professors R. Claude Woods, Robert W. Woody, and scientists I met through Gordon Research Conferences, who provided encouragements and discussions when I most needed these things.

I was fortunate to work alongside many people in the chem department, biochem department, and medical school. Kat Myhre and Chelsea Hanewall were unfailingly helpful, as were other administrative staff, including Paul Cook, Dr. Rita Hannah, Dr. Matt Sanders, Karen Stephens, and Phill Bloedow. Teaching was a highlight of grad school, and I have very happy memories of evening labs with my students and Drs. Nick Hill and Matt Bowman. It has been a privilege to work with students and staff across campus while I've served on the student council (GSFLC). Matt Martin (machine shop) greatly lightened my load in maintaining lab equipment. It brightened late nights in lab to have the company of True Thao Yang (custodial staff). Others who helped make things run smoothly included the UW-Madison Interlibrary Loan team, Jim Zernicke (stockroom), Dr. Bob Shanks (instrument technologist), Dr. Desiree Bates (computer center), Alan Silver (computer center), Tracy Drier (glass shop), Chad Skemp (purchasing office), Bruce Goldade (copy center), and Betty Harwood (business services).

I have been blessed beyond measure in my family and friends, who are constant sources of inspiration, strength, and joy. Above all, I thank my parents and sister, whose prayers and love sustained me during the hardest stretches.

NMR: Gabriel Cornilescu, Charles D. Schwieters (NIH), Charles G. Fry, Marco Tonelli, W. Milo Westler, Monika Ivancic, Mark E. Anderson, Vincent B. Chen, Hongyang Yao, Claudia C. Cornilescu, Woonghee Lee, Hans J. Reich

Biophysics: Darrell R. McCaslin, M. Thomas Record, Robert W. Woody (Colorado State University)

Mass Spectrometry: Martha M. Vestling

Bioinformatics: Gail J. Bartlett, Derek N. Woolfson (University of Bristol)

Microscopy: Randall J. Massey, Benjamin K. August

X-Ray Crystallography: Ilia A. Guzei

Other Collaborators: Nicole L. Burke, Andrew F. DeBlase, Timothy S. Zwier (Purdue University); Wenjiang Ma, Deane F. Mosher.

Funding: NIH R01 GM061238 (S.H.G.), NIH T32 GM008692 (Medical Scientist Training Program), UW-Madison Distinguished Graduate Fellowship

Table of Contents

Chapter 1. Introduction

1.1. Statement of Research Purpose

Understanding protein structures is important to understanding how proteins function, and how these functions might be enhanced or inhibited. A better understanding of protein structures and folding may also lead to the development of medical tools for amyloid diseases, which arise from protein misfolding. Interestingly, the majority of amyloid proteins appear to be composed of parallel β-sheets. Of the common protein secondary structures, parallel β-sheets are the least characterized, in large part because they are difficult to isolate. Minimal folding units of parallel β-sheets – i.e., short peptides that can fold into these structures in the absence of other structural contexts – can be used to garner important information about sheets, but are challenging design and synthesize. Here we report the first experimental model of a three-stranded parallel β-sheet (**Chapter 2**), and use this model system to study the effect of the number of strands on parallel β-sheet stability. The relationship between size and stability for the parallel β-sheet is important to characterize from a basic science perspective, and is also of interest because amyloid structures appear to grow via a propagation of parallel β-sheet structure. In **Chapter 3** we demonstrate how bioinformatics studies can inform the *de novo* design of minimal models of protein secondary and tertiary structures. We design a water-soluble, minimal model of amyloid-like structure, and use it to explore hydrophobic interactions involved in amyloid formation. Model systems like ours may contribute to understandings of protein structure, stability, and interactions.

1.2. Protein Structure and Nomenclature

Of the four major types of biological macromolecules (proteins, carbohydrates, nucleic acids, and lipids), proteins are the most abundant, and have a vast array of biological functions.[1] Proteins, and shorter chains called (poly)peptides, are polymers of amino acid residues. The 20 standard amino acids are depicted in **Figure 1.1**.

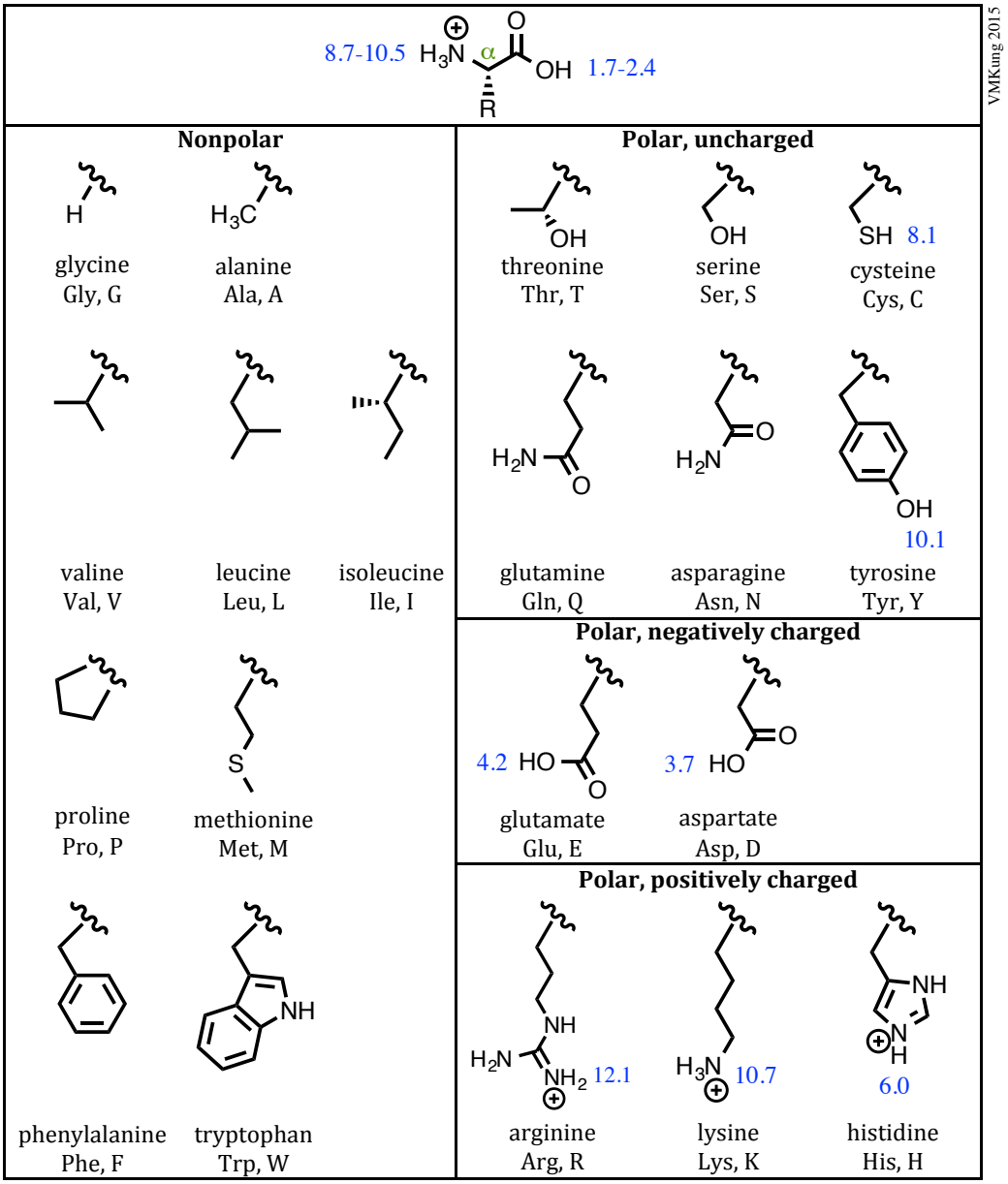

Figure 1.1. The 20 standard amino acids, listed with their 3- and 1-letter abbreviations, side-chain (R-group) structures, and pKa values (blue)[2]. Note that all are L-stereoisomers at the α-position (green).

Protein structure is organized hierarchically into four levels: primary, secondary, tertiary, and quaternary.[1] Primary structure is the sequence of amino acid residues that are connected by covalent bonds (e.g., peptide bonds, disulfide bonds). Secondary structure is defined by the pattern of hydrogen bonds between amide hydrogen atoms and carbonyl oxygen atoms of the protein/peptide backbone (e.g., **Figure 1.2B** dashed lines), or a pattern of backbone dihedral angles in a Ramachandran basin (**Figure 1.3**), discussed in more detail below. The most common protein secondary structures are the α-helix and the β-sheet, the latter of which can have antiparallel or parallel orientations (**Figure 1.2A**). Tertiary structure describes all aspects of the three-dimensional fold of a single protein/peptide. Quaternary structure arises from inter-molecular interactions between two or more folded protein/peptide chains. Instead of adopting a defined (secondary) structure, a chain or segment of it can be unstructured as a random coil.

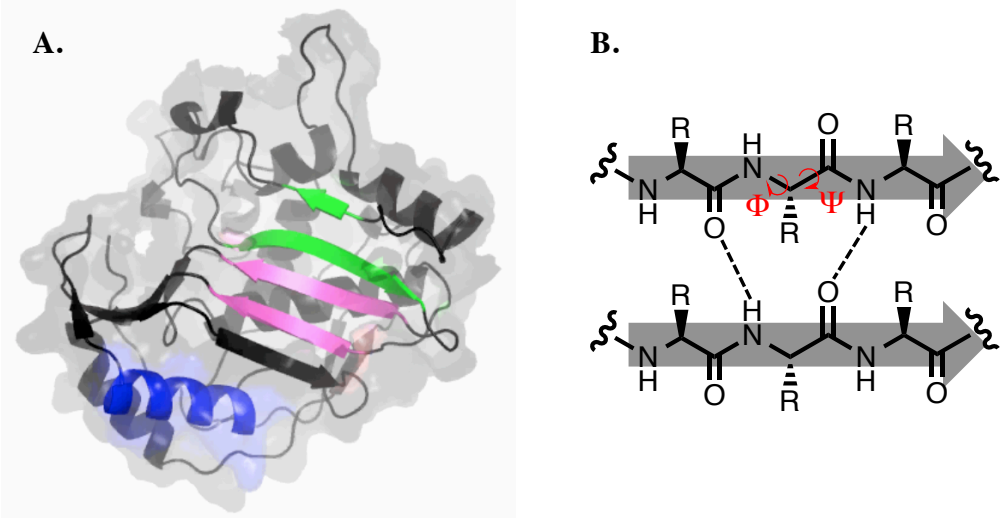

Figure 1.2. Ribbon diagrams of protein secondary structure. (**A**) An arbitrarily chosen protein (1SYN.pdb), containing α-helix (blue), antiparallel β-sheet (green), and parallel β-sheet (pink) secondary structures. (**B**) Ribbon diagram of a parallel β-sheet, demonstrating how cartoon arrows point from N- to C-termini of chains, have width in the plane of the backbone atoms, such that there are 'edges' where hydrogen bonds (dashed lines) form, and wide 'faces' from which side-chains (R-groups) protrude. Conformation-defining backbone dihedral angles (Φ,Ψ) for one residue are shown in red.

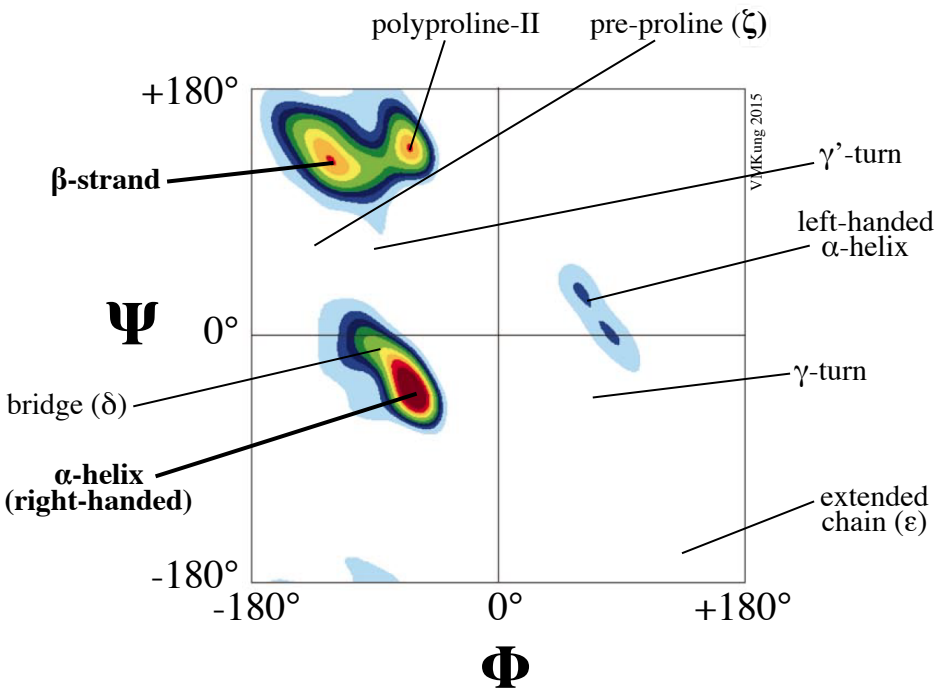

Figure 1.3. Ramachandran plot (Φ,Ψ) for all 20 of the standard amino acids. Ramachandran basins are labeled, with the highest frequency protein secondary structures in bold font. Figure adapted from Hovmoller, *et al.*,[3] with permission from the International Union of Crystallography.

The amino acid residues found in natural proteins and peptides are almost always L stereoisomers, as depicted in **Figure 1.1**.[1] The backbone conformations of amino acid residues are determined by the dihedral angles Φ and Ψ, defined in **Figure 1.2B**. A Ramachandran plot graphs the frequency of occurrence of these backbone dihedral angles in proteins/peptides (**Figure 1.3**).[3,4] This plot reveals basins of frequently-occurring backbone conformations, such as those for residues in (right-handed) α-helices or in β-strands. There is symmetry in the Ramachandran plot. For instance, the left-handed α-helix basin mirrors the (right-handed) α-helix basin; however, the left-handed α-helix basin is infrequently occupied, because naturally occurring L-amino acid residues favor right-handed α-helical conformations (**Figure 1.3**).

1.3. Significance of Studying Protein Folding and Misfolding

The term "protein" comes from the Greek word for "first place" (*proteios*), and indeed, proteins are vital to almost every task of cellular life, with functions ranging from reaction catalysis (e.g., enzymes of glycolysis), to structural support (e.g., in connective tissues), to cell signaling (e.g., receptors on a neuron), and much more.[5] Most available drugs have protein targets – i.e., they treat diseases by binding to specific protein structures, thereby inhibiting undesired or enhancing desired protein functions.[6,7] The function of a protein depends on its structure or fold. Thus, to understand and potentially harness protein functions requires an understanding of protein structures.

Proteins with a wide variety of native folds can misfold, aggregate, and deposit in tissues as assemblies called amyloid fibrils. Diseases involving amyloid formation (e.g., Alzheimer's disease, type 2 diabetes mellitus, **Table 1.1**) debilitate tens of millions of people worldwide.[8] Yet there currently are no medical tools available for early diagnosis or treatment of amyloid formation.[9,10] For instance, definitive diagnosis of amyloidosis requires invasive biopsy procedures, followed by microscopic examination of tissue samples (e.g., **Figure 1.4**). Agents designed to stabilize protein/peptide structures off-pathway to amyloid formation (e.g., tafamidis)[11] have generated great interest as potential therapeutics, though so far none have been approved by the U.S. Food and Drug Administration (FDA) for use in humans. A major obstacle in the design of medical tools that can effectively target amyloid formation has been limitations in our understanding of amyloid structure and mechanisms of amyloid (mis)folding. The relatively little information that we do have about amyloid structures suggests that those important to disease pathogeneses contain in-register parallel β-sheet structure, and cross-β structure

from interdigitation of side-chains extending orthogonal to the β-sheets (discussed in more detail in **Chapter 3**).[12-17] A better understanding of amyloid structures may enable the design of agents that detect, stabilize, and/or destabilize these structures, which could lead to the development of much-needed medical tools. In addition, the extraordinary mechanical properties, including elasticity and tensile strength, of amyloid fibrils (similar to spider silk) have prompted interest in the design of materials based on these structures.[18]

Table 1.1. Amyloid formation can be found in a wide variety of human diseases. This list is not comprehensive. List compiled by V.M. Kung; other lists available in reviews[19].

	Disease	**Protein or Peptide**
Neurologic	Alzheimer's disease	amyloid β
	Parkinson's disease	α-synuclein
	spongiform encephalopathies	prion
	Huntington's disease	huntingtin
Cardiovascular	restrictive/obliterative cardiomyopathy	
Respiratory	interstitial lung disease	
Endocrine	thyroid medullary carcinoma	calcitonin
	type 2 diabetes mellitus	amylin/IAPP
Musculoskeletal	carpal tunnel syndrome	
Renal	type II renal tubular acidosis	
	2° glomerular disease	
Hematologic	AL amyloidosis (e.g., multiple myeloma)	Ig light chain
	AA amyloidosis	amyloid A
Gastrointestinal	inflammatory bowel disease	

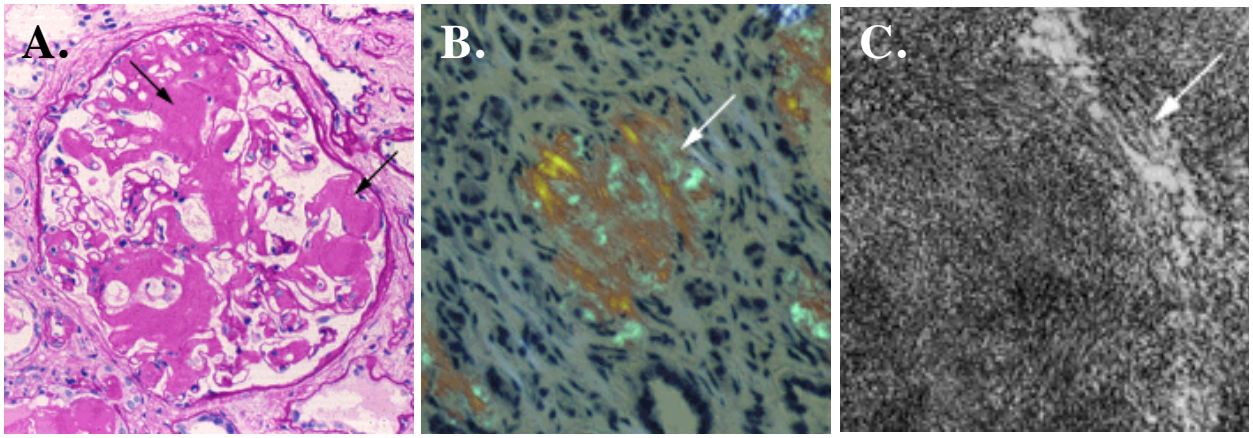

Figure 1.4. Example micrographs of kidney glomeruli, showing renal amyloidosis. Diagnosis of renal amyloidosis requires the demonstration of (**B**) apple green birefringence with Congo red staining and cross-polarized light, or (**C**) amyloid fibrils by electron microscopy. (**A**) Light micrograph with H&E staining. Micrographs by H.G. Rennke,[20] reproduced with permission from UpToDate, Inc.

1.4. Amyloid Mechanisms of Formation, Structures, and Toxicities

It is thought that regardless of their sequences or native folds (α-helix, β-sheet, and/or intrinsically disordered), most proteins are capable of misfolding into amyloid (β-sheet) structures.[21,22] There are also non-pathological proteins that appear to have amyloid fibril structures as their functional, native folds.[19,23] Yet amyloid structures are not typically found in healthy human tissues. Normally, native folds of proteins bury the polypeptide backbones and side-chains that when exposed would be prone to aggregation. Exquisite cellular regulation (e.g., molecular chaperones) and the cooperative nature of protein folding prevent the prolonged exposure and aggregation of incompletely folded, misfolded, and partially degraded proteins. When the regulation of protein folding or protein levels is impaired (e.g., by genetic mutations, inflammation), aggregation and amyloid formation can occur. Some normal cellular proteins and processes may also increase the risk of amyloid formation; for instance, there are proteins that are intrinsically

disordered in their native forms, and protein unfolding can be required for translocation across membranes.

A simplified diagram of current theories on the mechanisms of amyloid formation can be found in **Chapter 3** (**Figure 3.1**). The amyloid state may be more thermodynamically stable than the native state of a protein/peptide (especially one with < 150 residues), even at physiologic concentrations and conditions, such that only high kinetic barriers prevent conversion to amyloid structures.[24,25] Amyloid fibril formation likely occurs via a nucleated growth mechanism.[19] Typically there is a rate-limiting lag phase (nucleation) with formation of oligomers, followed by a rapid growth phase with formation of fibrils. Seeding preformed fibrils (with sequence identity >30%) into a sample can shorten or eliminate the lag phase.[26-28] Nucleation and elongation events both depend on the formation of backbone hydrogen-bonds (β-sheets) and contacts among hydrophobic side-chains.[29,30] Amino acid substitution studies have shown that the propensity of a sequence to form amyloid structures depends on its hydrophobicity.[29-31] Likewise, algorithms that depend heavily on hydrophobicity parameters can predict the propensities of protein regions to form amyloids.[32,33] Oligomer formation appears to involve coalescence around hydrophobic groups, followed by rearrangement into more stable oligomers, which have some exposed hydrophobic side-chains and cross-β structure (including hydrogen-bonded β-sheets).[27,34-37] As oligomers grow into fibers, the surface-to-volume ratios and the amounts of exposed hydrophobic regions decrease.[34,35] The ultrastructural characteristics (e.g., width, twist) of amyloid fibrils depend on several forces, including the hydrophobic effect, intrinsic conformational preferences for twisting (arising from amino acid chirality), and the elastic strain from twisting of the fibrils.[34,38]

High-resolution structures of natural amyloids have been elusive, in large part because amyloid fibrils are not crystalline and are too large or insoluble to study by solution state NMR spectroscopy. Studies of amyloid structures (and mechanisms of formation) have been further complicated by the fact that single protein/peptide species can adopt multiple different amyloid conformations (generally depending in the sample preparation conditions), a phenomenon termed polymorphism. However, to the best of our current knowledge, the majority of amyloid polymorphs formed at relatively physiologic conditions contain parallel β-sheet structure and hydrophobic packing of side-chains between neighboring sheets ("steric zipper"), as depicted in **Figure 1.5**.[39]

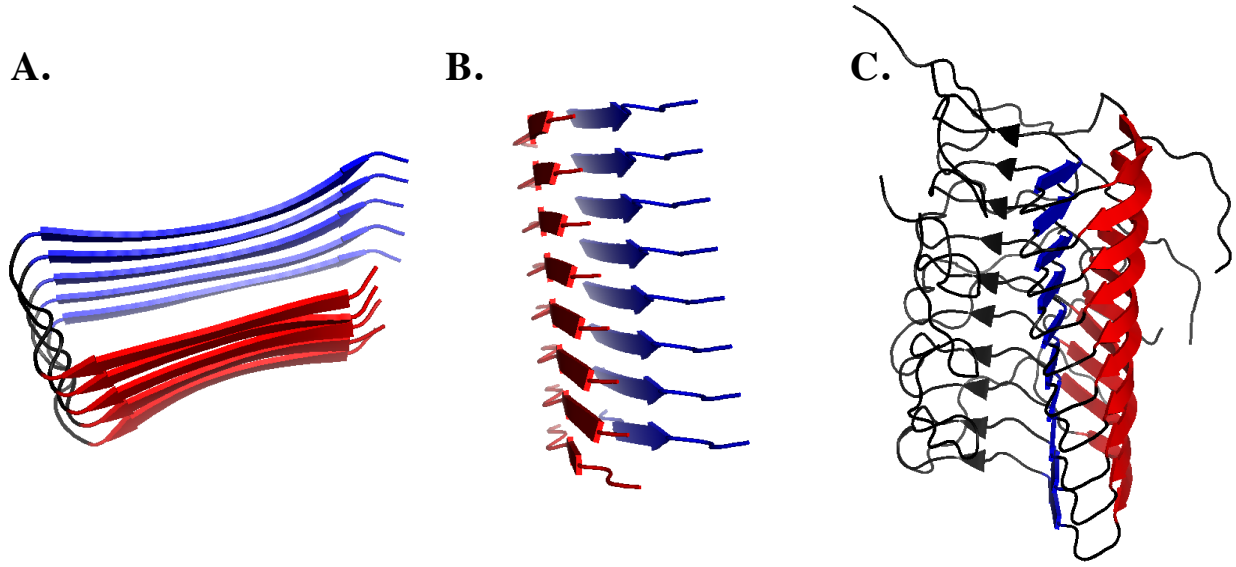

Figure 1.5. Example high-resolution structures of amyloid fibrils derived from fragments of (**A**) amyloid β,[40] and (**B**) transthyretin,[41] which may be involved in human diseases. A structure of a prion, (**C**) Het-s,[42] is also shown. Note the presence of parallel β-sheets, and that red β-sheets interact with blue ones via contacts among hydrophobic side-chains.

Combinations of the following methods have been employed to determine amyloid fibril structures at various resolutions. Low-resolution imaging techniques, such as electron microscopy (EM) and atomic force microscopy (AFM), can provide information such as the diameter and mass-per-unit-length of fibrils.[43,44] Circular dichroism (CD) and

Fourier transform infrared (FTIR)[45] spectroscopy provide information about the β-sheet content of structures. X-ray diffraction of amyloid fibrils (not crystals) shows a characteristic cross-β pattern, from which distances between β-strands can be extrapolated.[46] Single crystal X-ray diffraction of microcrystals has yielded the highest resolution data about amyloid structures, for instance elucidating side-chain packing arrangements within the cross-β motif; however, microcrystals are grown from amyloid peptide fragments (4-7 residues long) with undetermined medical relevance.[14,47] Fragments of amyloid peptides can have structures that differ significantly from those of their full-length counterparts. For example, the 40-42 residue long amyloid β peptides found in Alzheimer's disease have parallel β-sheet structures, but when their hydrophobic C-termini are removed, they instead have antiparallel β-sheet structures.[48] Recent advances in a variety of biophysical techniques, most notably solid-state NMR (ssNMR) spectroscopy, have permitted the determination of higher resolution structures that may be more biologically relevant. Unlike X-ray crystallography or solution state NMR spectroscopy, ssNMR spectroscopy can provide atomic-resolution data about the structures of non-crystalline solids and semi-solids, including intact amyloid fibrils. ssNMR has also been employed to distinguish the orientation (parallel or antiparallel) of β-sheets within amyloid fibrils, via measurements of ^{13}C-^{13}C dipole-dipole couplings in samples with ^{13}C-labeling at strategic sites.[49] Parallel β-sheet arrangements in amyloid fibrils have also been studied by electron paramagnetic resonance (EPR) spectroscopy,[50,51] and cryo-electron microscopy (cryo-EM).[52] Other clever techniques that have facilitated the determination of amyloid structures include hydrogen-deuterium exchange,[53-55] proline scanning mutagenesis,[56] site-directed labeling,[50,57,58] and limited proteolysis[59,60]

experiments.

Even less structural information exists for amyloid oligomers and protofibrils than for amyloid fibrils. Spherical oligomers and protofibrils have been observed by TEM and AFM (e.g., **Figure 1.6**). These precursors to fibril formation are thought to share structural features with the fibrils that they form. They can bind dyes (e.g., Congo red, thioflavin T) also bound by fibrils.[61] CD spectroscopy experiments have shown that oligomers are mostly composed of β-sheets, which are increasingly stabilized as the oligomers grow in size.[61,62] Some oligomeric precursors to fibril formation have also been shown by NMR to contain rudimentary cross-β structure with parallel β-sheets, supporting the theory that molecular-level polymorphisms in oligomers propagate into fibrils.[13,63,64] It would not be surprising for oligomers on-pathway to the formation of a fibril to contain the same orientation of β-sheets as that of the fibril, since conversion between antiparallel and parallel β-sheets within aggregated structures would require the disruption of hydrogen-bonds and reorientation of polypeptide chains in a sterically crowded environment.[12] One particularly intriguing study found that polyclonal antibodies raised against amyloid β oligomers could also bind amyloid oligomers from other sources (e.g., α-synuclein, IAPP, PrP), suggesting that different amyloid oligomers may have common structural elements.[65]

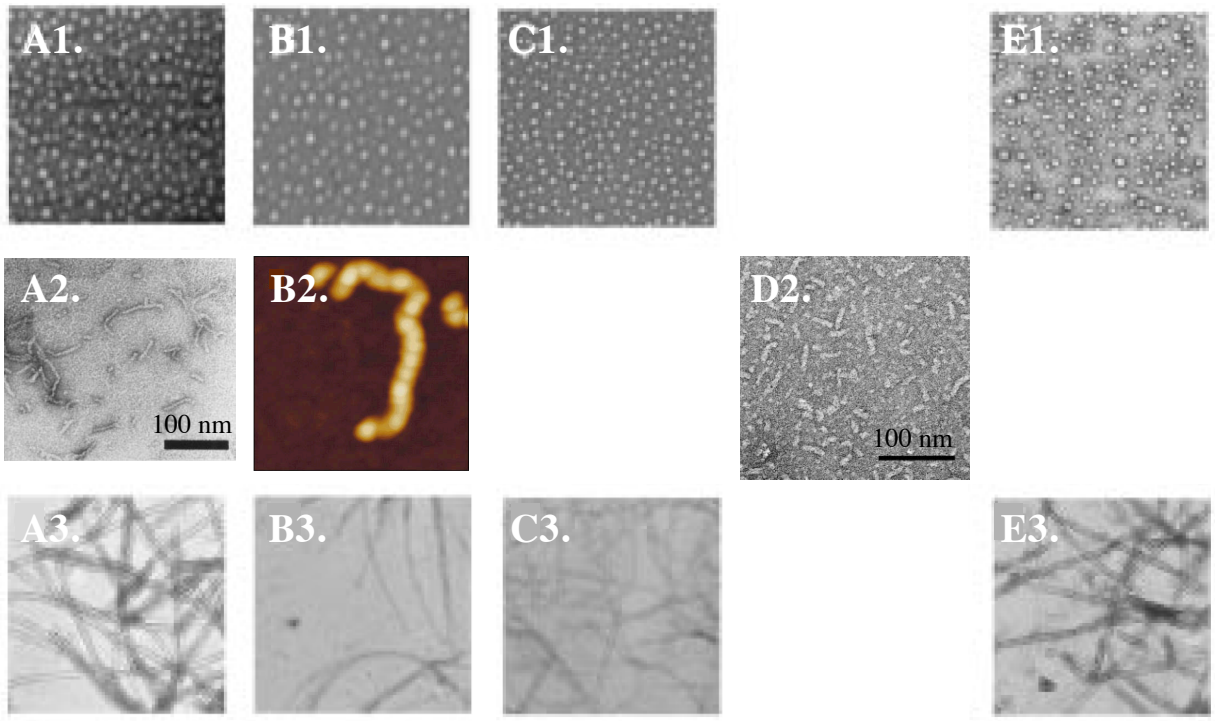

Figure 1.6. Example TEM and AFM images of spherical oligomers (row 1), protofibrils (row 2), and fibrils (row 3) of (**A**) amyloid β,[61,66] (**B**) α-synuclein,[66,67] (**C**) amylin,[66] (**D**) transthyretin,[68] and (**E**) prion PrP.[66] Figures reproduced with permissions from the American Society for Biochemistry and Molecular Biology (ASBMB) and the American Chemical Society (ACS).

Soluble amyloid oligomers are generally thought to be the cytotoxic species in amyloidoses (e.g., Alzheimer's disease,[69-72] familial amyloid polyneuropathy[73-75]), based on both *in vitro* and *in vivo* studies. Oligomers are also thought to be the most infectious form of the mammalian prion protein PrP.[76] Inflammatory responses, which both contribute to and result from amyloid formation, can also cause pathologies.[77] Large deposits of amyloid fibrils will physically damage cell or tissue architectures, and may also contribute to disease pathogeneses by more indirect routes (e.g., serve as reservoirs of toxic oligomers, sequester proteins that may be needed for normal biological functions). There is literature asserting that amyloid fibrils are the toxic species in amyloid diseases (including scattered reports that amyloid fibrils can interact with various receptors to trigger cellular stress

responses[78,79]). However, any role that fibrils have in amyloid disease pathogeneses is very likely dwarfed by the role of oligomers. Indeed, clinical observations of amyloidoses (e.g., Alzheimer's disease,[80-83] Parkinson's disease,[84] AL amyloidosis[85,86]) suggest that many symptoms correlate with levels of oligomers, not fibrils.

1.5. Methods of Peptide Design, Synthesis, and Structural Characterization

Methods employed for our studies are introduced and described in more detail in later chapters and appendices. The design of peptides was largely based on bioinformatics studies of proteins containing particular β-sheet structures, including new work described in **Chapter 3**. Standard Fmoc solid phase peptide synthesis (SPPS) methods were employed, and the development of synthetic routes for particularly difficult syntheses is described in **Appendix 1**. In **Chapters 2** and **3**, aqueous solution structures of peptide monomers were determined by nuclear magnetic resonance (NMR) spectroscopy, with assignment of ^{1}H signals *via* a combination of two-dimensional COSY, TOCSY, and ROESY or NOESY data. Peptides were not heavy isotope-labeled, due to the difficulty of the peptide syntheses, and natural abundance ^{13}C NMR data were not obtained, due to limited access to high-sensitivity probes at the time of these experiments. Monomeric and oligomeric peptide structures were further characterized by circular dichroism (CD) spectroscopy and transmission electron microscopy (TEM). Thermodynamic parameters of peptide folding and/or oligomerization were derived from thermal denaturation experiments, monitored by CD and ^{1}H NMR. Peptide oligomerization and/or aggregation, or a lack thereof, was also studied by sedimentation equilibrium analytical ultracentrifugation (AUC), NMR diffusion experiments, and variable concentration NMR and CD experiments.

1.6. Previous Reports by Gellman Laboratory of Peptide Models of β-Sheet Structure

The authors of Ref. 87 employed the peptides depicted in **Figure 1.7** to study the effect of strand length on antiparallel β-sheet stability.[87] They reported that secondary chemical shift analyses of select α-^{1}H and α-^{13}C NMR signals showed that a two-stranded antiparallel β-sheet is stabilized when strands are lengthened from 5 to 7 residues, but that further strand lengthening to 9 residues does not lead to further stabilization in most cases. Peptides were described to be monomeric at all studied conditions. This conclusion was reportedly reached via aggregation studies restricted to sedimentation equilibrium AUC and variable concentration CD experiments.[88] Only two rotor speeds, 44k and 56k rpm, were reportedly used.

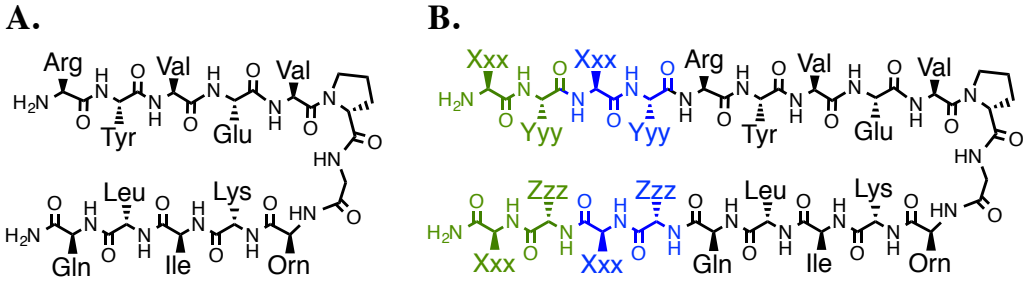

Figure 1.7. Peptides employed to study the effects of strand length on antiparallel β-sheet stability in Ref. 87. The [A] 5 residue long "core" (black) was extended to 7 (blue) or [B] 9 (green) residues, where residues (Xxx, Yyy, Zzz) = (Thr, Thr, Thr), (Ser, Thr, Thr), (Thr, Ala, Ala), (Ser, Thr, Tyr).

The authors of Ref. 89 employed the peptides depicted in **Figure 1.8** to study the effect of strand number on antiparallel β-sheet stability.[89] In Ref. 90, peptide **1.8A** was compared to peptides containing ^{L}Pro-Gly linkers (rather than ^{D}Pro-Gly linkers). The authors of Ref. 89 also reported that their data supported a cooperative, two-state model for folding of two-stranded β-sheets, and a "four-state" model for folding of three-stranded β-sheets.

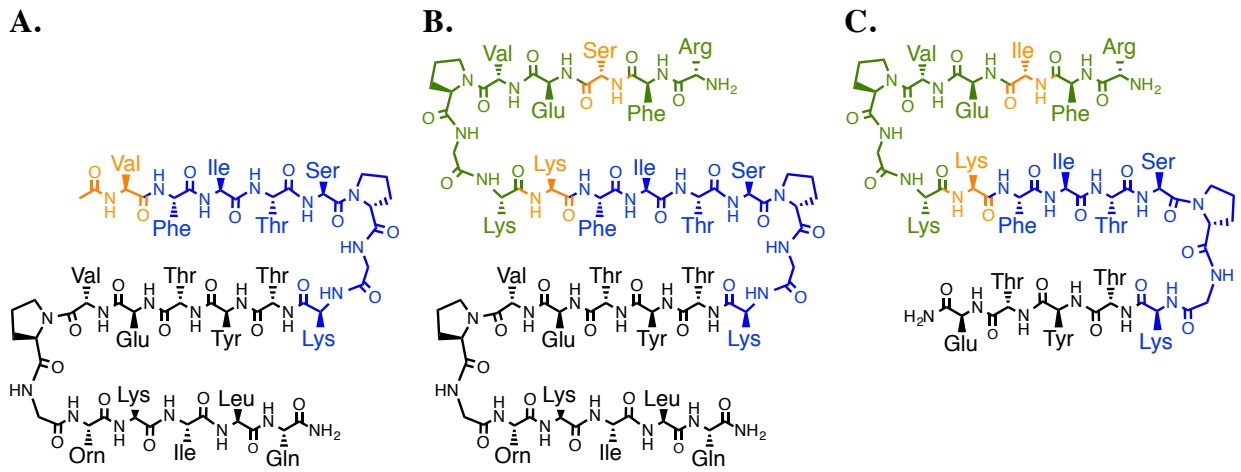

Figure 1.8. Peptides employed to study the effects of strand number on antiparallel β-sheet stability Ref. 89. Differences in peptide sequences are highlighted in orange. The folding of peptide [A] was also studied in Ref. 90.

The authors of Ref. 91 employed the peptides depicted in **Figure 1.9** to study the effect of strand length on parallel β-sheet stability.[91] The authors reported that β-sheet populations increased, or the parallel β-sheet was stabilized, as strand lengths were increased from 6 to 10 residues.[91]

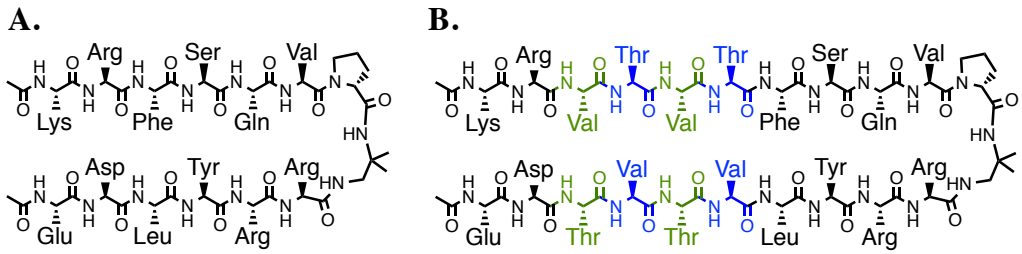

Figure 1.9. Peptides employed to study the effects of strand length on parallel β-sheet stability in Ref. 91. The [A] 6 residue long "core" (black) was extended to 7 (blue), 8 (green), 9 (blue), or [B]10 (green) residues.

Other peptides that have been reported to form monomeric two-stranded parallel β-sheets in aqueous solution include those depicted in **Figure 1.10** and **Figure 1.11**. The peptide depicted in **Figure 1.10** was reported to adopt a well-folded parallel β-sheet structure.

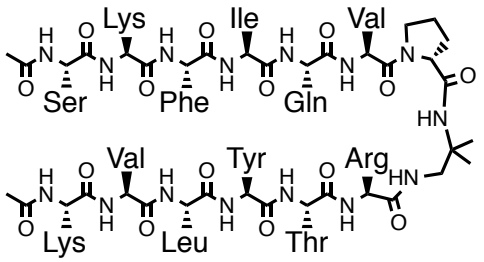

Figure 1.10. A peptide studied in Ref. 92.

The peptide depicted in **Figure 1.11** was reported to adopt a monomeric, parallel β-sheet structure, based on α-^{1}H secondary chemical shifts.[93]

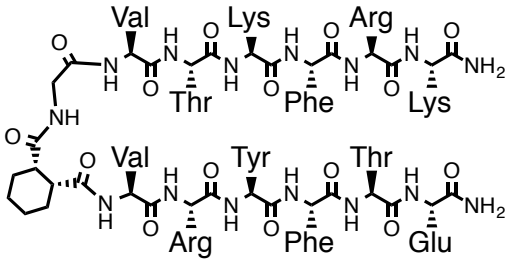

Figure 1.11. A peptide studied in Ref. 93.

1.7. Utility of Peptide Models of Protein Structure and Folding

Peptide models of protein structures, and other forms of *de novo* protein design, are fundamental to the study of protein folding, structure, and function, and may also lead to the development of biotechnology.[94-96] Peptide models of α-helices were developed in the 1980's,[97,98] and peptide models of antiparallel β-sheets were developed in the 1990's.[99-103] These model systems enabled the identification of components required for the folding and stability of these secondary structures. Peptide models of parallel β-sheets have been slower to emerge, largely because non-natural linkers must be employed to connect β-strands in a parallel manner, and promote their folding in the absence of a stabilizing, tertiary framework. Examples of non-natural, parallel β-sheet-promoting linkers – including linkers based on norbornene,[104] phenoxathiin,[105] dibenzofuran,[106-108] and oligoureas[109,110] – are shown in **Figure 1.12**. These linkers have been challenging to design

and synthesize, and so far only a few have been reported to be compatible with the development of water-soluble peptide model systems.[92,93,107,111] The linkers shown in **Figures 1.12G** and **1.12E**, developed by John D. "Nick" Fisk (Samuel H. Gellman Laboratory) and Felix Freire (Samuel H. Gellman Laboratory), respectively, have previously been employed in reports of water-soluble models of two-stranded parallel β-sheets.[92,93,111] We employ these linkers to develop novel, water-soluble peptides containing two to four parallel β-strands.

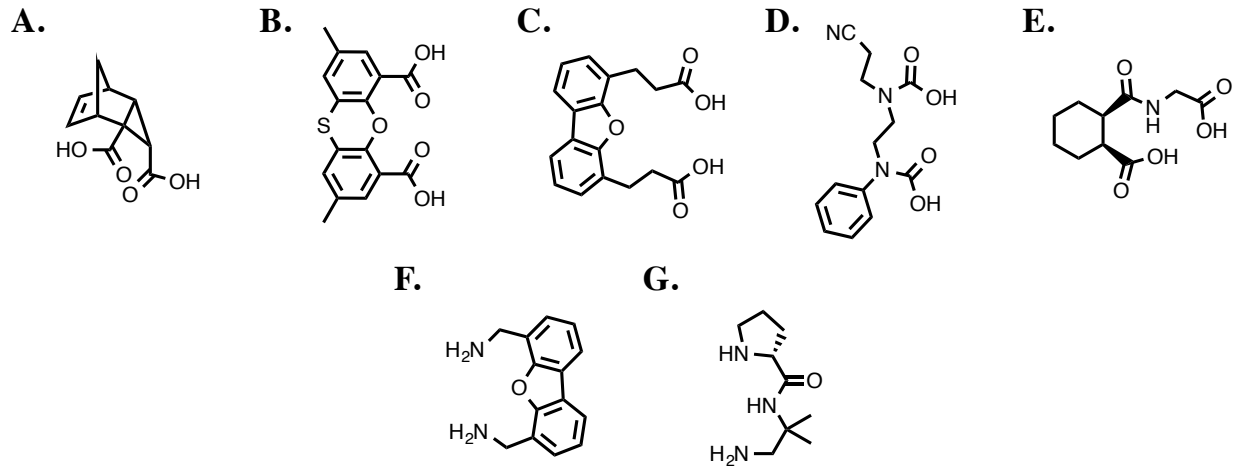

Figure 1.12. Examples of non-natural diacid (A-E) and diamine (F-G) linkers that can connect and promote hydrogen-bonding between two parallel β-strands.

The design of β-sheet structures, especially parallel ones, and the design of tertiary structures are major challenges in *de novo* protein design.[112-114] A shortage of minimal models of parallel β-sheets (i.e., peptides that fold autonomously into parallel β-sheets, in the absence of protein tertiary structural context) has both stemmed from and contributed to a poorer understanding of parallel β-sheet structure compared to other secondary structures. In **Chapter 2** we report the first water-soluble, three-stranded parallel β-sheet model. We employ this model to explore the impact of the number of β-strands ("strand number") on parallel β-sheet stability, and demonstrate how model systems like ours may

be used to determine spectroscopic characteristics of parallel β-sheets. The relationship between strand number and parallel β-sheet stability is of interest from a basic science perspective, and is also intriguing because amyloid fibrils appear to grow in the direction of increasing strand number.[39,115,116] In **Chapter 3** we report the design and characterization of a water-soluble, minimal folding unit of amyloid-like tertiary structure, containing both parallel β-sheet structure and interdigitation of hydrophobic side-chains. Model systems such as ours may have many potential applications, including in studies of the forces driving amyloid oligomerization, as agents to seed amyloid growth in *in vitro* studies, as amyloid substitutes in screens for diagnostic and therapeutic tool design, or even as scaffolds of diagnostic or therapeutic agents.

1.8. References

(1) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; 4th ed.; W.H. Freeman: New York, 2005.

(2) *CRC Handbook of Chemistry and Physics*; 96 ed.; CRC Press/Taylor and Francis: Boca Raton, FL, 2015.

(3) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr D* **2002**, *58*, 768.

(4) Hollingsworth, S. A.; Karplus, P. A. *Biomol Concepts* **2010**, *1*, 271.

(5) Campbell, N. A.; Reece, J. B. *Biology*; 7 ed.; Pearson Education, Inc.: San Francisco, CA, 2005.

(6) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. *Nat Rev Drug Discov* **2006**, *5*, 993.

(7) Imming, P.; Sinning, C.; Meyer, A. *Nat Rev Drug Discov* **2006**, *5*, 821.

(8) Prince, M.; Bryce, R.; Albanese, E.; Wimo, A.; Ribeiro, W.; Ferri, C. P. *Alzheimers Dement* **2013**, *9*, 63.

(9) Merlini, G.; Bellotti, V. *New Engl J Med* **2003**, *349*, 583.

(10) Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.; Merlini, G.; Saraiva, M. J.; Westermark, P. *Amyloid* **2014**, *21*, 221.

(11) Johnson, S. M.; Connelly, S.; Fearns, C.; Powers, E. T.; Kelly, J. W. *J Mol Biol* **2012**, *421*, 185.

(12) Tycko, R. *Neuron* **2015**, *86*, 632.

(13) Chimon, S.; Shaibat, M. A.; Jones, C. R.; Calero, D. C.; Aizezi, B.; Ishii, Y. *Nat Struct Mol Biol* **2007**, *14*, 1157.

(14) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773.

(15) Fandrich, M. *J Mol Biol* **2012**, *421*, 427.

(16) Kajava, A. V.; Baxa, U.; Steven, A. C. *FASEB J* **2010**, *24*, 1311.

(17) Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. *J Mol Biol* **2010**, *395*, 717.

(18) Knowles, T. P.; Buehler, M. J. *Nat Nanotechnol* **2011**, *6*, 469.

(19) Chiti, F.; Dobson, C. M. *Annu Rev Biochem* **2006**, *75*, 333.

(20) Gorevic, P. D. In *UpToDate.com* 2012; Vol. 2012.

(21) Dobson, C. M. *Philos T Roy Soc B* **2001**, *356*, 133.

(22) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. *Proc Natl Acad Sci USA* **2010**, *107*, 3487.

(23) Greenwald, J.; Riek, R. *Structure* **2010**, *18*, 1244.

(24) Gazit, E. *Angew Chem Int Ed* **2002**, *41*, 257.

(25) Baldwin, A. J.; Knowles, T. P.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. *J Am Chem Soc* **2011**, *133*, 14160.

(26) Naiki, H.; Hashimoto, N.; Suzuki, S.; Kimura, H.; Nakakuki, K.; Gejyo, F. *Amyloid* **1997**, *4*, 223.

(27) Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L. *Science* **2000**, *289*, 1317.

(28) Wright, C. F.; Teichmann, S. A.; Clarke, J.; Dobson, C. M. *Nature* **2005**, *438*, 878.

(29) Chiti, F.; Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Proc Natl Acad Sci USA* **2002**, *99 Suppl 4*, 16419.

(30) Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Nat Struct Biol* **2002**, *9*, 137.

(31) Wurth, C.; Guimard, N. K.; Hecht, M. H. *J Mol Biol* **2002**, *319*, 1279.

(32) Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat Biotechnol* **2004**, *22*, 1302.

(33) Pawar, A. P.; Dubay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J Mol Biol* **2005**, *350*, 379.

(34) Auer, S.; Meersman, F.; Dobson, C. M.; Vendruscolo, M. *PLoS Comput Biol* **2008**, *4*, e1000222.

(35) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G. *PLoS Comput Biol* **2007**, *3*, 1727.

(36) Cremades, N.; Cohen, S. I.; Deas, E.; Abramov, A. Y.; Chen, A. Y.; Orte, A.; Sandal, M.; Clarke, R. W.; Dunne, P.; Aprile, F. A.; Bertoncini, C. W.; Wood, N. W.; Knowles, T. P.; Dobson, C. M.; Klenerman, D. *Cell* **2012**, *149*, 1048.

(37) Campioni, S.; Mannini, B.; Zampagni, M.; Pensalfini, A.; Parrini, C.; Evangelisti, E.; Relini, A.; Stefani, M.; Dobson, C. M.; Cecchi, C.; Chiti, F. *Nat Chem Biol* **2010**, *6*, 140.

(38) Knowles, T. P.; De Simone, A.; Fitzpatrick, A. W.; Baldwin, A.; Meehan, S.; Rajah, L.; Vendruscolo, M.; Welland, M. E.; Dobson, C. M.; Terentjev, E. M. *Phys Rev Lett* **2012**, *109*, 158101.

(39) Toyama, B. H.; Weissman, J. S. *Annu Rev Biochem* **2011**, *80*, 557.

(40) Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. *Proc Natl Acad Sci USA* **2005**, *102*, 17342.

(41) Fitzpatrick, A. W.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Muller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V.; Griffin, R. G.; Dobson, C. M. *Proc Natl Acad Sci USA* **2013**, *110*, 5468.

(42) Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. *Science* **2008**, *319*, 1523.

(43) Serpell, L. C.; Sunde, M.; Benson, M. D.; Tennent, G. A.; Pepys, M. B.; Fraser, P. E. *J Mol Biol* **2000**, *300*, 1033.

(44) Harper, J. D.; Lieber, C. M.; Lansbury, P. T., Jr. *Chem Biol* **1997**, *4*, 951.

(45) Zandomeneghi, G.; Krebs, M. R.; McCammon, M. G.; Fandrich, M. *Protein Sci* **2004**, *13*, 3314.

(46) Eisenberg, D.; Jucker, M. *Cell* **2012**, *148*, 1188.

(47) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. *Proc Natl Acad Sci USA* **2005**, *102*, 315.

(48) Gordon, D. J.; Balbach, J. J.; Tycko, R.; Meredith, S. C. *Biophys J* **2004**, *86*, 428.

(49) Tycko, R. *Annu Rev Phys Chem* **2011**, *62*, 279.

(50) Torok, M.; Milton, S.; Kayed, R.; Wu, P.; McIntire, T.; Glabe, C. G.; Langen, R. *J Biol Chem* **2002**, *277*, 40810.

(51) Der-Sarkissian, A.; Jao, C. C.; Chen, J.; Langen, R. *J Biol Chem* **2003**, *278*, 37530.

(52) Jimenez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V.; Dobson, C. M.; Saibil, H. R. *Proc Natl Acad Sci USA* **2002**, *99*, 9196.

(53) Kheterpal, I.; Zhou, S.; Cook, K. D.; Wetzel, R. *Proc Natl Acad Sci USA* **2000**, *97*, 13597.

(54) Hoshino, M.; Katou, H.; Hagihara, Y.; Hasegawa, K.; Naiki, H.; Goto, Y. *Nat Struct Biol* **2002**, *9*, 332.

(55) Yamaguchi, K.; Katou, H.; Hoshino, M.; Hasegawa, K.; Naiki, H.; Goto, Y. *J Mol Biol* **2004**, *338*, 559.

(56) Williams, A. D.; Portelius, E.; Kheterpal, I.; Guo, J. T.; Cook, K. D.; Xu, Y.; Wetzel, R. *J Mol Biol* **2004**, *335*, 833.

(57) Krishnan, R.; Lindquist, S. L. *Nature* **2005**, *435*, 765.

(58) Chen, M.; Margittai, M.; Chen, J.; Langen, R. *J Biol Chem* **2007**, *282*, 24970.

(59) Kheterpal, I.; Williams, A.; Murphy, C.; Bledsoe, B.; Wetzel, R. *Biochemistry* **2001**, *40*, 11757.

(60) Monti, M.; Principe, S.; Giorgetti, S.; Mangione, P.; Merlini, G.; Clark, A.; Bellotti, V.; Amoresano, A.; Pucci, P. *Protein Sci* **2002**, *11*, 2362.

(61) Walsh, D. M.; Hartley, D. M.; Kusumoto, Y.; Fezoui, Y.; Condron, M. M.; Lomakin, A.; Benedek, G. B.; Selkoe, D. J.; Teplow, D. B. *J Biol Chem* **1999**, *274*, 25945.

(62) Modler, A. J.; Gast, K.; Lutsch, G.; Damaschun, G. *J Mol Biol* **2003**, *325*, 135.

(63) Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. *Nat Struct Mol Biol* **2010**, *17*, 561.

(64) Petkova, A. T.; Leapman, R. D.; Guo, Z.; Yau, W. M.; Mattson, M. P.; Tycko, R. *Science* **2005**, *307*, 262.

(65) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. *Science* **2003**, *300*, 486.

(66) Kayed, R.; Sokolov, Y.; Edmonds, B.; McIntire, T. M.; Milton, S. C.; Hall, J. E.; Glabe, C. G. *J Biol Chem* **2004**, *279*, 46363.

(67) Conway, K. A.; Harper, J. D.; Lansbury, P. T., Jr. *Biochemistry* **2000**, *39*, 2552.

(68) Quintas, A.; Vaz, D. C.; Cardoso, I.; Saraiva, M. J.; Brito, R. M. *J Biol Chem* **2001**, *276*, 27207.

(69) Hardy, J.; Selkoe, D. J. *Science* **2002**, *297*, 353.

(70) Lambert, M. P.; Barlow, A. K.; Chromy, B. A.; Edwards, C.; Freed, R.; Liosatos, M.; Morgan, T. E.; Rozovsky, I.; Trommer, B.; Viola, K. L.; Wals, P.; Zhang, C.; Finch, C. E.; Krafft, G. A.; Klein, W. L. *Proc Natl Acad Sci USA* **1998**, *95*, 6448.

(71) Hartley, D. M.; Walsh, D. M.; Ye, C. P.; Diehl, T.; Vasquez, S.; Vassilev, P. M.; Teplow, D. B.; Selkoe, D. J. *J Neurosci* **1999**, *19*, 8876.

(72) Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, M. J.; Selkoe, D. J. *Nature* **2002**, *416*, 535.

(73) Sousa, M. M.; Cardoso, I.; Fernandes, R.; Guimaraes, A.; Saraiva, M. J. *Am J Pathol* **2001**, *159*, 1993.

(74) Andersson, K.; Olofsson, A.; Nielsen, E. H.; Svehag, S. E.; Lundgren, E. *Biochem Biophys Res Commun* **2002**, *294*, 309.

(75) Reixach, N.; Deechongkit, S.; Jiang, X.; Kelly, J. W.; Buxbaum, J. N. *Proc Natl Acad Sci USA* **2004**, *101*, 2817.

(76) Silveira, J. R.; Raymond, G. J.; Hughson, A. G.; Race, R. E.; Sim, V. L.; Hayes, S. F.; Caughey, B. *Nature* **2005**, *437*, 257.

(77) Rogers, J.; Webster, S.; Lue, L. F.; Brachova, L.; Civin, W. H.; Emmerling, M.; Shivers, B.; Walker, D.; McGeer, P. *Neurobiol Aging* **1996**, *17*, 681.

(78) Yan, S. D.; Zhu, H.; Zhu, A.; Golabek, A.; Du, H.; Roher, A.; Yu, J.; Soto, C.; Schmidt, A. M.; Stern, D.; Kindy, M. *Nat Med* **2000**, *6*, 643.

(79) Sousa, M. M.; Du Yan, S.; Fernandes, R.; Guimaraes, A.; Stern, D.; Saraiva, M. J. *J Neurosci* **2001**, *21*, 7576.

(80) Lue, L. F.; Kuo, Y. M.; Roher, A. E.; Brachova, L.; Shen, Y.; Sue, L.; Beach, T.; Kurth, J. H.; Rydel, R. E.; Rogers, J. *Am J Pathol* **1999**, *155*, 853.

(81) McLean, C. A.; Cherny, R. A.; Fraser, F. W.; Fuller, S. J.; Smith, M. J.; Beyreuther, K.; Bush, A. I.; Masters, C. L. *Ann Neurol* **1999**, *46*, 860.

(82) Wang, J.; Dickson, D. W.; Trojanowski, J. Q.; Lee, V. M. *Exp Neurol* **1999**, *158*, 328.

(83) Nilsberth, C.; Westlind-Danielsson, A.; Eckman, C. B.; Condron, M. M.; Axelman, K.; Forsell, C.; Stenh, C.; Luthman, J.; Teplow, D. B.; Younkin, S. G.; Naslund, J.; Lannfelt, L. *Nat Neurosci* **2001**, *4*, 887.

(84) Kitada, T.; Asakawa, S.; Hattori, N.; Matsumine, H.; Yamamura, Y.; Minoshima, S.; Yokochi, M.; Mizuno, Y.; Shimizu, N. *Nature* **1998**, *392*, 605.

(85) Comenzo, R. L.; Vosburgh, E.; Simms, R. W.; Bergethon, P.; Sarnacki, D.; Finn, K.; Dubrey, S.; Faller, D. V.; Wright, D. G.; Falk, R. H.; Skinner, M. *Blood* **1996**, *88*, 2801.

(86) Dember, L. M.; Sanchorawala, V.; Seldin, D. C.; Wright, D. G.; LaValley, M.; Berk, J. L.; Falk, R. H.; Skinner, M. *Ann Intern Med* **2001**, *134*, 746.

(87) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriatt, I.; Muir, T.; Gellman, S. H. *Proc Natl Acad Sci USA* **2001**, *98*, 12015.

(88) Stanger, H. E., Doctoral Dissertation Thesis, University of Wisconsin - Madison, 1999.

(89) Syud, F. A.; Stanger, H. E.; Mortell, H. S.; Espinosa, J. F.; Fisk, J. D.; Fry, C. G.; Gellman, S. H. *J Mol Biol* **2003**, *326*, 553.

(90) Schenck, H. L.; Gellman, S. H. *J Am Chem Soc* **1998**, *120*, 4869.

(91) Freire, F.; Almeida, A. M.; Fisk, J. D.; Steinkruger, J. D.; Gellman, S. H. *Angew Chem Int Ed* **2011**, *50*, 8735.

(92) Fisk, J. D.; Gellman, S. H. *J Am Chem Soc* **2001**, *123*, 343.

(93) Freire, F.; Fisk, J. D.; Peoples, A. J.; Ivancic, M.; Guzei, I. A.; Gellman, S. H. *J Am Chem Soc* **2008**, *130*, 7839.

(94) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. *Nature* **2012**, *491*, 222.

(95) Floudas, C. A.; Fung, H. K.; McAllister, S. R.; Monnigmann, M.; Rajgaria, R. *Chem Eng Sci* **2006**, *61*, 966.

(96) Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew Chem Int Ed* **2013**, *52*, 5700.

(97) Chakrabartty, A.; Baldwin, R. L. *Adv Protein Chem* **1995**, *46*, 141.

(98) Baldwin, R. L.; Rose, G. D. *Trends Biochem Sci* **1999**, *24*, 26.

(99) Nesloney, C. L.; Kelly, J. W. *Bioorgan Med Chem* **1996**, *4*, 739.

(100) Gellman, S. H. *Curr Opin Chem Biol* **1998**, *2*, 717.

(101) Searle, M. S.; Ciani, B. *Curr Opin Struc Biol* **2004**, *14*, 458.

(102) Hughes, R. M.; Waters, M. L. *Curr Opin Struc Biol* **2006**, *16*, 514.

(103) Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. *Proc Natl Acad Sci USA* **2010**, *107*, 10466.

(104) Ranganathan, D.; Haridas, V.; Kurur, S.; Thomas, A.; Madhusudanan, K. P.; Nagaraj, R.; Kunwar, A. C.; Sarma, A. V. S.; Karle, I. L. *J Am Chem Soc* **1998**, *120*, 8448.

(105) Winningham, M. J.; Sogah, D. Y. *Macromolecules* **1997**, *30*, 862.

(106) Diaz, H.; Kelly, J. W. *Tetrahedron Lett* **1991**, *32*, 5725.

(107) Chitnumsub, P.; Fiori, W. R.; Lashuel, H. A.; Diaz, H.; Kelly, J. W. *Bioorgan Med Chem* **1999**, *7*, 39.

(108) Skar, M. L.; Svendsen, J. S. *Tetrahedron* **1997**, *53*, 17425.

(109) Nowick, J. S.; Smith, E. M.; Noronha, G. *J Org Chem* **1995**, *60*, 7386.

(110) Nowick, J. S. *Acc Chem Res* **1999**, *32*, 287.

(111) Fisk, J. D.; Powell, D. R.; Gellman, S. H. *J Am Chem Soc* **2000**, *122*, 5443.

(112) Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. *Annu Rev Biophys* **2013**, *42*, 315.

(113) Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. *Trends Biotechnol* **2014**, *32*, 99.

(114) Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W. *Curr Opin Struc Biol* **2015**, *33*, 16.

(115) Tycko, R.; Wickner, R. B. *Acc Chem Res* **2013**, *46*, 1487.

(116) Hamley, I. W. *Chem Rev* **2012**, *112*, 5147.

Chapter 2. Impact of Strand Number on Parallel β-Sheet Stability

Parts of this work were published by Kung VM, Cornilescu G, and Gellman SH, in *Angew Chem Int Ed Engl* **2015**, *54*, 14336-14339. DOI: 10.1002/anie.201506448.

2.1. Abstract

Here we examine whether parallel β-sheet secondary structure becomes more stable as the number of β-strands increases. We have explored this question with peptides designed to adopt two- or three-stranded parallel β-sheet conformations in aqueous solution. The three-stranded peptide in this report is the first experimental model of a three-stranded parallel β-sheet. We probed, using nuclear magnetic resonance (NMR) spectroscopy and circular dichroism (CD) spectroscopy, the effect on peptide conformation of β-sheet lengthening perpendicular to the strand direction. Our results support the hypothesis that extension, from two strands to three, increases the stability of the parallel β-sheet. We present the first experimental evidence of cooperative folding in the three-stranded parallel β-sheet. Finally, we show how minimal model systems may enable experimental documentation of characteristic properties, such as CD spectra, of parallel β-sheets.

2.2. Introduction

Natural proteins are predominantly comprised of α-helices and β-sheets, the latter of which can be subcategorized into antiparallel and parallel β-sheets. "Minimal models" of these protein secondary structures – i.e., short peptides that can fold autonomously into α-helices or β-sheets, in the absence of protein tertiary structural context – can facilitate the study of the parameters guiding protein folding and protein-protein interactions. In addition to contributing to fundamental understandings of protein structure, stability, and interactions, minimal models of parallel β-sheets are of particular interest for modeling amyloid structures, which appear to contain parallel β-sheets that grow in the direction of increasing strand number.[1-3] The development of minimal models of parallel β-sheets has been challenging, however, and it has consequently lagged behind the development of minimal models of α-helices[4,5] and of

antiparallel β-sheets[6-10]. β-Sheets with any strand orientation are notoriously prone to aggregation, making them difficult to study.[3] An added challenge in the design and synthesis of parallel β-sheet model systems is that non-peptidic turn units (some of which decrease the water-solubility of peptides) must be employed to promote folding in the absence of a stabilizing, tertiary framework.[11-15] In contrast, autonomously-folding α-helices and antiparallel β-sheets can be generated from short sequences composed entirely from α-amino acid residues. Here we report the first three-stranded parallel β-sheet that folds autonomously in aqueous solution. We use this model to explore the impact of the number of β-strands on parallel β-sheet stability.

Understanding how variables such as length affect the stability of a protein secondary structure is fundamental to understanding protein conformations. The relationship between length and stability has been characterized for the α-helix and for the antiparallel β-sheet. The α-helix becomes more stable as its length (along the helical axis) increases.[16-18] The β-sheet has two defining lengths, one along the strands (number of residues per strand, or "strand length") and one perpendicular to the strands (number of strands, or "strand number"). The stability of the antiparallel β-sheet is dependent on both lengths: there are limits to length-dependent stabilization with increasing strand length;[19-21] and stability increases with increasing number of strands, at least up to four strands.[22-24] The existing evidence, though limited (to studies of strand length), indicated that the size-stability relationship for parallel β-sheets is not necessarily identical to that for antiparallel β-sheets.[20,21,25] Parallel β-sheet secondary structure can be stabilized by lengthening the strands, without any limit yet detected, if the added residues have sufficient β-sheet propensity.[25] The relationship between strand number and parallel β-sheet stability has not previously been addressed experimentally.

2.3. Methods

2.3.1. Synthesis and Characterization of Peptides

The OAllyl-Gly-(1*R*,2*S*)-CHDA-Val-OH and Alloc-Glu-Val-^{D}Pro-DADME-Fmoc linkers were synthesized via the routes described in **Appendix 1**. Peptide **2.A** was synthesized via the route depicted in **Figure 2.1**, and peptides **2.B** and **2.C** were synthesized by analogous routes. Peptides were purified by reversed-phase high-performance liquid chromatography (HPLC), using a C18 preparatory column, and a mobile phase of CH_3CN: H_2O:CF_3COOH. Peptide identities were confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (**Figures 2.3**). MALDI-TOF MS showed the following m/z values for the $[M+H^+]^+$ species: peptide **2.A** calculated = 2888.67, observed = 2889.3; peptide **2.B** calculated = 1894.07, observed = 1895.1; and peptide **2.C** calculated = 1726.05, observed = 1726.8. Purity was determined by HPLC, and was >95% for all purified peptides (**Figures 2.2**).

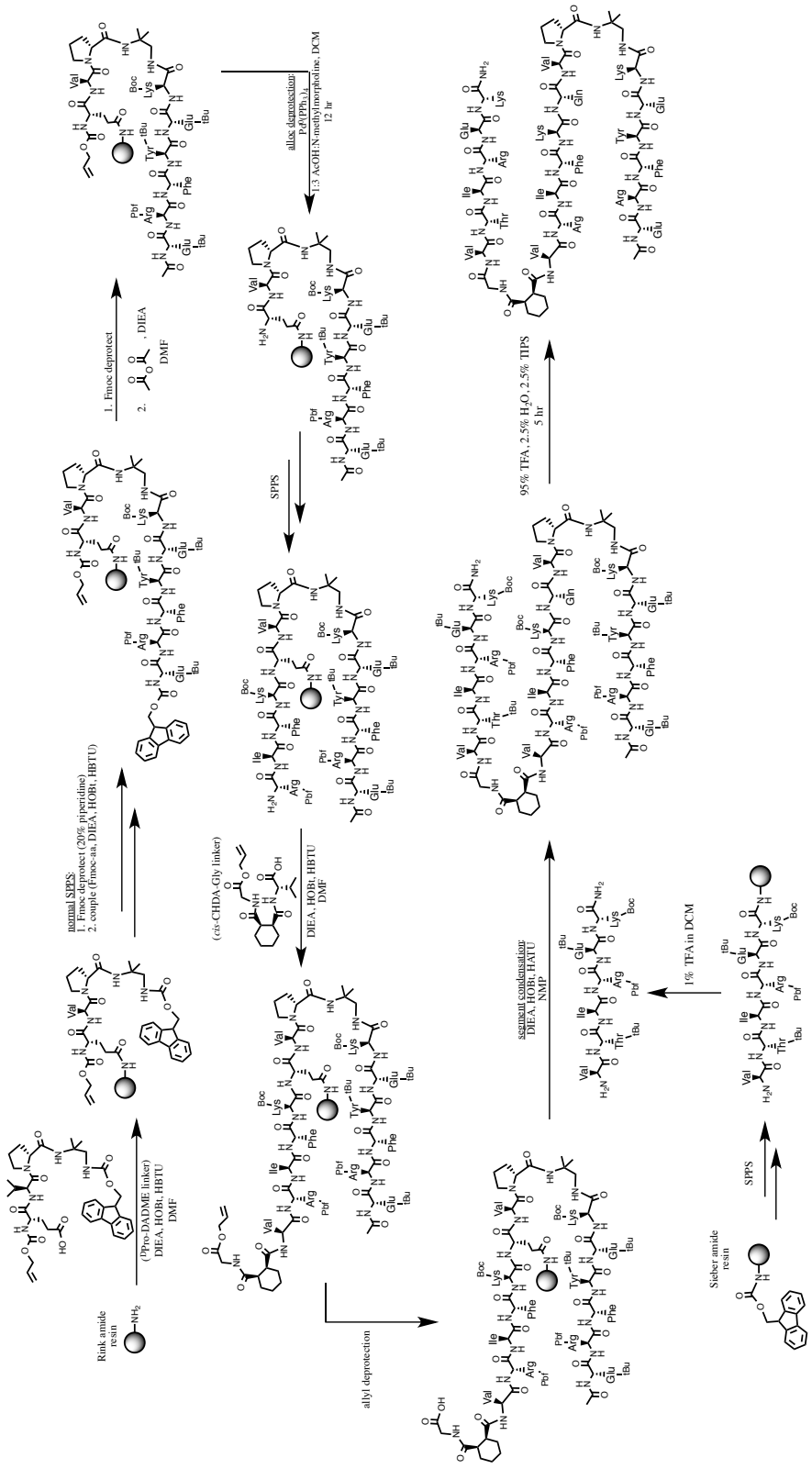

Figure 2.1. Synthetic route for peptide **2.A**. Analogous routes were employed to synthesize peptides **2.B** and **2.C**.

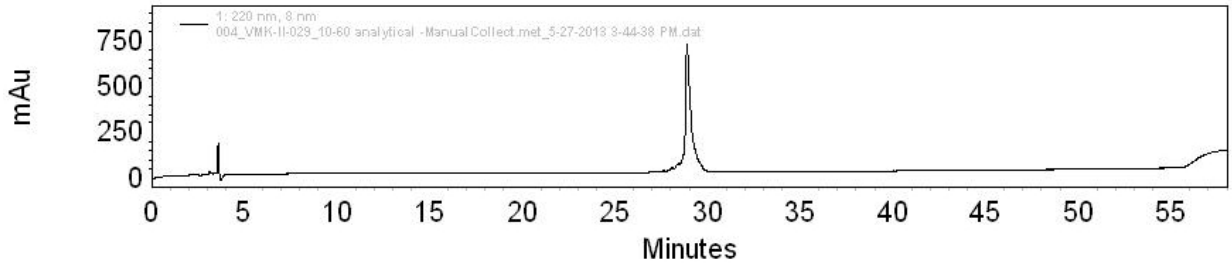

Figure 2.2A. Analytical HPLC trace for peptide **2.A**.

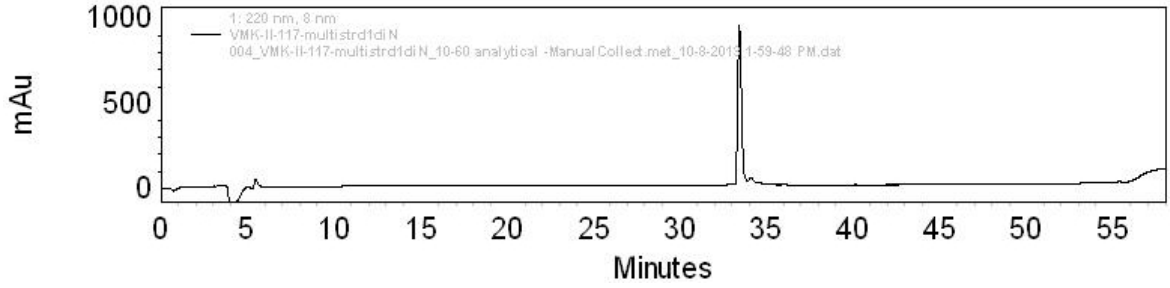

Figure 2.2B. Analytical HPLC trace for peptide **2.B**.

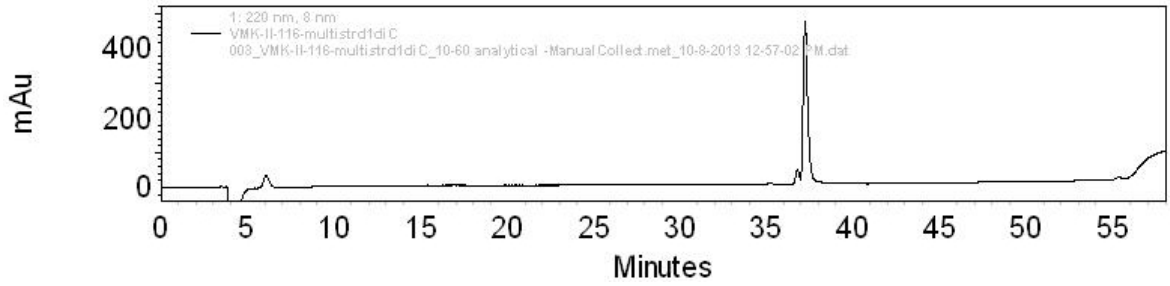

Figure 2.2C. Analytical HPLC trace for peptide **2.C**.

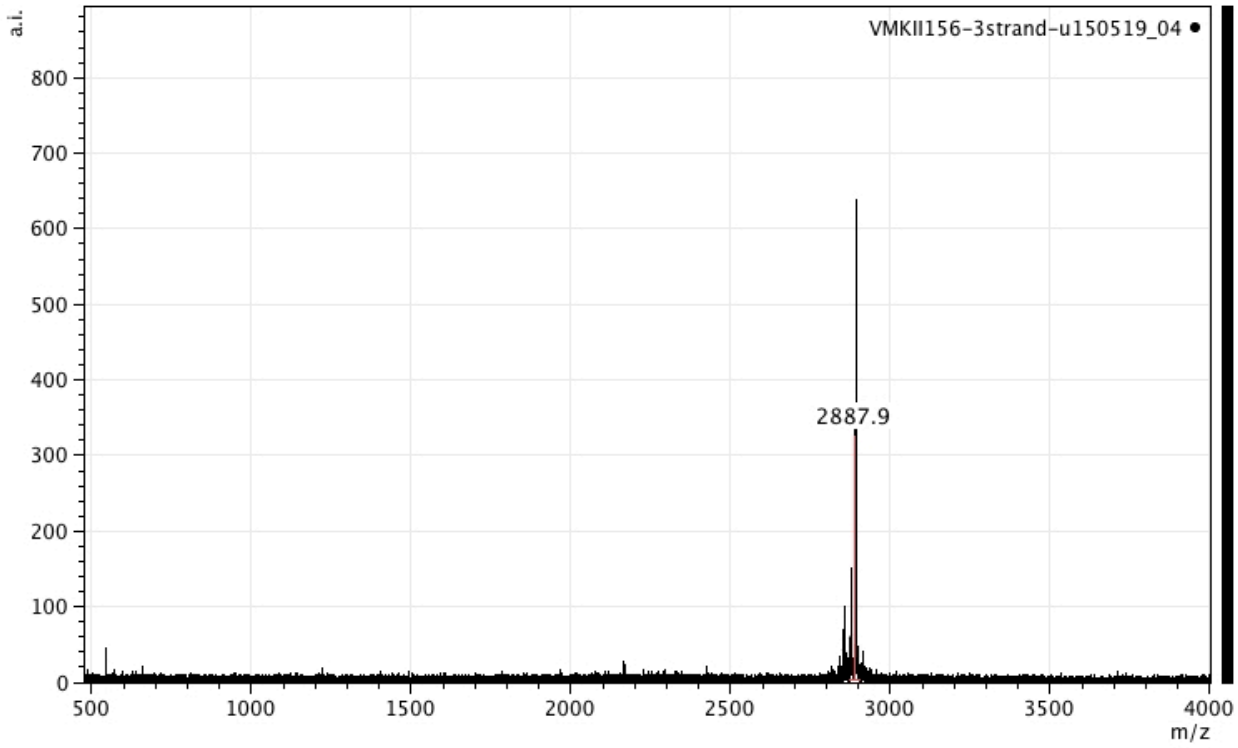

Figure 2.3A. MALDI-TOF MS of peptide **2.A**.

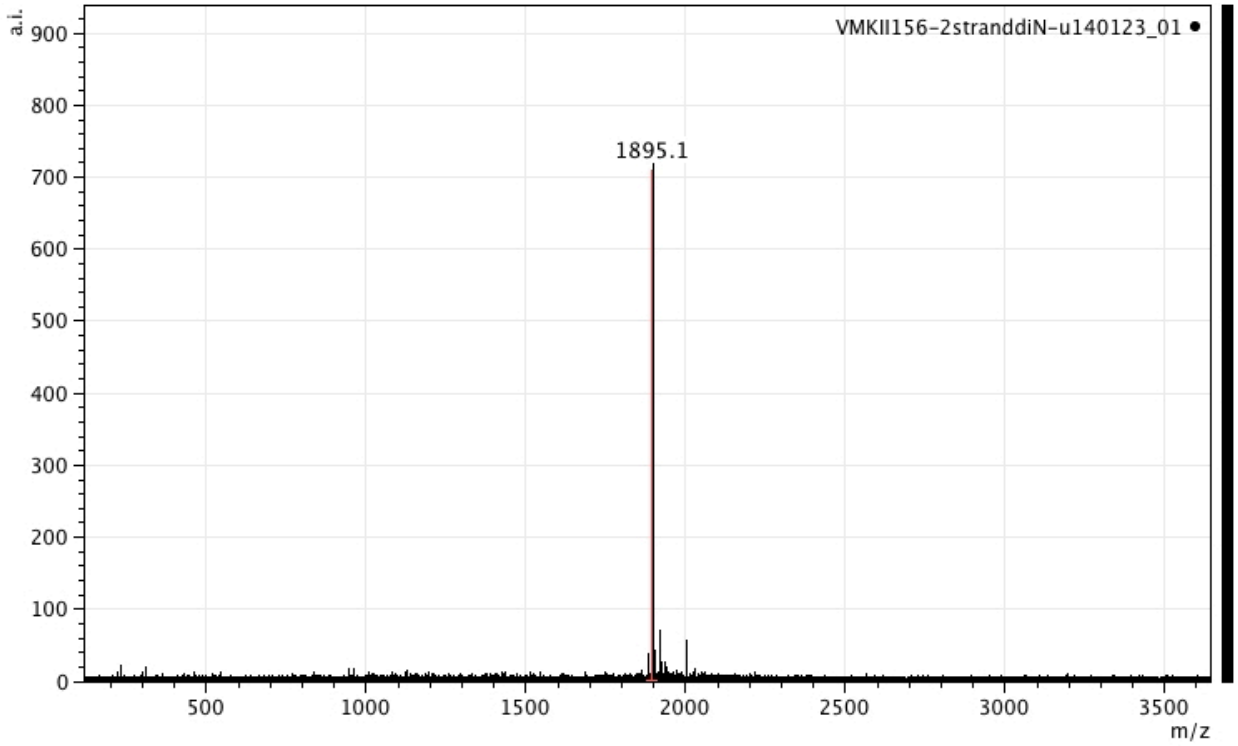

Figure 2.3B. MALDI-TOF MS of peptide **2.B**.

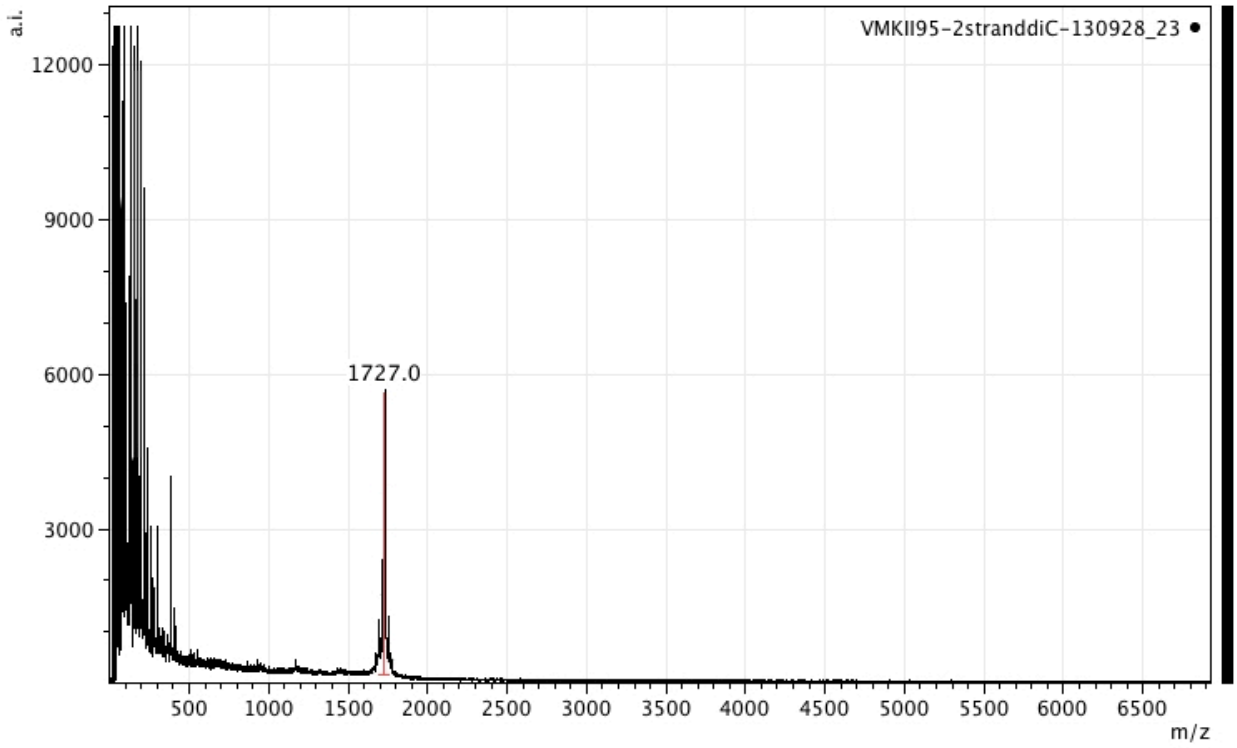

Figure 2.3C. MALDI-TOF MS of peptide **2.C**.

2.3.2. 2D-NMR data acquisition and NMR structure determination

Each NMR sample was prepared by dissolving lyophilized peptide to 2 mM concentration in 9:1 H_2O:D_2O, pH 3.8, 3 mM sodium acetate-d_3 buffer, with trace amounts of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as an internal reference. Peptide concentrations were determined by mass. Samples were prepared with total volumes of approximately 300 µL in 3 mm NMR tubes. Peptide samples were stable in solution for at least days, showing no visible precipitation of peptide (**Figure 2.4**), NMR line broadening, or decrease in NMR signal intensity over the entire period of study. Low salt conditions were employed to facilitate observation of weaker NMR signals (**Figures 2.5**, **2.6**, **2.7**). For each peptide, the following sets of 1D ^{1}H spectra were superimposable: [1] before and after collection of 2-D NMR spectra (COSY, TOCSY, ROESY), and [2] 2 mM peptide in low salt buffer (9:1 H_2O:D_2O, 3 mM

sodium acetate-d_3, pH 3.8, trace DSS) and 2 mM peptide in high salt buffer (9:1 H_2O:D_2O, 100 mM sodium acetate-d_3, pH 3.8, trace DSS).

A. **B.** **C.**

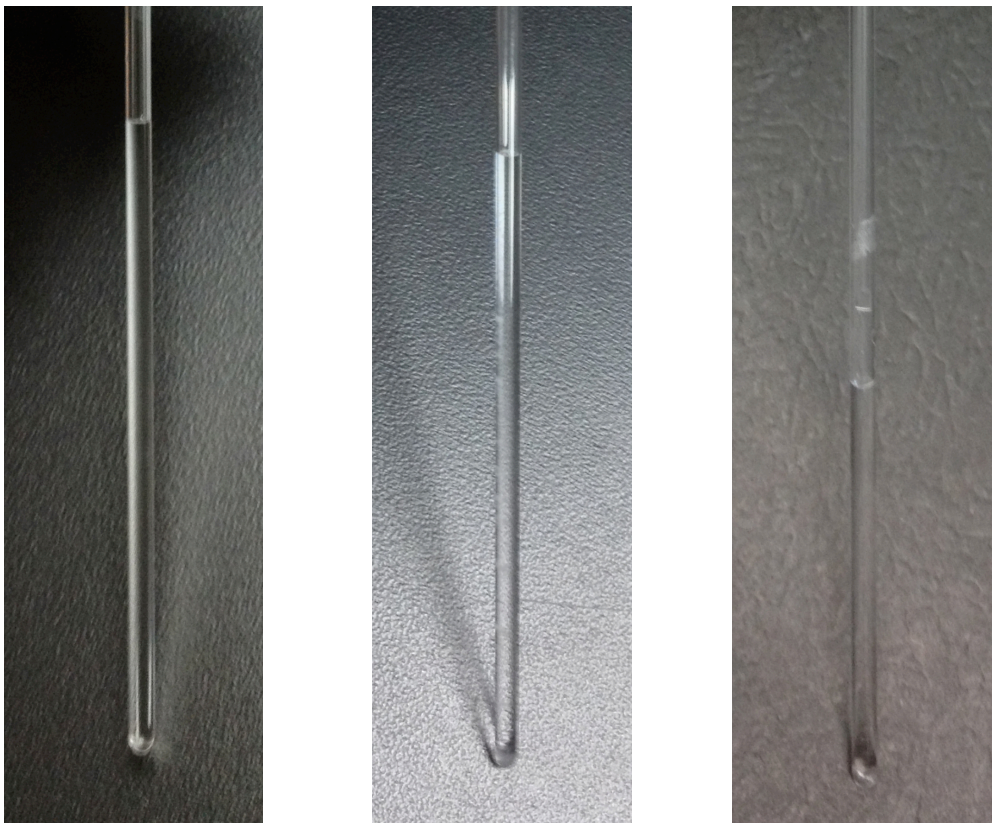

Figure 2.4. No visible heterogeneity, formation of viscous gels, or precipitation from NMR tubes, for duration of NMR experiments, for peptides **2.A**, **2.B**, and **2.C**.

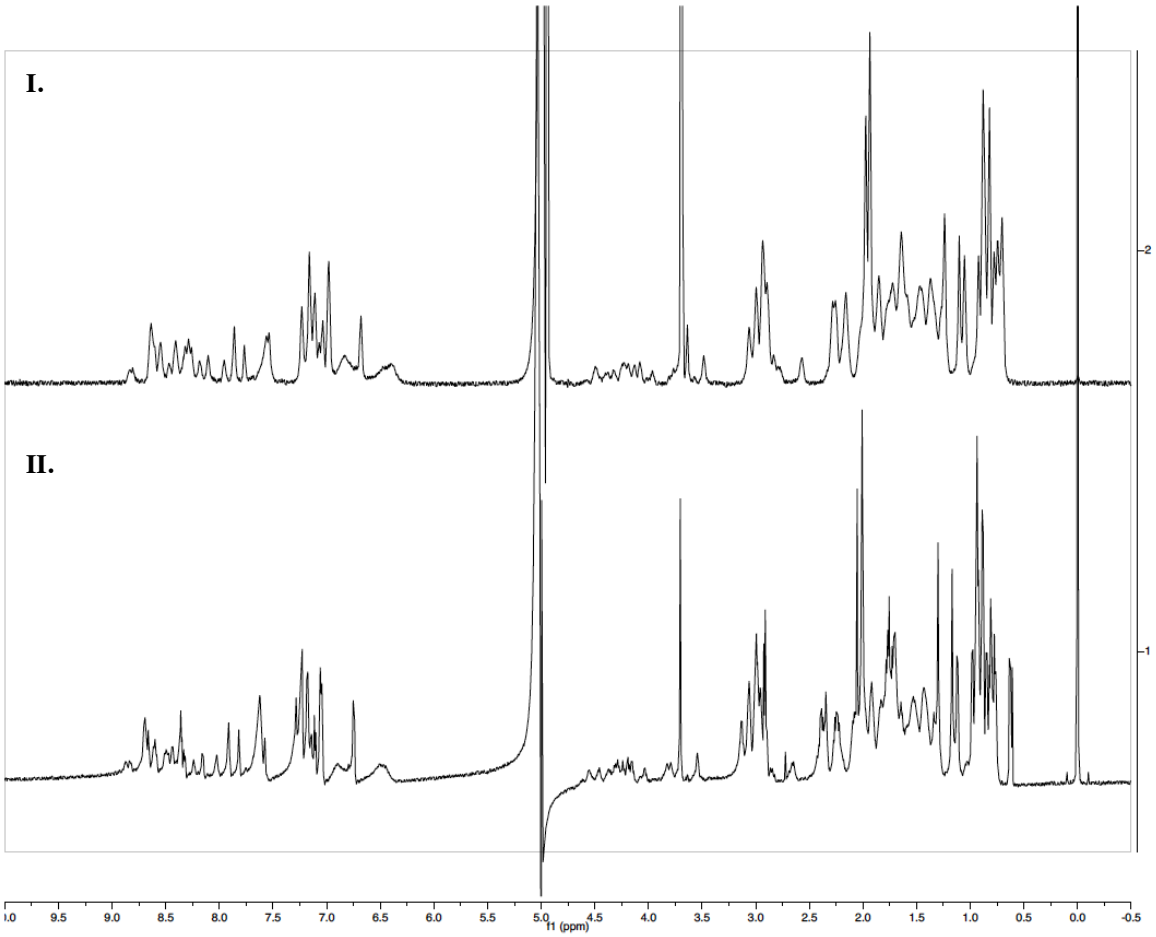

Figure 2.5. ^{1}H NMR spectra of 2 mM peptide **2.A**, at (I) 100 mM NaOAc-d_3, pH 3.8, and (II) 3 mM NaOAc-d_3, pH 3.8.

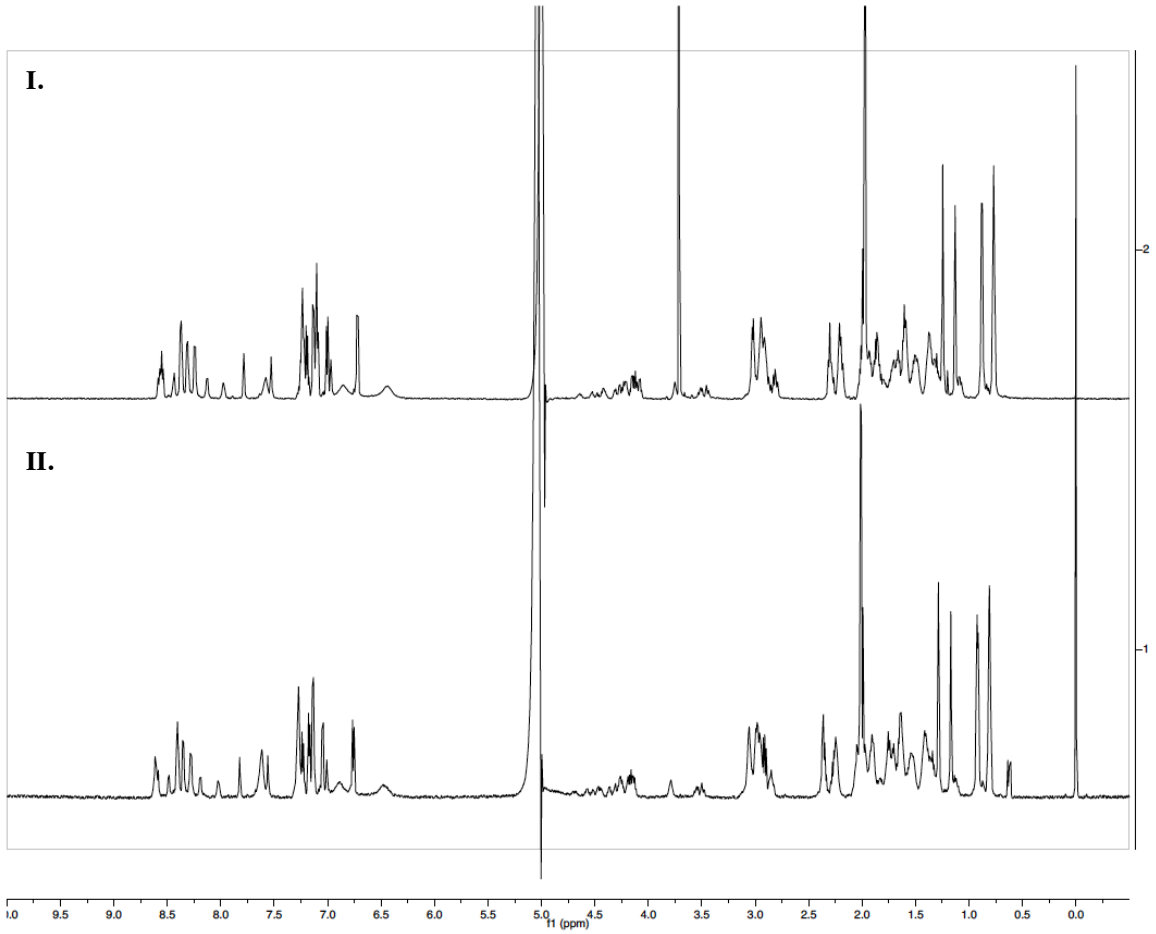

Figure 2.6. ^{1}H NMR spectra of 2 mM peptide **2.B**, at (I) 100 mM NaOAc-d_3, pH 3.8, and (II) 3 mM NaOAc-d_3, pH 3.8.

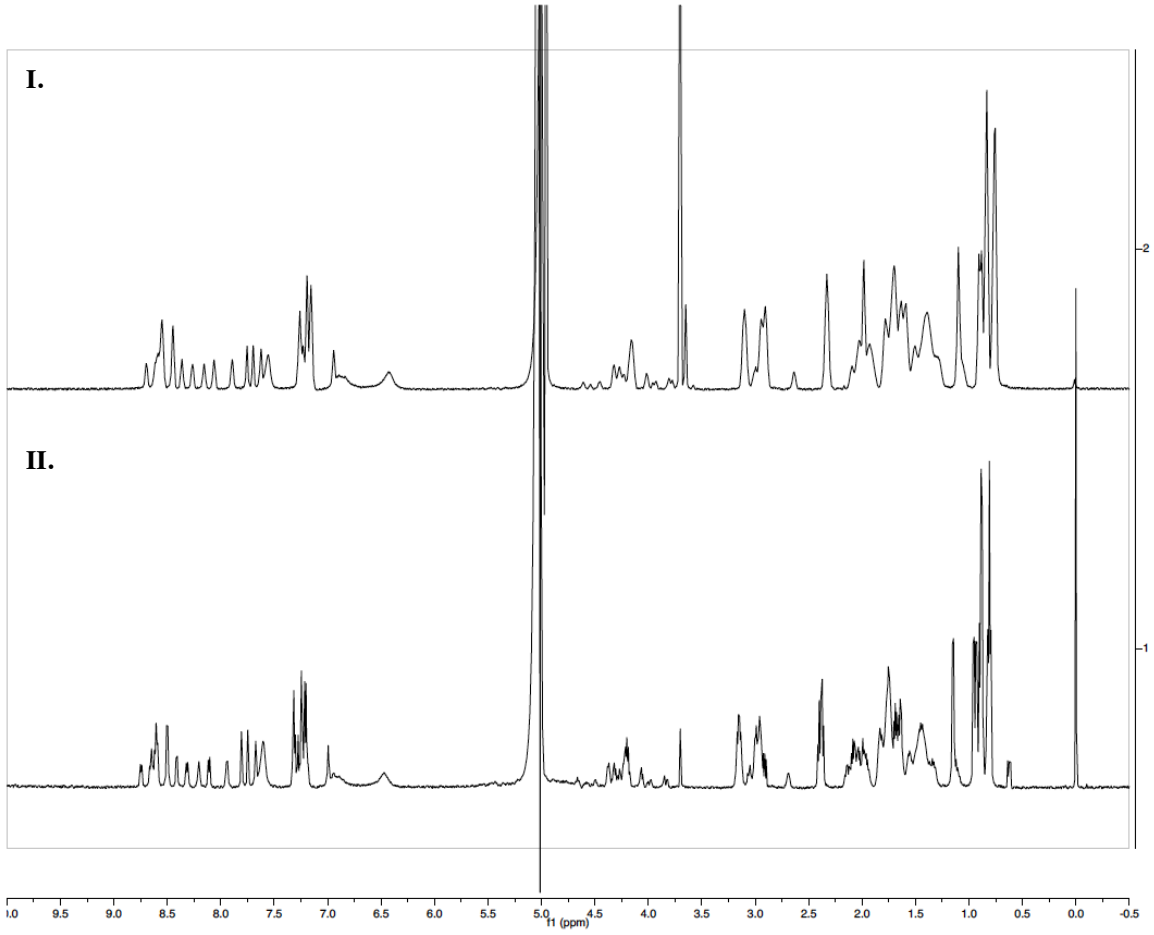

Figure 2.7. ^{1}H NMR spectra of 2 mM peptide **2.C**, at (I) 100 mM NaOAc-d_3, pH 3.8, and (II) 3 mM NaOAc-d_3, pH 3.8.

All NMR experiments were performed on a Varian INOVA 600 MHz spectrometer at 277.2 K (4.0 °C), using a 3 mm probe (Varian 3 mm inverse ^{1}H/^{13}C/^{15}N with 3-axis PFG and VT 0 to +50 C). Standard Varian pulse sequences – presat, gcosy, wgtocsy, and wgroesy – were employed. Solvent suppression was achieved using the following methods: 1.5 sec presaturation during relaxation delay for 1D, 2.0 sec presaturation during relaxation delay for COSY, and Watergate 3-9-19 solvent suppression (as implemented in the Varian BioPack suite of experiments) for TOCSY and ROESY. COSY spectra were acquired in the absolute value mode with gradient coherence selection; TOCSY and ROESY spectra were acquired in the phase-sensitive mode with hypercomplex phase cycling (States-Haberkorn method). TOCSY

experiments used an MLEV-17 spin-lock field of 8 kHz, and mixing time of 100 ms. ROESY experiments used a flip-flop spin-lock field of 6 kHz, and mixing time of 250 ms. Spectral widths of 7000 Hz were used, with collection of 1492-2982 points in the direct (f2) and 116-291 points in the indirect (f1) dimensions. Data were processed using Varian VNMR 6.1 software, with application of the following window functions (along both dimensions) before Fourier transformations: sinebell for COSY, Gaussian for TOCSY, and shifted sinebell for ROESY. Data were analyzed using the Sparky program,[26] with employment of sequential assignment procedures to assign chemical shifts of protons (**Tables 2.3**).[27]

ROE peak intensities were translated into a continuous distribution of interproton distance restraints, with a 35% distance uncertainty applied to take into account spin diffusion. For inferred hydrogen bonds, distance restraints of 1.9 Å and 2.9 Å were used for the H_N-O and N-O non-bonded distances, respectively.[27] To check for biases caused by hydrogen bond restraints, structure calculations were performed without these restraints; there were no major changes in the resulting strand registers or in the backbone conformations in the residues for which hydrogen bond inferences had been made. The secondary chemical shift, or difference between an observed chemical shift and random coil values, was calculated for each α-proton of each peptide (**Figure 2.22**). Calculated chemical shifts for peptides in random coil conformations (in H_2O, pH 3.80, 277.15 K) were obtained from the CSDb[2] program, which takes into account nearest-neighbor and other effects.[28] Compared to random coil values, amino acid residues experience a mean α-proton downfield shift of 0.37 ppm when placed in a β-strand or extended configuration (and a mean α-proton upfield shift of 0.39 ppm when placed in a helix).[29] A gross pattern of downfield α-proton chemical shifts can be a sensitive indicator of β-strand backbone dihedral angles.[30-32] For regions in which a gross pattern of downfield α-proton

chemical shifts was observed, β-strand backbone dihedral angle ($\Phi \in [-180, -110]$, $\Psi \in [90, 180]$)[33] restraints were applied for residues with ^{1}H secondary chemical shifts > 0.1 ppm. To check for biases caused by backbone dihedral angle restraints, structure calculations were performed without these restraints; there were no major changes in the resulting backbone conformations.

Three-dimensional structure calculations and refinements, with the described structural restraints, were carried out using the torsion angle molecular dynamics and the internal variable dynamics modules of Xplor-NIH (v. 2.35), with patches for the non-natural residues.[34] The target function minimized was comprised of the experimental NMR restraints (ROE-derived interproton distances, inferred hydrogen bonds and torsion angles), a repulsive van der Waals potential for the non-bonded contacts,[35] a torsion angle database potential of mean force,[36] and a gyration volume potential[37]. The initial topology and parameter files for non-natural residues (DADME, CHDA) were generated using the PRODRG2 Server.[38]

An ensemble of the 10 lowest energy (out of 100) structures for each peptide is shown in **Figure 2.18**, and the average of each ensemble is shown in **Figure 2.21**. For each peptide, there were no (0) consistent violations of restraints (ROE-derived distance, hydrogen bond, dihedral angle). Structural statistics for all (not just ordered) residues were generated using the Protein Structure Validation Suite (PSVS) version 1.5,[39] and are reported in **Tables 2.4**. The structures of these peptides in solution are more dynamic than is typical of well-ordered globular proteins, such that the ROE intensities represent averages over multiple conformers, but we make the simplifying assumption of there being a single set of structures. One result of this assumption is that reported clash scores are poorer than those typical of standard PDB structures.

2.3.3. Circular dichroism (CD) spectroscopy

CD measurements were taken with an AVIV Model 420 Circular Dichroism Spectrometer, using a 1.00 nm bandwidth and 3.000-10.000 sec averaging times, in 1.00 mm quartz cuvettes. For each sample measurement, the corresponding buffer blank measurement was subtracted, and the baseline CD signal at 320 nm was set to zero. Peptide solutions, with five concentrations ranging from 0.5 mM to 0.05 mM, were prepared by serial dilution in 10 mM sodium acetate buffer, pH 3.8. Peptide concentrations were determined by mass, and verified with UV absorption measurements (Cary 50 scan UV Vis spectrophotometer) if a Tyr was present in the sequence. For each peptide, CD spectra were superimposable after normalization for concentration, supporting that peptide structure was not concentration-dependent in the 0.05-0.5 mM range. CD spectra for 0.2 mM peptide samples in 10 mM sodium acetate buffer, pH 3.8, at 20.0 °C (293.2 K) are reported (**Figure 2.16** solid lines). For each peptide (except **2.C** in the absence of TFE), a single minimum in the CD signal was observed near 214 nm (214 nm for **2.A**, 216 nm for **2.B**, and 213 nm for **2.C**), and the CD signal at this minimum was monitored as a function of temperature for thermal denaturation experiments. Addition of 50% (v/v) 2,2,2-trifluoroethanol (TFE) co-solvent enhanced CD signals (**Figure 2.16** dashed lines), consistent with TFE-mediated promotion of secondary structure.[40] Thermal melt curves for samples in the presence of TFE are reported in **Figure 2.24**. Measurements were taken every 2 °C, in the 2.0-78.0 °C (275.2-351.2 K) temperature range in the presence of TFE, and in the 2.0-98.0 °C (275.2-371.2 K) temperature range in the absence of TFE. CD signals were also measured every 2 °C as solutions were cooled, and heating and cooling transitions were found to be superimposable, indicating that unfolding reactions were reversible.

2.3.4. Sedimentation Equilibrium Analytical Ultracentrifugation (AUC)

Sedimentation equilibrium data were collected using a Beckman Coulter Model XL-A Analytical Ultracentrifuge at 4 °C. Peptide solutions were prepared by simple dissolution in buffer (3 mM acetate, pH 3.8, with peptide concentrations of 0.1, 0.3, and 0.4 mM) and used without further manipulation. Initial peptide concentrations were determined from UV absorption spectra of samples diluted into 6 M GdmCl, using an extinction coefficient (ε) of 1280 M^{-1} cm^{-1}.[41] Note that the peptide concentrations studied were chosen to span the absorbance range detectable by AUC. AUC experiments were performed for peptides **2.A** and **2.B**, each of which contains a Tyr residue; however, experiments were not conducted on peptide **2.C**, due to its lack of aromatic chromophores. For each sample, 100 μL of a peptide solution was placed in the sample sector of a 1.2 cm pathlength (l) charcoal filled Epon centerpiece, with 110 μL of the buffer in the reference sector. Gradients were monitored at the wavelength of maximal absorbance (275 nm for **2.A**, and 272 nm for **2.B**). Data were collected at rotor speeds of 24k, 42k, 48k, 54k, 60k rpm for **2.A**, and 36k, 42k, 50k, 58k rpm for **2.B**. At speeds > 45k rpm, as yet unexplained spectral anomalies were observed at high radial positions, so we restricted all analyses to lower rotor speeds (< 45k rpm). For each rotor speed, we assumed samples had reached equilibrium when gradients collected at least 3 hours apart were superimposable. After data collection at the highest rotor speed, gradients at one of the lower speeds were measured again and found to be superimposable with original measurements, indicating that there had been no irreversible loss of material during the course of the experiment.

The equilibrium data were analyzed using software written by Darrell R. McCaslin, in a manner analogous to previously-described analysis methods.[42] In the analysis, the solvent

density (ϱ) at 4 °C was computed using density increment functions to be 1.001 g/mL.[43] The partial specific volumes ($\bar{v}$) of **2.A** (0.748 $cm^3\ g^{-1}$), and **2.B** (0.734 $cm^3\ g^{-1}$) were calculated using consensus values for the natural amino acid residues, and corrections for acetyl, carboxamide, DADME, and CHDA groups.[44,45] The molecular weights (M_S) of **2.A** (2888.5 Da) and **2.B** (1894.3 Da) were calculated based on the peptide sequences, including non-natural linkers and termini.

Final estimates of the weight-average molecular weights (M_P, or MW_{obs}), reported in **Table 2.1**, were obtained from global fits of the data at the varied loading concentrations and speeds to the distribution of a single macromolecular species, according to **Equation 2.1**.

Equation 2.1:
$$c\left(r^2\right) = BOD + c_r \exp\left(\frac{M_P(1-\bar{v}\rho)\omega^2}{2RT}\left(r^2 - r_r^2\right)\right)$$

The variables in **Equation 2.1** are defined as follows: c = concentration at radial position r, c_r = concentration at an arbitrary reference position r_r (one c_r for each speed and loading concentration), $\bar{v}$ = partial specific volume, ϱ = solvent density, ω = rotor speed (in radians/s), R = gas constant, T = absolute temperature, and baseline optical density (BOD) is a baseline absorbance correction for non-sedimenting species. Due to the small molecular weights of the peptides studied, monomeric forms could not be completed depleted from the solution column even at maximal attainable speeds (60k rpm), preventing an experimental evaluation of the contributions of non-sedimenting materials. The BOD was allowed to float in some fits, but returned values larger than those measured at the highest speed; therefore a BOD value of 0 was assumed. The reduced molecular weight, $M_R = M_P(1-\bar{v}\,\varrho)$, was used as the fitting parameter and converted to M_P or M_P/M_S after the fit. This permits the evaluation of the impact of errors in $\bar{v}$ and ϱ during the fit, and permits post hoc evaluation of the aggregation state with various values

of these parameters. Inclusion of a term for a second macromolecular species in the above worsened fits, and generally yielded M_R's for the two species not significantly different from eachother and/or gave c_r's that were negative for one species.

To probe the contribution of electrostatic repulsion to the nonideal behavior of the peptides observed at low-salt conditions, AUC studies were performed with buffer containing 100 mM neutral salt (NaCl). Data collection and analysis was performed as above. Note that the density of the "with salt" buffer (3 mM acetate, 100 mM NaCl) at 4 °C was computed to be 1.005 g/mL.[43] For the peptide **2.A** (0.4 mM) sample with salt, data were collected at rotor speeds of 24k, 36k, 40k, 42k, 48k (**Figure 2.9II**).

2.3.5. NMR diffusion experiments

Experiments measuring peptide diffusion by NMR were performed, using the water-suppressed LED (water-sLED) pulse program on a Varian INOVA 600 MHz spectrometer.[46] Samples were prepared by dissolving lyophilized peptide (2 mM) in buffer (9:1 H_2O:D_2O, pH 3.8, 100 mM d_3-acetate, 0.6 mM dioxane). Serial dilutions (10x, 100x) were made of these peptide solutions into buffer, to yield 0.2 mM and 0.02 mM peptide solutions. For each peptide condition, NMR spectra were acquired at 4.0 °C, with 11 pulsed field gradient (PFG) strengths in the 0-57.5 G/cm range, and diffusion delay of 0.1 s. Variations in the natural log of NMR signal intensities (lnA) as a function of the square of the gradient strength (G^2) are shown for the three peptides at 2 mM concentration (**Figure 2.11**) and 0.2 mM concentration (**Figure 2.12**). The R^2 value was > 0.95, for every linear regression line. The slope of the (lnA vs. G^2) plot is proportional to the molecule's diffusion coefficient (D), or the inverse of the hydrodynamic radius (R_H), given **Equation 2.2**, in which A = NMR signal intensity, γ = gyromagnetic ratio of ^{1}H, δ = duration of PFG, G = PFG strength, D = constant of self-diffusion, and Δ = time between

beginnings of two gradient pulses. For each condition studied, the hydrodynamic radius (R_H) of the peptide was extracted by comparison of peptide diffusion data with that of an internal standard, 1,4-dioxane (R_H = 2.12 Å).[47]

Equation 2.2: $$A = A_0 e^{\left[-(\gamma\delta G)^2\left(\Delta-\frac{\delta}{3}\right)D\right]}$$

For the above NMR diffusion experiments, buffer containing 100 mM salt was employed to control for salt effects – i.e., such that [salt] >> [peptide], for all peptide concentrations studied. A second set of NMR diffusion experiments was conducted using 3 mM NaOAc, and those results are also reported in **Table 2.2**. Of note, due to changes in equipment availability, the low [salt] NMR diffusion experiments could not be performed using the spectrometer and probe that were used for all other NMR experiments reported herein. They were instead performed on an Agilent NMR System 600II with a $^1H(^{13}C/^{15}N)$ cryogenic probe, in 5 mm Shigemi tubes matched with D_2O, and employed the Doneshot pulse program provided by VnmrJ.[48] NMR spectra were acquired at 4.0 °C, with 14 pulsed field gradient (PFG) strengths in the 0-43.1 G/cm range, and diffusion delay of 0.1 s.

2.3.6. NMR dilution experiments

See also **section 2.3.5**. 1D ^1H NMR spectra were taken on a Varian INOVA 600 MHz spectrometer, for peptides **2.A** (**Figure 2.13**), **2.B** (**Figure 2.14**), and **2.C** (**Figure 2.15**), at three concentrations: 2 mM (64 scans), 0.2 mM (64 scans), and 0.02 mM (640 scans).

2.4. Results

2.4.1. Peptide design

We designed the model system depicted in **Figure 2.8**, studying parallel β-sheet interactions between "strand 1" (residues 1-6) and "strand 2" (residues 14-9) in peptides **2.A** and

2.B, and "strand 2" (residues 15-10) and "strand 3" (residues 18-23) in peptides **2.A** and **2.C**. The sequences of the strands were designed to have high propensities for parallel β-sheet folding, and with some effort to preserve sequences used in other parallel β-sheet model systems, such that comparisons might be made between studies.

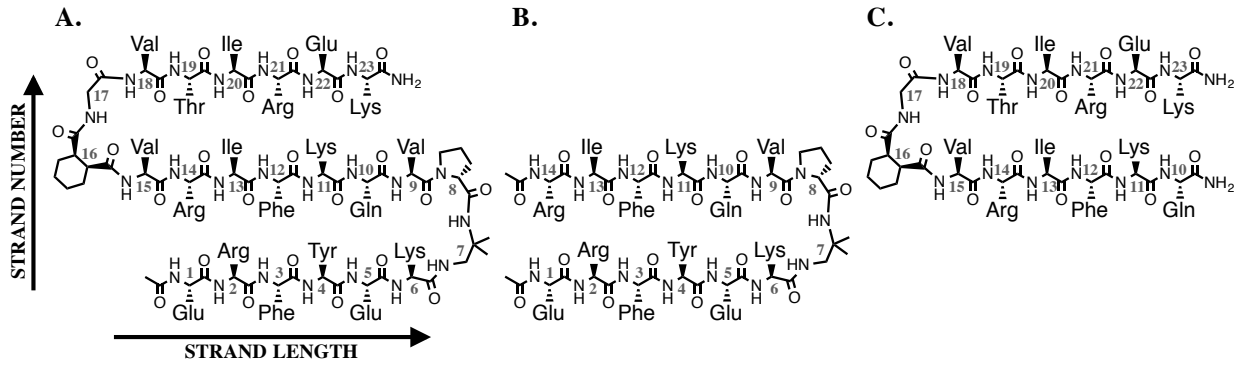

Figure 2.8. *De novo* design of (A) peptide **2.A**, devised as described in the text to adopt a three-stranded parallel β-sheet conformation. Peptides (B) **2.B** and (C) **2.C** are the two possible truncations of peptide **2.A** to two-stranded peptides, with **2.B** containing a non-natural diamine linker, and **2.C** containing a non-natural diacid linker. We employ these peptides to study the effect of increasing the number of parallel β-strands from two to three.

To connect strands, we used the following non-natural linkers, which have been shown to promote, but not strictly enforce, parallel β-sheet interactions: ^{D}Pro-DADME, a diamine linker derived from D-proline (^{D}Pro) and 2-methylpropane-1,2-diamine (a.k.a. 1,2-diamino-1,1-dimethylethane, or DADME), and (1*R*,2*S*)-CHDA-Gly, a diacid linker derived from (1*R*,2*S*)-1,2-cyclohexanedicarboxylic acid ((1*R*,2*S*)-CHDA) and glycine (Gly).[13,14] Note that there are two possible ways to connect three strands with one diamine linker and one diacid linker: one as drawn in **Figure 2.8A**, and the other with a diacid (rather than diamine) linker connecting the first and second strands, and a diamine (rather than diacid) linker connecting the second and third strands. In theory either of the two possible connectivities should be suitable for *de novo* design of three-strand models. We employed the former connectivity here, because it supported strand sequences more similar to those used in previously developed two-strand

models. In particular, strand 2 (residues 15-9) was designed to maximally preserve β-strand sequences used in multiple previous model systems.[14,25] Compared to the longer, peptidic segments that link parallel β-strands in natural proteins, our linking units are likely less flexible, so are expected to favor parallel β-sheet interactions entropically. Moreover, natural linking segments can engage in tertiary interactions, whereas our linking segments enable the study of secondary interactions in isolation.

The primary sequences of the β-strands were chosen with several considerations in mind. First, they were chosen to promote parallel β-sheet folding, based on intrinsic propensities of amino acids to assume β-strand conformations,[49,50] amino acid pairing preferences between adjacent strands of parallel β-sheets,[51] and published sequences for parallel β-sheet model systems that fold in aqueous solution[14,25]. Second, charged residues were incorporated to promote water-solubility and prevent aggregation; positively-charged residues were favored over negatively-charged residues because the former have slightly higher β-strand propensities.[6,50] Third, given the difficulty of synthesis of parallel β-sheet model peptides with more than two strands (**Figure 2.1**), amino acids that are known to add difficulty peptide syntheses (e.g., Asp)[52] were generally avoided, as long as this avoidance did not significantly affect to other design criteria. Fourth, the peptide included as much residue diversity as was allowed by the aforementioned considerations, to facilitate NMR interpretation.

2.4.2. Peptides are monomeric

Sedimentation equilibrium analytical ultracentrifugation (AUC) (**Figures 2.9I** and **2.10**, **Table 2.1**), NMR diffusion (**Table 2.2**), and variable concentration CD and NMR experiments (**Figures 2.13**, **2.14**, **2.15**) suggest that peptides **2.A**, **2.B**, and **2.C** do not oligomerize or aggregate under the aqueous solution conditions used to acquire

spectroscopic data for structural studies (**section 2.4.3**).

The AUC experiments were designed to examine the peptide association state at low-salt buffer conditions similar to those employed for NMR studies. For both **2.A** (**Figure 2.9I**) and **2.B** (**Figure 2.10**), plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center of rotation were linear, at all speeds and for all loading concentrations. Linearity in such plots is indicative of the presence of a single molecular weight entity. For both peptides, the weight-average molecular weight (MW_{obs}) was much smaller than the molecular weight based on the sequence (Ms). Furthermore, the apparent weight decreased with increasing concentration. These results suggest that at the low-salt conditions studied, the peptides are monomeric, and not forming dimers or larger molecular weight aggregates. With salt, the MW_{obs} of **2.A** was still smaller than, but much closer to, the MW_{calc} (**Table 2.1**), again supporting a monomeric state. Thus we conclude that the observed $MW_{obs}/MW_{calc} \leq 1$ indicates a monomeric state, with electrostatic nonideality accounting for $MW_{obs}/MW_{calc} << 1$.

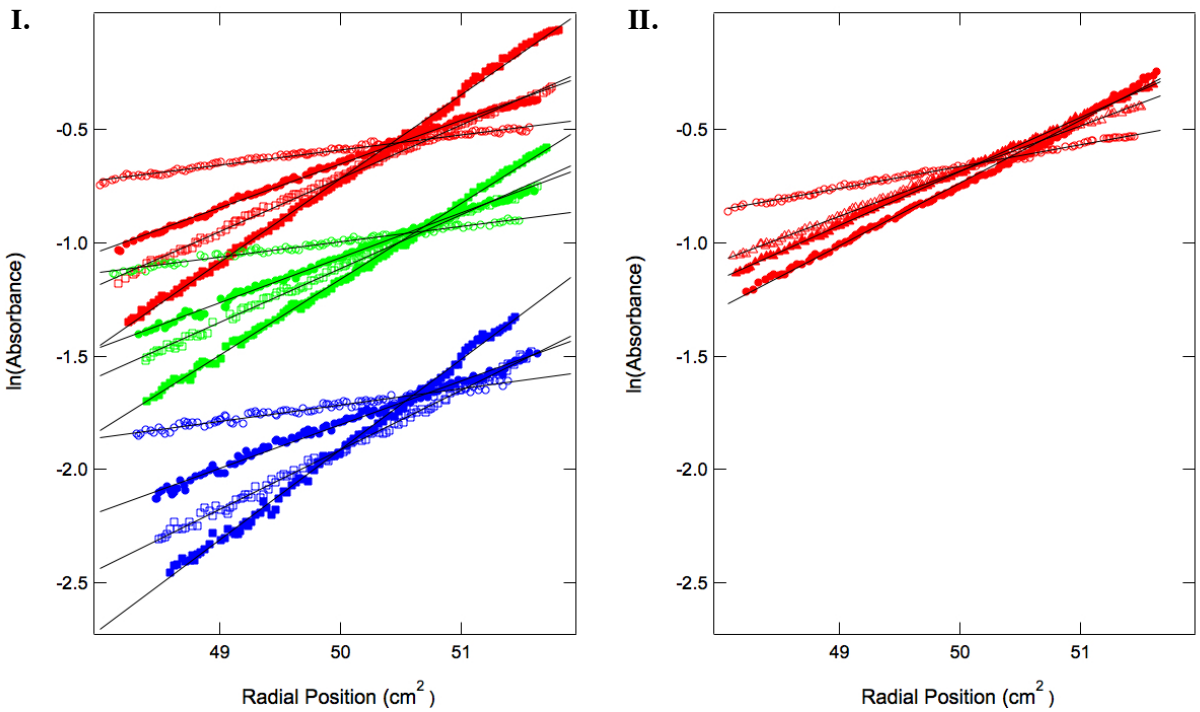

Figure 2.9. (I) Equilibrium AUC data for peptide **2.A** in 3 mM acetate, pH 3.8. The data shown are for speeds of 24k, 42k, 48k, 54k rpm, at initial loading concentrations of 0.1 mM (blue), 0.3 mM (green), and 0.4 mM (red). (II) Equilibrium AUC data, with an additional 100 mM NaCl in buffer. The data shown are for speeds of 24k, 36k, 40k, 42k rpm, at an initial loading concentration of ~0.4 mM peptide.

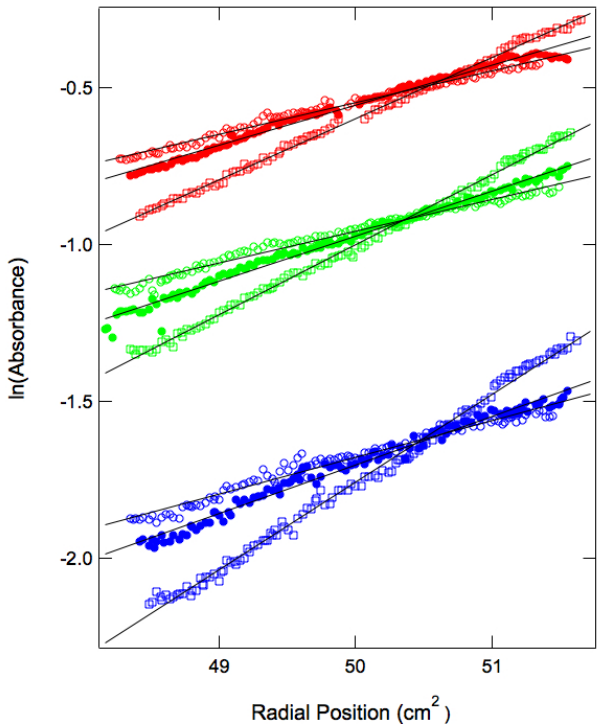

Figure 2.10. Equilibrium AUC data for peptide **2.B** in 3 mM acetate, pH 3.8. The data shown are for speeds of 36k, 42k, 50k rpm, at initial loading concentrations of 0.1 mM (blue), 0.3 mM (green), and 0.4 mM (red).

Table 2.1. Sedimentation equilibrium results for **2.A** and **2.B**. Weight-average molecular weight (MW_{obs}) from fits to a single-species model, reported with least-squares fitting error in the fitted parameter.

peptide	**buffer**	**MW_{obs}/M_S**	**MW_{obs} (Da)**
2.A	3 mM NaOAc, pH 3.8	0.631 ± 0.004	1822
	3 mM NaOAc, 100 mM NaCl, pH 3.8	0.911 ± 0.007	2632
2.B	3 mM NaOAc, pH 3.8	0.686 ± 0.006	1300

NMR diffusion experiments were performed in buffer containing 100 mM salt, to control for salt effects – i.e., such that [salt] >> [peptide], for all peptide concentrations studied. The hydrodynamic radii derived from NMR diffusion experiments (obs R_H) are reported in **Table 2.2**. The predicted hydrodynamic radii of the monomeric peptides were calculated using the atomic-level (a = 2.84 Å) shell-model of HYDROPRO 10,[53] and are in good agreement with the hydrodynamic radii derived from NMR diffusion data (**Table 2.2**). Furthermore, variable concentration NMR diffusion analysis of the peptides show little to no increase in R_H values as

peptide concentrations are increased from 0.2 mM to 2 mM. These results support the conclusion that peptides are monomeric (no peptide aggregation occurs) at the conditions used to obtain NMR structures of the peptides.

A second set of NMR diffusion experiments was conducted using 3 mM NaOAc, and those results are also reported in **Table 2.2**. As with the high [salt] experiments, for each peptide at low [salt] conditions, there was no increase in the peptide's R_H value as the peptide concentration was increased, supporting the conclusion that the peptide is monomeric in the 0.2 to 2 mM peptide concentration range.

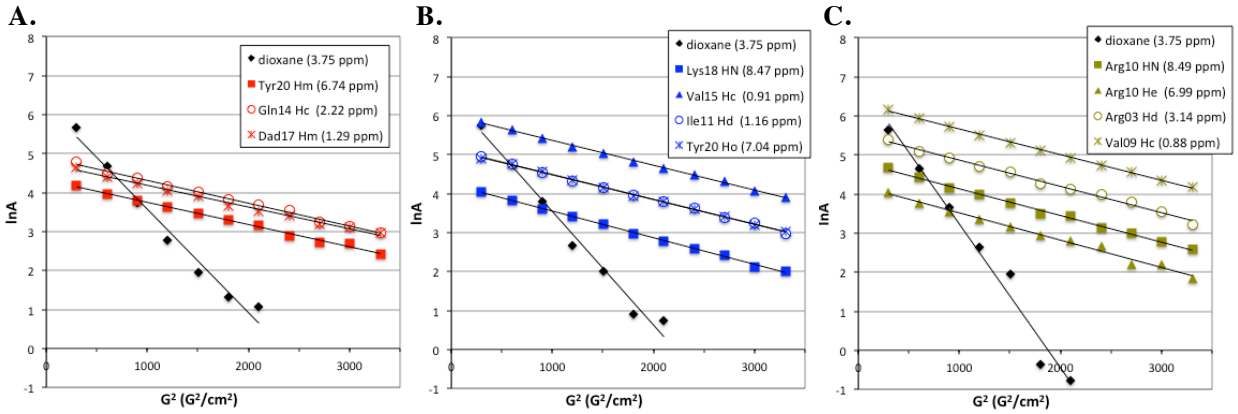

Figure 2.11. Plots of the natural log of ^{1}H NMR signal intensities (lnA) as a function of the square of the gradient strength (G^2), for peptides **2.A**, **2.B**, and **2.C**, at 2 mM concentration in buffer (9:1 H_2O:D_2O, pH 3.8, 100 mM d_3-acetate) at 4.0 °C. An internal standard of dioxane (0.6 mM) was employed. Linear regressions are shown, and R^2 value > 0.95, for every line.

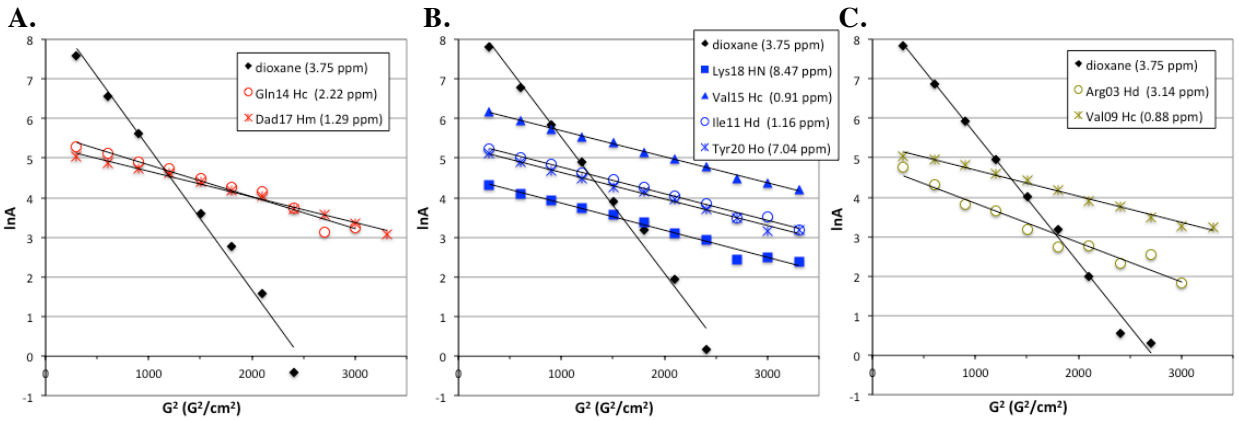

Figure 2.12. Plots of the natural log of ^{1}H NMR signal intensities (lnA) as a function of the square of the gradient strength (G^2), for peptides **2.A**, **2.B**, and **2.C**, at 0.2 mM concentration in buffer (9:1 H_2O:D_2O, pH 3.8, 100 mM d_3-acetate) at 4.0 °C. An internal standard of dioxane (0.6 mM) was employed. Linear regressions are shown, and R^2 value > 0.95, for every line.

Table 2.2. Hydrodynamic radii of peptides, derived from NMR diffusion experiments (obs R_H) or HYDROPRO 10 calculations for monomeric peptides (calc R_H).

peptide	**buffer**	**obs R_H** (Å) at 2 mM peptide	**obs R_H** (Å) at 0.2 mM peptide	**calc R_H** (Å)
2.A	100 mM NaOAc, pH 3.8	9.8 ± 0.6	10.5 ± 0.7	13.2
	3 mM NaOAc, pH 3.8	7.7 ± 0.4	8.9 ± 0.8	
2.B	100 mM NaOAc, pH 3.8	9.5 ± 0.5	10.8 ± 0.5	11.3
	3 mM NaOAc, pH 3.8	7.4 ± 0.3	8.7 ± 0.3	
2.C	100 mM NaOAc, pH 3.8	11.5 ± 0.7	8.3 ± 0.4	11.0
	3 mM NaOAc, pH 3.8	6.8 ± 0.3	9.1 ± 0.2	

For each peptide, no changes in chemical shifts, peak shapes, or splitting are observed, as peptide concentrations are increased from 0.02 mM to 2 mM (**Figures 2.13**, **2.14**, **2.15**). This supports the hypothesis that there is no aggregation of these peptides, in the 0.02 to 2 mM concentration range in aqueous solution.

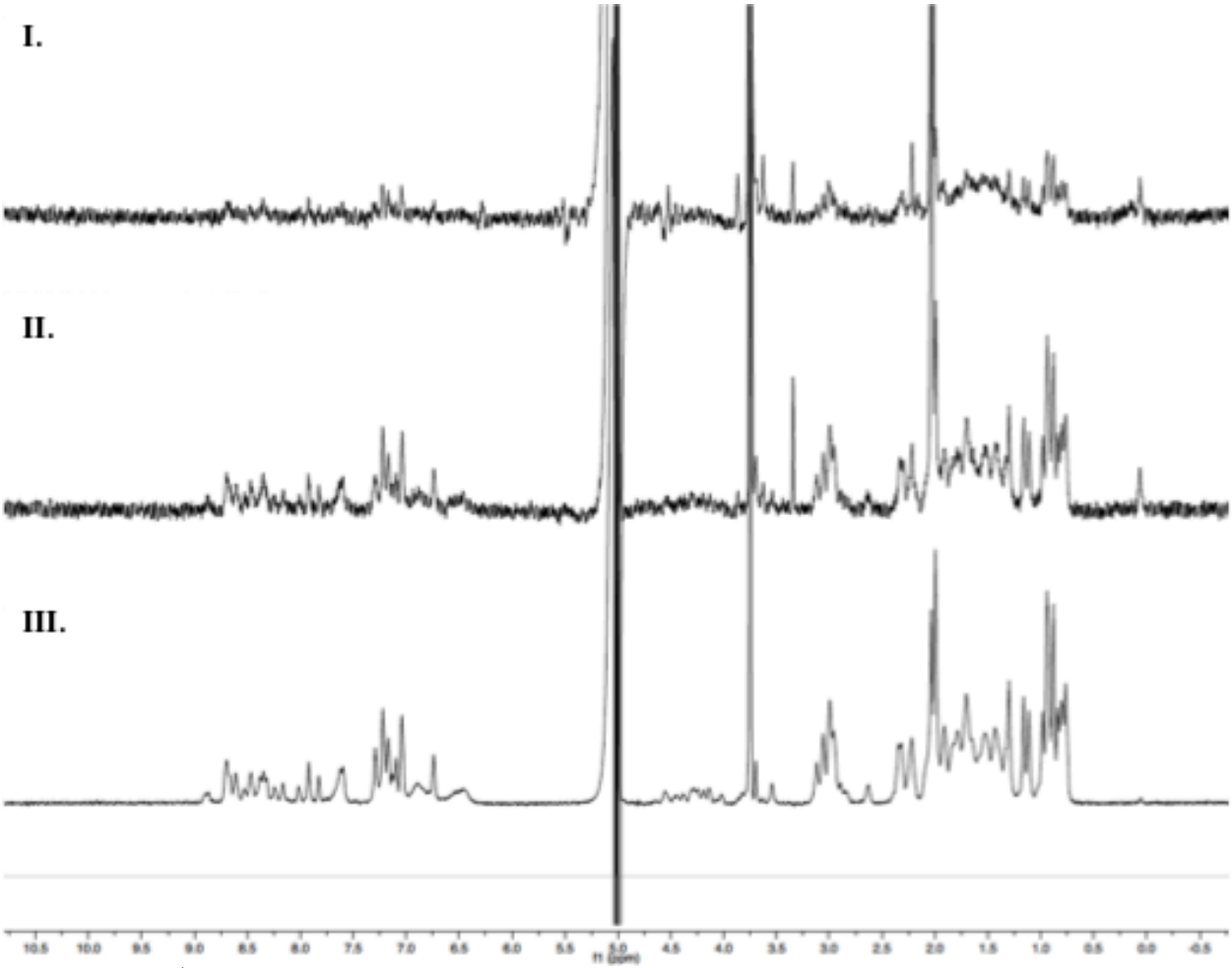

Figure 2.13. ^{1}H NMR spectra of peptide **2.A**, at (III) 2 mM, (II) 0.2 mM, and (I) 0.02 mM concentrations.

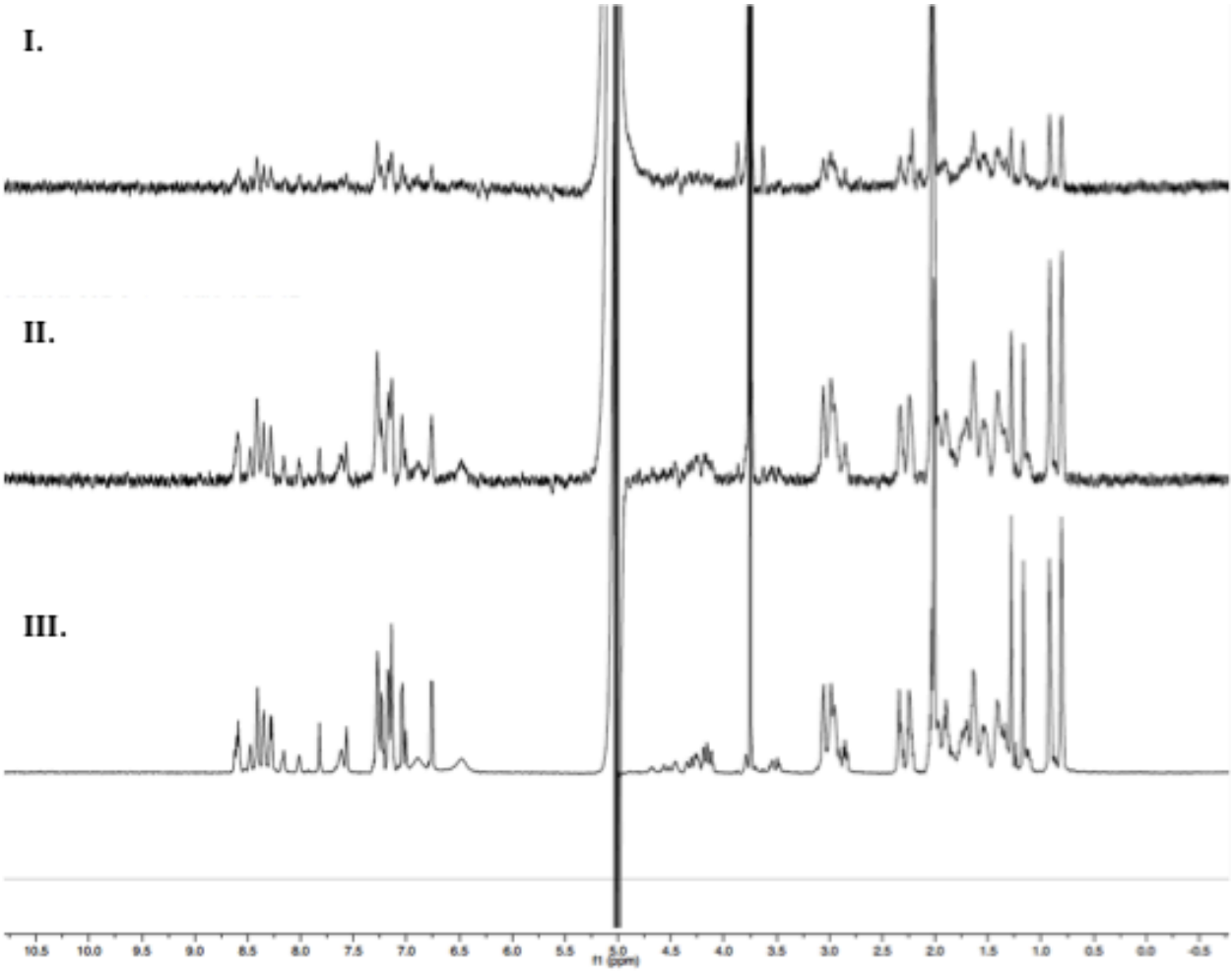

Figure 2.14. ^{1}H NMR spectra of peptide **2.B**, at (III) 2 mM, (II) 0.2 mM, and (I) 0.02 mM concentrations.

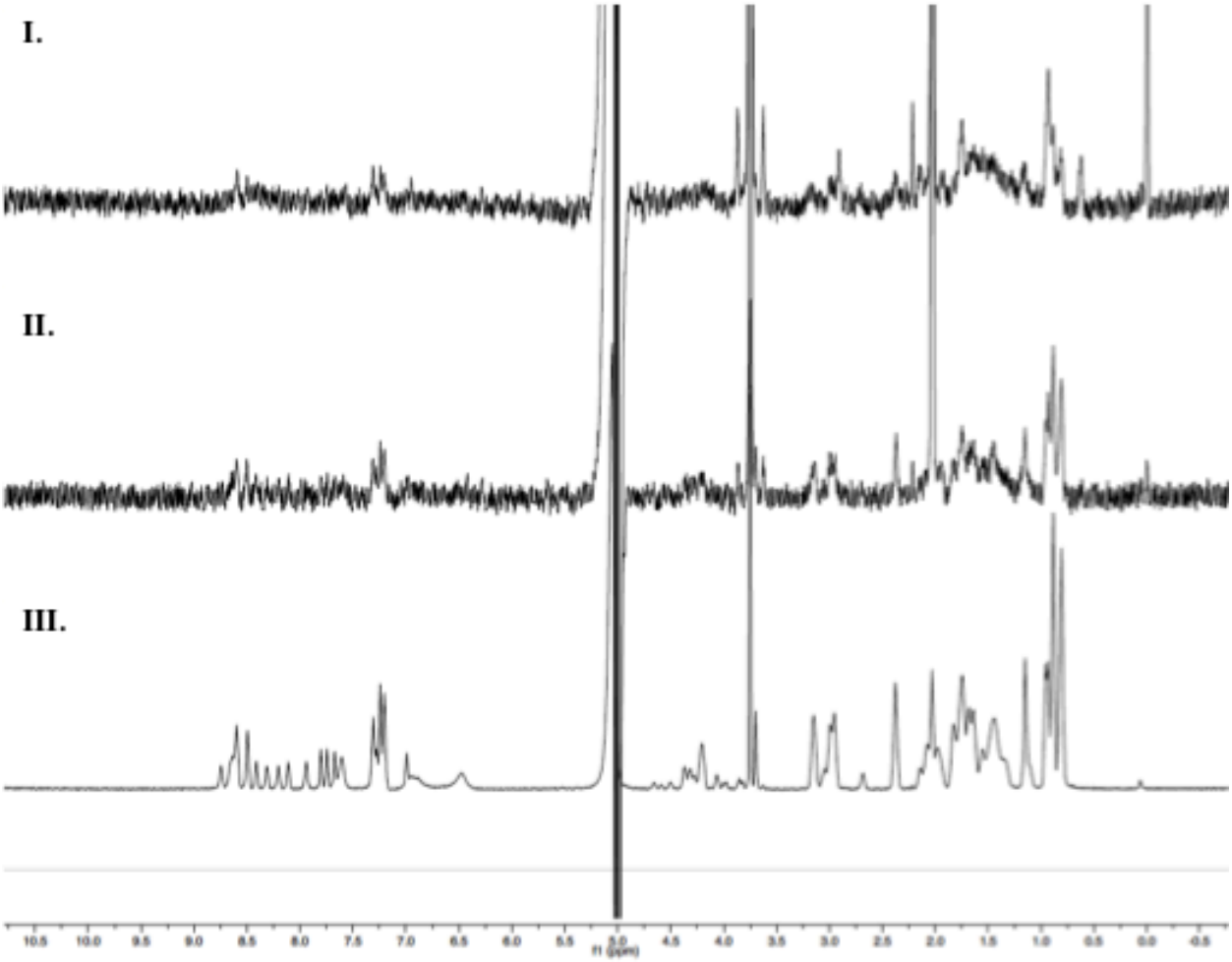

Figure 2.15. ^{1}H NMR spectra of peptide **2.C**, at (III) 2 mM, (II) 0.2 mM, and (I) 0.02 mM concentrations.

2.4.3. Structural characterization of model parallel β-sheet peptides

The extent of β-sheet formation was monitored by nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy.

Circular dichroism can be used to assess protein folding in solution, with the association of α-helix (208 nm, 222 nm) and antiparallel β-sheet (218 nm) secondary structures with characteristic minima in CD spectra.[54] The CD signature of parallel β-sheets is not as well characterized,[55,56] due to the difficulty of parallel β-sheet isolation. Here we report CD wavelength scans for isolated, monomeric parallel β-sheets in aqueous solution (**Figure 2.16** solid lines). Triple-strand design **2.A** displays a single minimum of modest intensity at 214 nm

(**Figure 2.16**), which is consistent with β-sheet secondary structure. Double-strand design **2.B** displays a weaker minimum in the same region, which suggests a smaller population of β-sheet. Double-strand design **2.C** has a negative intensity signal but no minimum in this region, a CD signature that might be consistent with a small population with β-sheet secondary structure and a population with random coil structure. Addition of 50% (v/v) 2,2,2-trifluoroethanol (TFE) co-solvent enhanced the minimum near 214 nm for all three peptides (**Figure 2.16** dashed lines), consistent with alcohol-mediated promotion of (β-sheet) secondary structure.[40] For purely aqueous conditions and with TFE co-solvent, comparison of the CD data for the three peptides suggests that the residue-normalized population of β-sheet structure is greater in three-stranded design **2.A** relative to either of the two-stranded designs, **2.B** or **2.C**. Thus, the CD data suggest that parallel β-sheet secondary structure becomes more favorable as the number of strands is increased. This presentation of CD wavelength scans (**Figures 2.16**) of parallel beta-sheets with defined parameters (e.g., strand number, strand twist (**Figure 2.21**), aromatic residue composition) also demonstrates how model systems such as ours may be employed to experimentally probe computational predictions of the CD spectra of parallel β-sheets.[56,57]

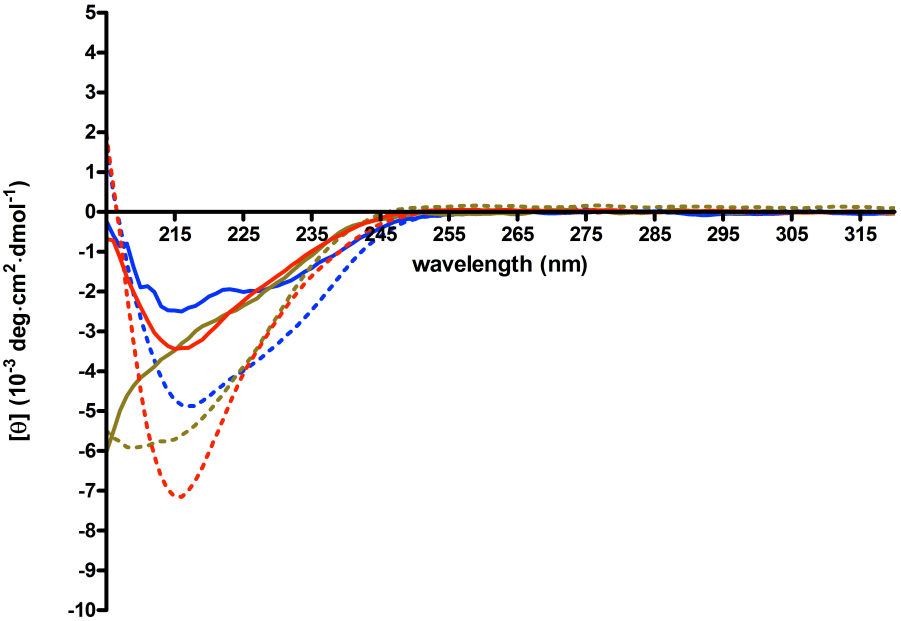

Figure 2.16. Mean residue ellipticity of peptides **2.A** (red), **2.B** (blue), and **2.C** (yellow) at 20.0 °C (293.2 K). Spectra are reported for 0.2 mM peptide samples in 10 mM NaOAc buffer, pH 3.8, without (solid lines) or with (dashed lines) addition of 50% (v/v) 2,2,2-trifluoroethanol.

Two-dimensional COSY, TOCSY, and ROESY spectra were obtained for each peptide at 2 mM concentration in aqueous solution (9:1 H_2O:D_2O, 3 mM acetate-d_3, pH 3.8) at 4 °C, and were sufficient for assigning almost all ^{1}H NMR signals (**Tables 2.3**). ROE peak intensities (**Figures 2.17**) were translated into a continuous distribution of interproton distance restraints, and these restraints were used to carry out simulated annealing of each peptide. The average of the 10 lowest energy (out of 100 calculated) structures for each peptide is shown in **Figure 2.21**. These NMR structures show a parallel β-sheet conformation for peptides **2.A** and **2.B**, and a less ordered fold or more strand fraying for peptide **2.C**.

Table 2.3A. Proton resonances (ppm) for peptide **2.A**.

		NH	αH	βH		γH		δH		εH		others	
23	**Lys**	8.66	4.31	1.81	1.81	1.47	1.47	1.69	1.69	3.00	3.00	7.62 (εNH_3^+)	
22	**Glu**	8.50	4.45	2.05	1.90	2.34	2.34						
21	**Arg**	8.69	4.60	1.77	1.70	1.58	1.52	3.12	3.12			7.24 (εNH)	7.04 (ηNH)
20	**Ile**	8.69	4.45	2.06		1.42	1.13	0.78				0.93 (γCH_3)	
19	**Thr**	8.67	4.74	4.02		1.12							
18	**Val**	8.82	4.23	1.83		0.87	0.87						
17	**Gly**	8.23	3.80, 4.05										
16	**CHDA**	3.79, 2.92 (αH), 2.23, 2.07, 1.98, 1.90, 1.76 (βH), 0.62											
15	**Val**	7.91	4.35	1.98		0.86	0.86						
14	**Arg**	8.48	4.54	1.74	1.68	1.56	1.56	3.06	3.06			7.16 (εNH)	
13	**Ile**	8.44	4.53			1.69	1.05	0.75					
12	**Phe**	8.67	4.74	2.89	2.89			7.27 (o)				7.81 (p)	
11	**Lys**	8.44	4.46	1.72	1.72	1.31	1.31	1.60	1.60	2.96	2.96	7.66 (εNH_3^+)	
10	**Gln**	8.61	4.60	1.98	1.98	2.21	2.21			6.47	6.90		
9	**Val**	8.15	4.19	2.17		0.96	0.96						
8	**Pro**			2.41	2.03			3.50	3.50				
7	**DADME**	8.03	3.53									1.28 (CH_3), 1.17 (CH_3)	
6	**Lys**	8.58	4.53	1.81	1.81			1.71	1.71	3.00	3.00	7.62 (εNH_3^+)	
5	**Glu**	8.69	4.56	1.93	1.93	2.39	2.39						
4	**Tyr**	8.36	4.55	2.97	2.84			7.06 (o)		6.74 (m)			
3	**Phe**	8.36	4.88	2.96	2.67			7.02 (o)				7.57 (p)	
2	**Arg**	8.33	4.37	1.53	1.53	1.41	1.36	3.05	3.05			7.17 (εNH)	
1	**Glu**	8.36	4.15	2.26	1.91								
	Ac		2.36										

Table 2.3B. Proton resonances (ppm) for peptide **2.B**.

		NH	αH	βH		γH		δH		εH		others	
	Ac												
14	**Arg**	8.35	4.20	1.64	1.53	1.41	1.41	3.06	3.06			7.14 (εNH)	
13	**Ile**	8.28	4.16	1.75		1.42	0.82	1.14				0.63 (γCH_3)	
12	**Phe**	8.60	4.68	2.99	2.92			7.17 (o)		7.30 (m)		7.27 (p)	
11	**Lys**	8.35	4.32	1.73	1.73	1.37	1.37	1.66	1.66	2.95	2.95	7.61 (NH_3^+)	
10	**Gln**	8.61	4.48	1.98	1.98	2.26	2.26						
9	**Val**	8.63	4.53	2.03		0.93	0.93						
8	**Pro**		4.28	2.24	1.92	0.93	0.93	3.80	3.80				
7	**DADME**	8.02	3.54, 3.55									1.28 (CH_3), 1.18 (CH_3)	
6	**Lys**	8.49	4.45	1.83	1.81	1.52	1.44	1.71	1.71	2.99	2.99	7.62 (NH_3^+)	
5	**Glu**	8.41	4.37	2.05	2.05	2.37	2.37						
4	**Tyr**	8.19	4.72	2.97	2.86			7.04 (o)		6.76 (m)			
3	**Phe**	8.29	4.57	3.01	2.85			7.00 (o)		7.55 (m)		7.28 (p)	
2	**Arg**	8.40	4.26	1.62	1.57	1.39	1.39	3.06	3.06			7.14 (εNH)	7.82 (ηNH)
1	**Glu**	8.41	4.13	1.91	1.91	2.37	2.28						
	Ac												

Table 2.3C. Proton resonances (ppm) for peptide **2.C**.

		NH	αH	βH		γH		δH		εH		others	
	NH2												
23	Lys	8.62	4.24	1.81	1.81	1.45	1.45	1.69	1.69	2.98	2.98	7.63 (εNH_3^+)	
22	Glu	8.60	4.38	1.96	1.96	2.40	2.10						
21	Arg	8.64	4.50	1.80	1.73	1.61	1.57	3.17	3.17			7.23 (εNH)	
20	Ile	8.67	4.21	1.83		1.47	1.43	1.16				0.89 (γCH_3)	
19	Thr	8.60	4.58	4.07		1.16							
18	Val	8.11	4.19	2.14		0.95	0.95						
17	Gly	8.21	3.85, 3.99										
16	CHDA	3.01 (αH), 2.70 (αH), 2.03 (βH), 1.84, 1.75 (βH), 1.51, 1.41, 0.62											
15	Val	7.95	4.27	2.04		0.89	0.89						
14	Arg	8.50	4.38	1.73	1.73	1.56	1.48	3.14	3.14			7.19 (εNH)	7.31 (ηNH)
13	Ile	8.32	4.31	1.76		1.43	0.81	1.12					
12	Phe	8.75	4.66	3.05	2.96			7.24 (o)		7.80 (m)		7.74 (p)	
11	Lys	8.42	4.32	1.75	1.75	1.35	1.35	1.66	1.66	2.95	2.95	7.60 (εNH_3^+)	
10	Gln	8.50	4.22	1.98	1.98	2.37	2.08			6.99	7.67		
	Ac												

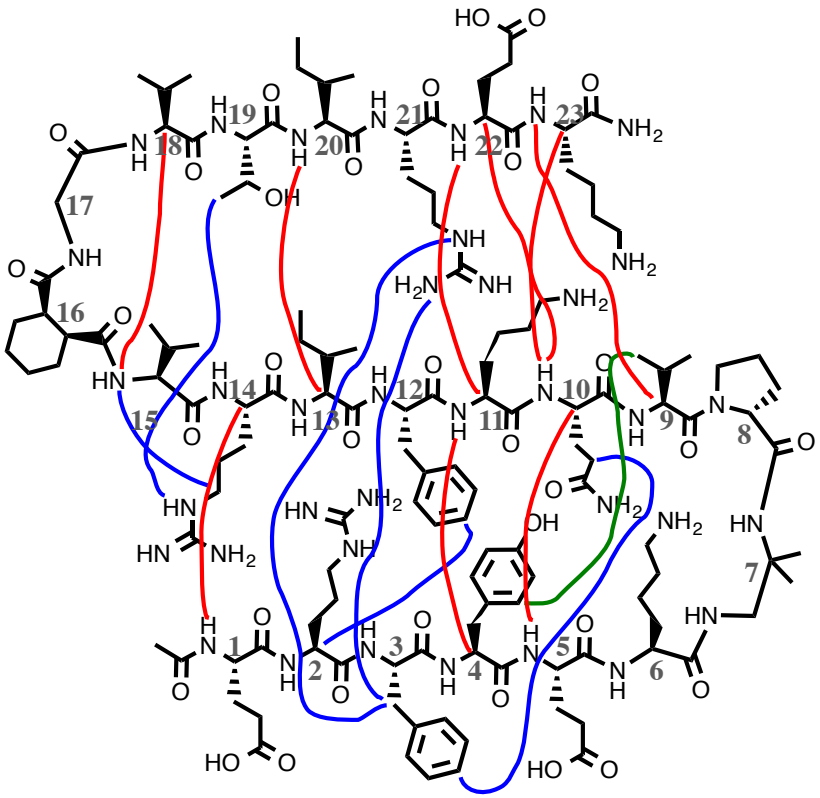

Figure 2.17A. ROE's observed in peptide **2.A**, in backbone (red), between sidechains in "top face" (green), and between sidechains in "bottom face" (blue) of peptide.

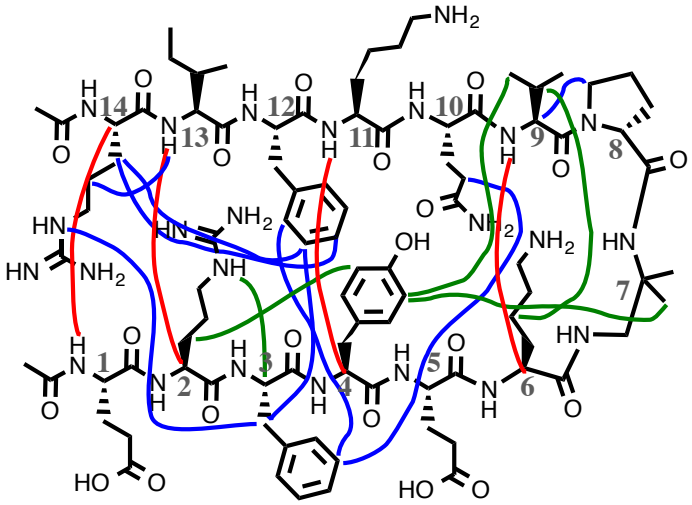

Figure 2.17B. ROE's observed in peptide **2.B**, in backbone (red), between sidechains in "top face" (green), and between sidechains in "bottom face" (blue) of peptide.

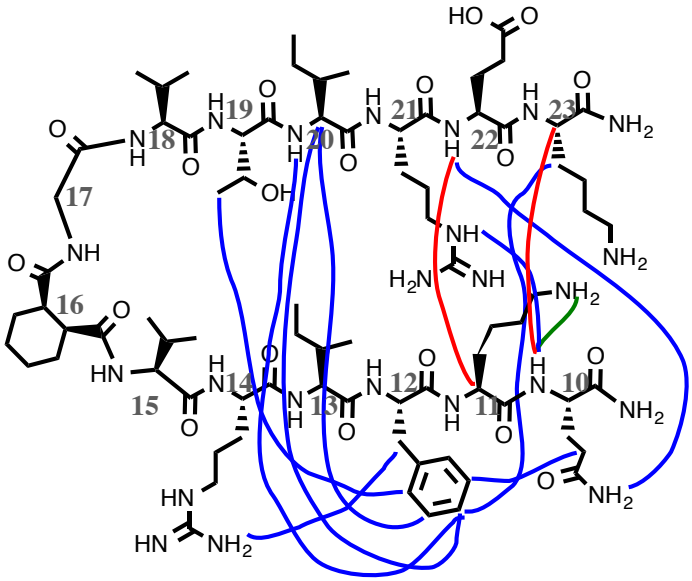

Figure 2.17C. ROE's observed in peptide **2.C**, in backbone (red), between sidechains in "top face" (green), and between sidechains in "bottom face" (blue) of peptide.

Table 2.4A. Structural statistics for peptide **2.A**, 10 (of 100 calculated) structures with the overall lowest energies. Reported statistics are for all (not just ordered) residues.

Conformationally restricting restraints	Experimental ^{1}H-^{1}H distance restraints	
	Intraresidue (i = j)	129
	Sequential (|i –j| = 1)	13
	Medium range (1 < |i –j| < 5)	1
	Long range |i –j| ≥ 5	14
	Hydrogen-bond restraints	10
	Dihedral angle restraints	12
Residual restraint violations	Ave no. of ROE restraint violations per structure (>0.5 Å)	1.2
	Ave no. of H-bond violations per structure (>0.5 Å)	0.1
	Ave no. of dihedral angle violations per structure (>10º)	0.1
Model quality	RMSD backbone atoms (Å)	1.4
	RMSD heavy atoms (Å)	2.4
	RMSD bond lengths (Å)	0.014
	RMSD bond angles (º)	1.4
MolProbity Ramachandran statistics	Most favored regions (%)	84.7
	Allowed regions (%)	12.0
	Disallowed regions (%)	3.3
Global quality scores (mean/Z score)	Verify3D	-0.02/ -7.70
	ProsaII	-1.09/ -7.20
	PROCHECK (**ϕ-ψ**)	-1.43/ -5.31
	PROCHECK (all)	-0.85/ -5.03
	MolProbity clash score	87.73/-13.53
Model contents	BMRB accession number	25673
	PDB ID code	2N4N

Table 2.4B. Structural statistics for peptide **2.B**, 10 (of 100 calculated) structures with the overall lowest energies. Reported statistics are for all (not just ordered) residues.

Conformationally restricting restraints	Experimental ^{1}H-^{1}H distance restraints	
	Intraresidue (i = j)	120
	Sequential (|i –j| = 1)	11
	Medium range (1 < |i –j| < 5)	6
	Long range |i –j| ≥ 5	9
	Hydrogen-bond restraints	6
	Dihedral angle restraints	4
Residual restraint violations	Ave no. of ROE restraint violations per structure (>0.5 Å)	2.2
	Ave no. of H-bond violations per structure (>0.5 Å)	0.2
	Ave no. of dihedral angle violations per structure (>10º)	0
Model quality	RMSD backbone atoms (Å)	1.3
	RMSD heavy atoms (Å)	2.4
	RMSD bond lengths (Å)	0.014
	RMSD bond angles (º)	1.6
MolProbity Ramachandran statistics	Most favored regions (%)	63.3
	Allowed regions (%)	23.3
	Disallowed regions (%)	13.3
Global quality scores (mean/Z score)	Verify3D	-0.04/ -8.03
	Prosall	-1.80/-10.13
	PROCHECK (**ϕ-ψ**)	-2.12/ -8.03
	PROCHECK (all)	-1.02/ -6.03
	MolProbity clash score	78.40/-11.93
Model contents	BMRB accession number	25764
	PDB ID code	2N6H

Table 2.4C. Structural statistics for peptide **2.C**, 10 (of 100 calculated) structures with the overall lowest energies. Reported statistics are for all (not just ordered) residues.

Conformationally restricting restraints	Experimental ^{1}H-^{1}H distance restraints	
	Intraresidue (i = j)	117
	Sequential (|i –j| = 1)	11
	Medium range (1 < |i –j| < 5)	1
	Long range |i –j| ≥ 5	9
	Hydrogen-bond restraints	3
	Dihedral angle restraints	4
Residual restraint violations	Ave no. of ROE restraint violations per structure (>0.5 Å)	2.5
	Ave no. of H-bond violations per structure (>0.5 Å)	0
	Ave no. of dihedral angle violations per structure (>10°)	0
Model quality	RMSD backbone atoms (Å)	1.8
	RMSD heavy atoms (Å)	2.4
	RMSD bond lengths (Å)	0.012
	RMSD bond angles (°)	1.5
MolProbity Ramachandran statistics	Most favored regions (%)	66.7
	Allowed regions (%)	21.1
	Disallowed regions (%)	12.2
Global quality scores (mean/Z score)	Verify3D	-0.15/ -9.79
	Prosall	-1.04/ -6.99
	PROCHECK (ϕ-ψ)	-1.78/ -6.69
	PROCHECK (all)	-1.05/ -6.21
	MolProbity clash score	123.83/-19.72
Model contents	BMRB accession number	25765
	PDB ID code	2N6I

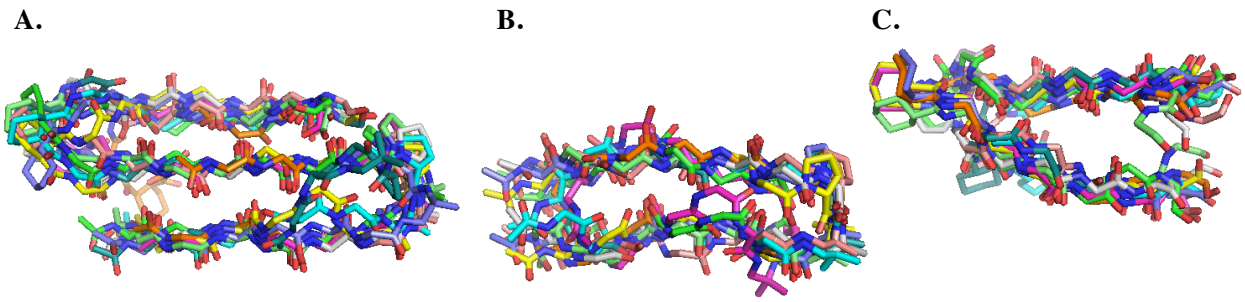

Figure 2.18. Structure ensembles for 10 lowest energy (out of 100 calculated) structures, for peptides **2.A**, **2.B**, and **2.C**. PDB ID codes: 2N4N, 2N6H, 2N6I.

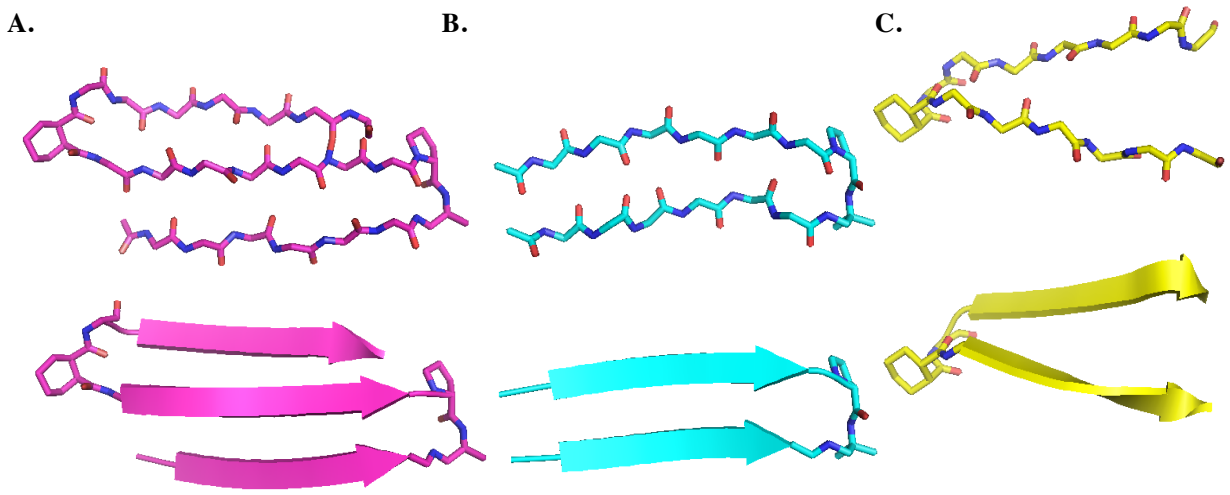

Figure 2.19. Average of 10 lowest energy (out of 100 calculated) structures, for peptides **2.A** (magenta), **2.B** (cyan), and **2.C** (yellow). In the second row, residues 1-7, 15-9, and 18-23 of peptides are depicted using cartoon ribbon diagrams, representing backbone atom positions but not necessarily canonical β-strand structure. PDB ID codes: 2N4N, 2N6H, 2N6I.

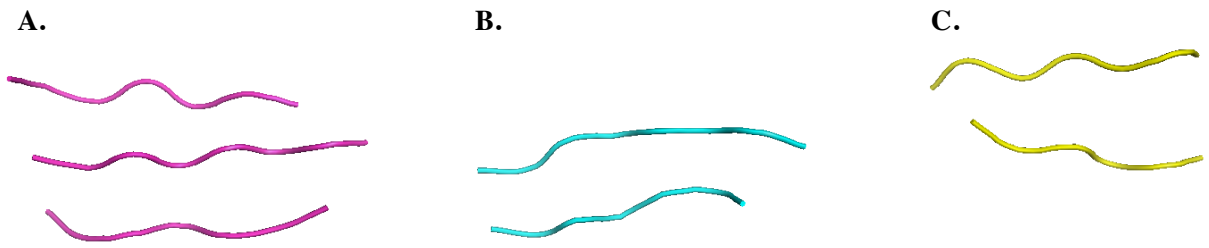

Figure 2.20. Average of 10 lowest energy (out of 100 calculated) structures, for peptides **2.A** (magenta), **2.B** (cyan), and **2.C** (yellow), using so-called "tube" depictions of backbone atom positions. Non-natural residues DADME and CHDA are not shown.

Figure 2.21. Superposition of aqueous solution state structures of peptides **2.A** (magenta), **2.B** (cyan), and **2.C** (yellow). Each structure represents the average of the 10 lowest energy (out of 100) structures calculated from NMR ROE restraints. To facilitate visual comparison, residues 1-7, 15-9, and 18-23 of all peptides are depicted using cartoon ribbon diagrams, representing backbone atom positions but not necessarily canonical β-strand structure. PDB ID codes: 2N4N, 2N6H, 2N6I.

2.4.4. Impact of strand number on parallel β-sheet stability

Chemical shifts of α-protons ($\delta_{\alpha H}$) are sensitive to secondary structure, with residues in β-sheet or extended conformations generally occurring downfield of those in random coils.[29] A gross pattern of downfield α-proton secondary chemical shifts was observed for peptides **2.A**, **2.B**, and **2.C** (**Figure 2.22**). Observation of $\Delta\delta_{\alpha H}$ ($= \delta_{\alpha H,observed} - \delta_{\alpha H,random\ coil}$) > 0.1 ppm for three consecutive residues can be evidence of β-strand structure at these positions.[58] This criterion offers further evidence that β-strands 2 (residues 10-15) and 3 (residues 19-21) do form in peptide **2.A**. The residues of strand 1 in peptide **2.A** also have downfield $\Delta\delta_{\alpha H}$ values, but only a weakly downfield $\Delta\delta_{\alpha H}$ value for residue 4; it is possible that a nearby aromatic ring shields the α-proton of Tyr4. For two-stranded peptides **2.B** and **2.C**, $\Delta\delta_{\alpha H}$ at nearly every strand residue is less than $\Delta\delta_{\alpha H}$ for the corresponding residue in **2.A**. This trend suggests that the β-sheet population is higher for peptide **2.A** than for two-stranded peptides **2.B** or **2.C**. In other words, the presence of strand 1 increases the extent to which strand 3 in **2.A** populates the β-sheet state relative to strand 3 in **2.C**, and the presence of strand 3 increases the extent to which

strand 1 in **2.A** populates the β-sheet state relative to strand 1 in **2.B**. Thus, the ^{1}H secondary chemical shift data imply that the extent of β-sheet formation becomes larger as strand number is increased from two to three.

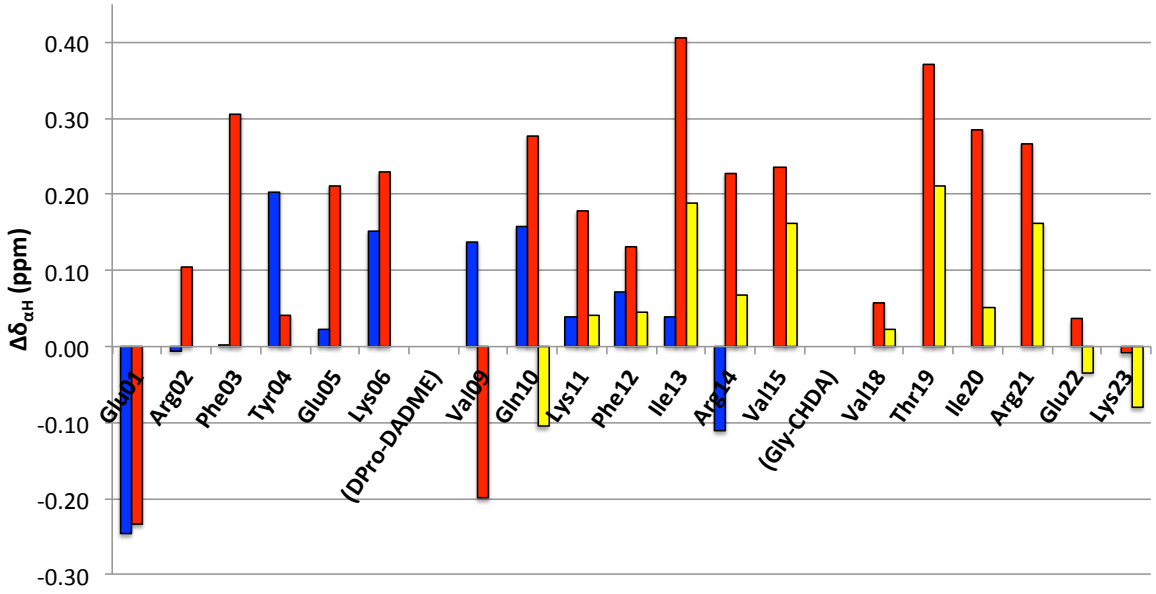

Figure 2.22. Secondary α-proton chemical shifts of peptides **2.A** (red), **2.B** (blue), and **2.C** (yellow), at 2 mM peptide concentrations in aqueous solution (9:1 H_2O:D_2O, 3 mM acetate-d_3, pH 3.8) at 4 °C. Random coil values were calculated with the CSDb[2] program.

In purely aqueous conditions (10 mM NaOAc, pH 3.8), a sigmoidal decrease in the CD signal at ~214 nm as a function of temperature was observed for peptide **2.A**, but not for either **2.B** or **2.C** (**Figure 2.23**). For all three peptides, CD signal intensities were weak. The two-stranded peptides, **2.B** and **2.C**, had particularly weak or no local minima at ~214 nm (**Figure 2.22**), so though this minimum broadened as the temperature increased, there is an apparent increase in the intensity of the CD signal at ~214 nm as a function of temperature, due to an increase in lower wavelength signals, likely consistent with random coil CD signatures.

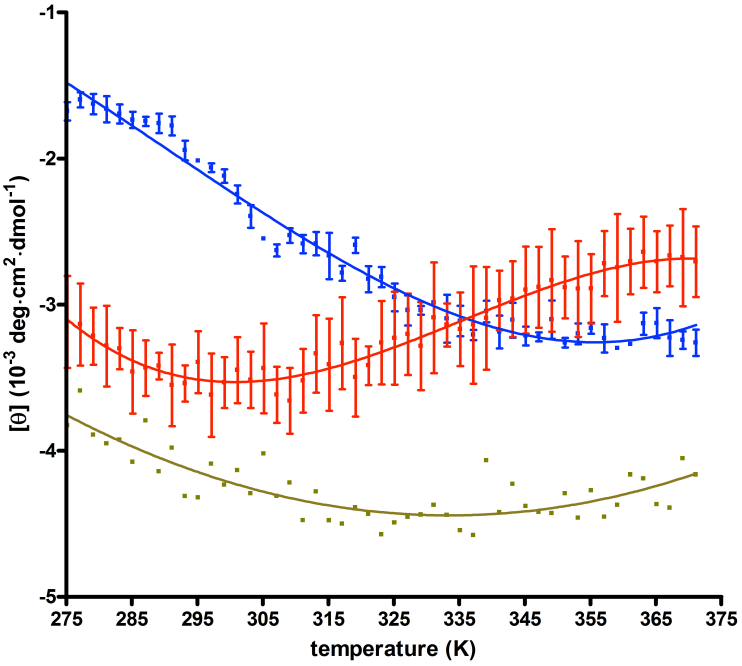

Figure 2.23. Mean residue ellipticity of peptides **2.A** (red), **2.B** (blue), and **2.C** (yellow) as a function of temperature. Solutions consisted of 0.2 mM peptide in aqueous buffer (10 mM NaOAc, pH 3.8). Points represent mean values (n = 2 for **2.A**, n = 3 for **2.B**, n = 1 for **2.C**), with accompanying standard error of the mean. Lines represents best fits of the data to polynomial models, which are intended to facilitate visualization of the data, and have no physical meaning.

Thermal denaturation experiments were also performed in the presence of a secondary structure-promoting co-solvent, TFE. For each peptide, the CD signal at ~214 nm was monitored as a function of temperature. Thermal melt curves for samples with TFE are reported in **Figure 2.24**. A sigmoidal transition was observed for the three-stranded peptide, which is evidence for cooperative folding. In contrast, more parabolic transitions were observed for both of the two-stranded peptides, with no well-defined baseline regions (in the temperature range examined) corresponding to fully folded or unfolded states.

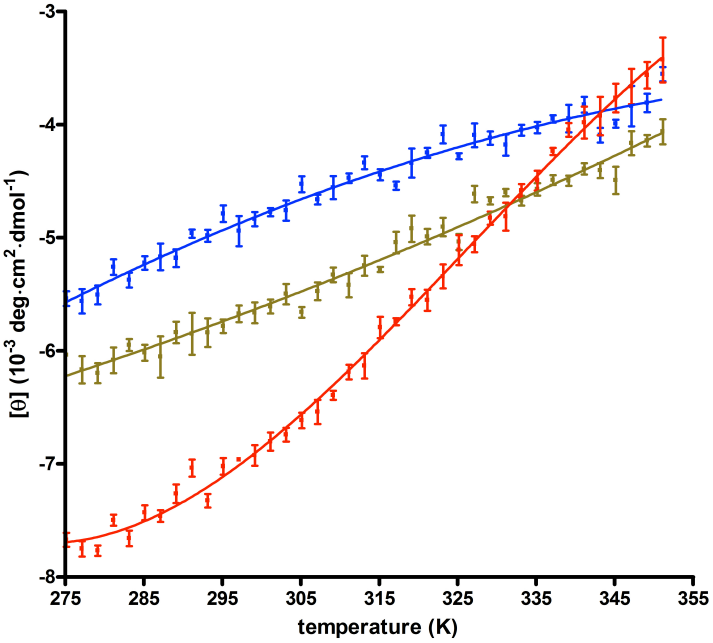

Figure 2.24. Mean residue ellipticity of peptides **2.A** (red), **2.B** (blue), and **2.C** (yellow) as a function of temperature. Solutions consisted of 0.2 mM peptide in 50% (v/v) aqueous buffer (10 mM NaOAc, pH 3.8) and 50% (v/v) 2,2,2-trifluoroethanol. Points represent mean values (n=3), with accompanying standard error of the mean. Lines represent best fits of the data to polynomial models, which are intended to facilitate visualization of the data, and have no physical meaning.

2.5. Conclusion

We have developed peptides that fold autonomously in aqueous solution into two- or three-stranded parallel β-sheets. There is no precedent for a three-stranded parallel β-sheet that folds autonomously. Our peptide models can be used to answer fundamental questions about β-sheets, which are otherwise difficult to study, because they tend to form insoluble aggregates when extracted from proteins, yet isolated study is desired to remove potential confounding effects of tertiary structure contacts. We have determined by AUC, NMR diffusion, and variable concentration NMR experiments that the three peptides developed here are monomeric, and do not aggregate at the aqueous solution NMR conditions used to study the peptide folding.

Protein folding is governed by a complex network of factors.[59] One key factor that has not been previously explored experimentally is the relationship between parallel β-sheet strand

number and conformational stability. In nature, parallel structure is rarely found in β-sheets of less than five strands, whereas antiparallel β-sheets of just two strands are common.[60] This observation may suggest a particularly robust correlation between strand number and stability for parallel β-sheets, or an energetically favorable propagation of parallel β-sheet structure beyond two strands. The propagation of parallel β-sheet structures is of great interest, as it has been implicated in the protein misfolding that is found in amyloid diseases.[61-63] In our model system, increasing the number of strands from two to three stabilizes the parallel β-sheet. Of note in our model system, in the absence of strand 1 (i.e., in peptide **2.C**), the two strands connected by the diacid linker have little parallel β-sheet structure. The NMR and CD results described here suggest that parallel β-sheet structure is stabilized with increasing strand number, but do not rule out the possibility that length-dependent stabilization of β-sheets perpendicular to the strand direction dissipates with increasing length. We also observe evidence for cooperative folding of the three-stranded peptide, but not for either of its two-stranded counterparts. The cooperativity of β-sheet (antiparallel or parallel) growth in the perpendicular direction has been predicted by computational studies, and attributed to electrostatic attractions among amide dipoles.[64] This cooperativity was reported for antiparallel β-sheets.[22,23,65,66] Here we present the first experimental evidence of cooperative parallel β-sheet growth in the perpendicular direction. The cooperative effect may be explained by a decrease in the entropic cost of the formation of the second parallel hairpin, once the first parallel hairpin has been formed (i.e., once the conformation of strand 2 has been preorganized). Here we also present the first experimental quantification of the contribution of parallel β-sheet secondary structure to CD spectra. Further characterizations could be made of the CD signatures of these peptides at lower wavelengths.

2.6. References

(1) Geddes, A. J.; Parker, K. D.; Atkins, E. D.; Beighton, E. *J Mol Biol* **1968**, *32*, 343.

(2) Nelson, R.; Eisenberg, D. *Curr Opin Struc Biol* **2006**, *16*, 260.

(3) Serpell, L. C. *BBA Mol Basis Dis* **2000**, *1502*, 16.

(4) Chakrabartty, A.; Baldwin, R. L. *Adv Protein Chem* **1995**, *46*, 141.

(5) Baldwin, R. L.; Rose, G. D. *Trends Biochem Sci* **1999**, *24*, 26.

(6) Nesloney, C. L.; Kelly, J. W. *Bioorgan Med Chem* **1996**, *4*, 739.

(7) Gellman, S. H. *Curr Opin Chem Biol* **1998**, *2*, 717.

(8) Searle, M. S.; Ciani, B. *Curr Opin Struc Biol* **2004**, *14*, 458.

(9) Hughes, R. M.; Waters, M. L. *Curr Opin Struc Biol* **2006**, *16*, 514.

(10) Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. *Proc Natl Acad Sci USA* **2010**, *107*, 10466.

(11) Chitnumsub, P.; Fiori, W. R.; Lashuel, H. A.; Diaz, H.; Kelly, J. W. *Bioorgan Med Chem* **1999**, *7*, 39.

(12) Nowick, J. S. *Acc Chem Res* **2008**, *41*, 1319.

(13) Fisk, J. D.; Gellman, S. H. *J Am Chem Soc* **2001**, *123*, 343.

(14) Freire, F.; Fisk, J. D.; Peoples, A. J.; Ivancic, M.; Guzei, I. A.; Gellman, S. H. *J Am Chem Soc* **2008**, *130*, 7839.

(15) Fisk, J. D.; Powell, D. R.; Gellman, S. H. *J Am Chem Soc* **2000**, *122*, 5443.

(16) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463.

(17) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1992**, *31*, 1263.

(18) Zimm, B. H.; Doty, P.; Iso, K. *Proc Natl Acad Sci USA* **1959**, *45*, 1601.

(19) Gibbs, A. C.; Kondejewski, L. H.; Gronwald, W.; Nip, A. M.; Hodges, R. S.; Sykes, B. D.; Wishart, D. S. *Nat Struct Biol* **1998**, *5*, 284.

(20) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriatt, I.; Muir, T.; Gellman, S. H. *Proc Natl Acad Sci USA* **2001**, *98*, 12015.

(21) Hong, L. *J Chem Phys* **2008**, *129*, 225101.

(22) Schenck, H. L.; Gellman, S. H. *J Am Chem Soc* **1998**, *120*, 4869.

(23) Sharman, G. J.; Searle, M. S. *J Am Chem Soc* **1998**, *120*, 5291.

(24) Sun, J. K.; Doig, A. J. *J Phys Chem B* **2000**, *104*, 1826.

(25) Freire, F.; Almeida, A. M.; Fisk, J. D.; Steinkruger, J. D.; Gellman, S. H. *Angew Chem Int Ed* **2011**, *50*, 8735.

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

(27) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(28) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J Am Chem Soc* **2004**, *126*, 7238.

(29) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J Mol Biol* **1991**, *222*, 311.

(30) Wishart, D. S. *Prog Nucl Mag Res Sp* **2011**, *58*, 62.

(31) Sharman, G. J.; Griffiths-Jones, S. R.; Jourdan, M.; Searle, M. S. *J Am Chem Soc* **2001**, *123*, 12318.

(32) Mielke, S. P.; Krishnan, V. V. *Prog Nucl Mag Res Sp* **2009**, *54*, 141.

(33) Hollingsworth, S. A.; Karplus, P. A. *Biomol Concepts* **2010**, *1*, 271.

(34) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J Magn Reson* **2003**, *160*, 65.

(35) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett* **1988**, *229*, 317.

(36) Clore, G. M.; Kuszewski, J. *J Am Chem Soc* **2002**, *124*, 2866.

(37) Schwieters, C. D.; Clore, G. M. *J Phys Chem B* **2008**, *112*, 6070.

(38) Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr D* **2004**, *60*, 1355.

(39) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins* **2007**, *66*, 778.

(40) Venkatraman, J.; Shankaramma, S. C.; Balaram, P. *Chem Rev* **2001**, *101*, 3131.

(41) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.

(42) Laue, T. M. *Methods Enzymol* **1995**, *259*, 427.

(43) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; The Royal Society of Chemistry: Cambridge, U.K., 1992, p 90.

(44) Perkins, S. J. *Eur J Biochem* **1986**, *157*, 169.

(45) Durchschlag, H.; Zipper, P. *Prog Colloid Polym Sci* **1994**, *94*, 20.

(46) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. *J Am Chem Soc* **1995**, *117*, 7566.

(47) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424.

(48) Pelta, M. D.; Morris, G. A.; Stchedroff, M. J.; Hammond, S. J. *Magn Reson Chem* **2002**, *40*, S147.

(49) Munoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301.

(50) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat Struct Biol* **1995**, *2*, 596.

(51) Fooks, H. M.; Martin, A. C. R.; Woolfson, D. N.; Sessions, R. B.; Hutchinson, E. G. *J Mol Biol* **2006**, *356*, 32.

(52) Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: New York, 2000.

(53) Ortega, A.; Amoros, D.; de la Torre, J. G. *Biophys J* **2011**, *101*, 892.

(54) Greenfield, N. J. *Nat Protoc* **2006**, *1*, 2876.

(55) Sieber, V.; Jurnak, F.; Moe, G. R. *Proteins* **1995**, *23*, 32.

(56) Woody, R. W. *Tetrahedron-Asymmetr* **1993**, *4*, 529.

(57) Micsonaia, A.; Wienb, F.; Kernyaa, L.; Leec, Y.-H.; Gotoc, Y.; Réfrégiersb, M.; Kardos, J. *Proc Natl Acad Sci USA* **2015**, *112*, E3095.

(58) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.

(59) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.

(60) Richardson, J. S. *Nature* **1977**, *268*, 495.

(61) Tycko, R.; Wickner, R. B. *Acc Chem Res* **2013**, *46*, 1487.

(62) Hamley, I. W. *Chem Rev* **2012**, *112*, 5147.

(63) Toyama, B. H.; Weissman, J. S. *Annu Rev Biochem* **2011**, *80*, 557.

(64) Zhao, Y. L.; Wu, Y. D. *J Am Chem Soc* **2002**, *124*, 1570.

(65) Roe, D. R.; Hornak, V.; Simmerling, C. *J Mol Biol* **2005**, *352*, 370.

(66) Hudson, F. M.; Andersen, N. H. *Biopolymers* **2006**, *83*, 424.

Chapter 3. A Designed Peptide Model of Early-Stage Amyloid Structures

This chapter is in preparation for publication by Kung VM, Bartlett GJ, Woolfson DN, and Gellman SH.

3.1. Abstract

Diseases involving amyloid proteins (e.g., Alzheimer's disease) debilitate tens of millions of people worldwide, yet there are no medical tools available for early diagnosis, prevention, or treatment of amyloid formation. This is partly because amyloid structures and mechanisms of formation remain enigmatic. Here we report the *de novo* design of a peptide containing both parallel β-sheet structure and β-arch structure in its monomeric form. This peptide monomer models the minimal folding unit of an amyloid polymorph thought to be medically relevant, and is stable in aqueous solution for days. Monomeric and oligomeric structures of the peptide were studied by solution state NMR, circular dichroism (CD), transmission electron microscopy (TEM), and sedimentation equilibrium analytical ultracentrifugation (AUC). Experiments employing our novel model system support the hypothesis that amyloid formation is driven by the hydrophobic effect. Model systems such as ours may help provide insights to amyloid disease pathogeneses.

3.2. Introduction

Despite the extremely high disease burden of conditions involving amyloid proteins (e.g., Alzheimer's disease, diabetes mellitus type 2), there currently are no noninvasive, definitive diagnostic methods or curative treatments for amyloid formation.[1,2] A key obstacle in the design of diagnostic and therapeutic tools for amyloid diseases has been limitations in our understanding of amyloid structure and mechanisms of formation. Peptide models of protein structures, and other forms of *de novo* protein design, are fundamental to the study of protein folding, structure, and function, and may lead to the development of biomedical tools.[3-5] The design of β-sheet structures, especially parallel ones, and of tertiary structures, are major challenges of *de novo* protein design.[6-8] Here we

report the first designed peptide that in aqueous solution has both parallel β-sheet and a specific tertiary structure, making it a promising model of amyloid polymorphs of medical relevance.

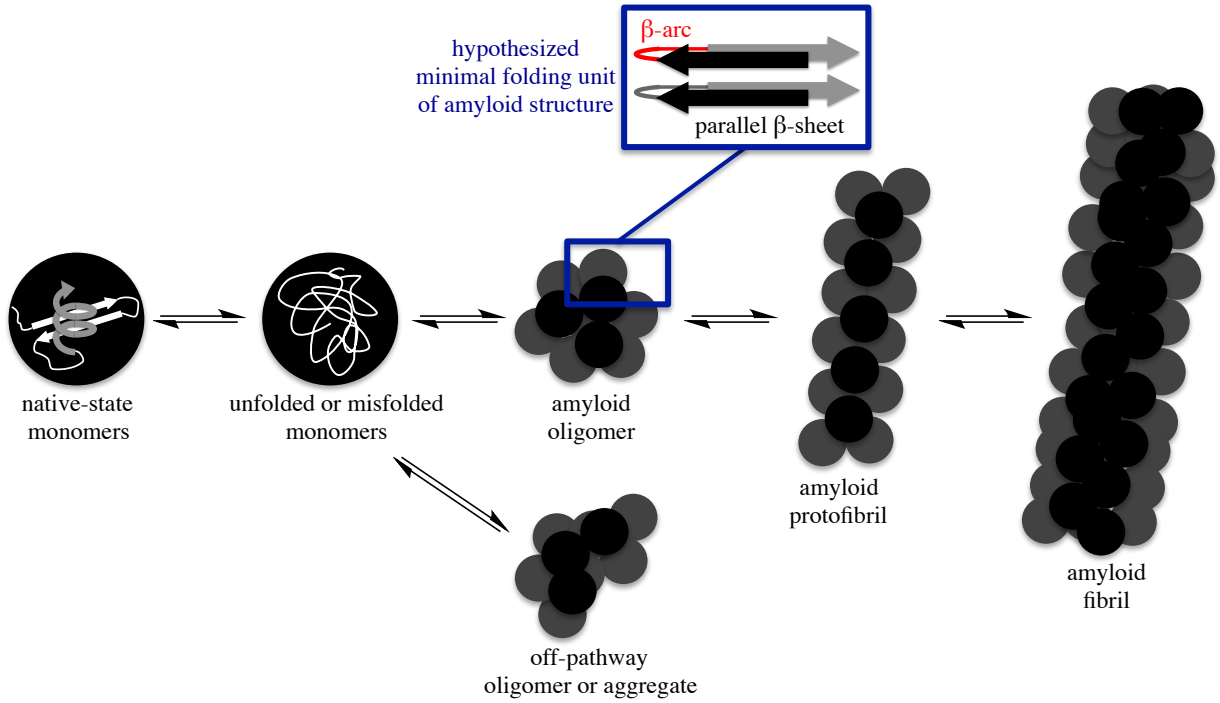

Figure 3.1. A simplified diagram of the mechanism of amyloid formation, depicting the spherical oligomer, protofibril, and fibril stages of amyloid growth. A hypothesized minimal folding unit of amyloid structure is shown in the blue box, with hydrogen-bonding between β-strands in a parallel orientation, and hydrophobic interactions between side-chains of β-strands linked by a β-arc (red).

Current understandings of amyloid structure and formation have been detailed in numerous reviews, and **Figure 3.1** is a simplified diagram of several key theories.[9-12] In brief, pathological amyloid proteins can misfold, aggregate, and deposit in tissues as assemblies called amyloid fibrils. Intact, natural fibrils are non-crystalline and insoluble, stymying efforts to determine their structures at high-resolution using conventional X-ray crystallography and multidimensional nuclear magnetic resonance (NMR) spectroscopy techniques. Recently, state of the art techniques such as solid state NMR (ssNMR) have accelerated progress in the determination of fibril structures.[9] These studies have

confirmed that fibrils are polymorphic (i.e., that fibrils formed from a single protein or peptide can have multiple distinct molecular structures), and that fibril polymorphism likely arises from polymorphism in soluble precursor structures, called amyloid oligomers and protofibrils.[9] The structures of amyloid oligomers and protofibrils are even less characterized than fibril structures, due to the transient nature and polymorphism of these precursors.[11] Yet it is vital to understand these early stage amyloid structures, as they may seed fibril growth (or lead to off-pathway structures), and moreover, may be the most toxic structures in amyloid diseases.[13,14] A better understanding of amyloid oligomer and protofibril structures may enable the design of agents that detect, stabilize, and/or destabilize these structures, which could lead to the development of medical tools that diagnose early stages of amyloid formation, prevent, and/or treat amyloid diseases.

It is still unclear which amyloid polymorphs are pathogenic, which are benign, and which are artifacts of different preparation protocols.[9,11] However, studies of amyloid fibrils suggest that polymorphs important to disease pathogeneses contain in-register parallel β-sheet structure, and cross-β structure from interdigitation of side-chains extending orthogonal to the β-sheets (**Figure 3.1** inset).[9,11,12,15-17] Here we design and characterize a water-soluble, minimal folding unit of amyloid-like tertiary structure, containing both parallel β-sheet structure and interdigitation of hydrophobic side-chains. To promote hydrogen-bonding between parallel β-strands, we employ diamine and diacid linkers previously designed by our laboratory. To promote interdigitation of hydrophobic side-chains of neighboring β-strands, we design a novel linker, based on a β-arc consensus sequence.

3.3. Methods

3.3.1. Bioinformatics Studies

Kajava and coworkers classified β-arches by the backbone dihedral angles of the β-arcs, and identified several classes that produce 180° turns: three-residue (*ppl*, *xbl*), four-residue (*bepl*, *bebl*, *gbpl*, *bgpp*), five-residue (*blbbl*), and six-residue (*baepep*, *bllpbl*) β-arcs, where *a* = (right-handed) α-helix, *b* = β-strand, *d* = bridge (δ), *e* = extended chain (ε), *g* = γ', *l* = left-handed α-helix, *p* = polyproline-II, and *x* = "around Φ = -80° and Ψ = -140°" regions of the Ramachandran plot.[18] We mined the RCSB Protein Data Bank (PDB) for structures containing these backbone dihedral angles, using DSSP to assign dihedral angles.[19-21] The PDB was culled by sequence identity (30%), resolution (2.0 Å), and R-factor (0.25), on December 30, 2012.[22] After visual inspection of the culled set, we obtained the following sets of β-arches: 26 *ppl*, 0 *xbl*, 0 *bepl*, 8 *bebl*, 5 *gbpl*, 2 *bgpp*, 12 *blbbl*, 7 *bllpbl*, and 1 *baepep* β-arches. Of these, 3 *ppl*, 3 *bebl*, 5 *gbpl*, 1 *bgpp*, 12 *blbbl*, and 7 *bllpbl* β-arches produced 180° turns (**Figure 3.5**). Sequence logos in **Figure 3.5** were constructed with WebLogo.[23]

Using a set of 16 *blbbl* β-arcs – the 12 *blbbl* β-arcs in our culled dataset, plus 4 *blbbl* β-arcs that were retrieved from our culled PDB upon slight broadening of our backbone angle constraints for the β-strands that flank *blbbl* (i.e., slight relaxation of the definition of *b*, for positions "1B1"-"1B3" and "2B1"-"2B3") – a new sequence logo was generated (**Table 3.1**, **Figure 3.6**). These data were normalized to account for the natural abundance of amino acids.[24]

To better characterize the influence of each element of the *blbbl* sequence on the β-arc's conformation, we mined the PDB (via DSSP) for sequences in which one element of *blbbl* was allowed to adopt any backbone conformation (*n*): *nlbbl*, *bnbbl*, *blnbl*, *blbnl*, and

blbbn. In addition to the *blbbl* set, these searches returned culled sets of 1 *nlbbl*, 30 *bnbbl*, 2 *blnbl*, 3 *blbnl*, and 24 *blbbn* sequences. Of these, 1/1 *nlbbl*, 1/30 *bnbbl*, 2/2 *blnbl*, 1/3 *blbnl*, and 3/24 *blbbn* sequences were β-arcs (not necessarily 180°). For comparison, 12/13 of the *blbbl* sequences were 180° β-arcs. Examination of all these sequences revealed possible contributions of backbone angles to the β-arc (**Figure 3.8**).

Kajava and coworkers identified single cases of non-*blbbl*, five-residue, 180° β-arcs: *blbbx*, *blbpl*, and *bldpl*.[18] The Ramachandran basins *p* and *d* neighbor the β-strand (*b*) basin, while we define here a region *x* ($\Phi \in \left[30^{\circ},110^{\circ}\right]$, $\Psi \in \left[90^{\circ},120^{\circ}\right]$) that neighbors the left-handed α-helix (*l*) basin. Our searches identified no *blbbx* β-arches in our *blbbn* set, and one *blbpl* β-arch in our *blbnl* set. Mining the culled PDB for *blnnl* (including *bldpl*) sequences did not return any β-arches (besides the aforementioned two *blnbl* β-arches and one *blbnl* β-arch).

For each β-arch in our culled dataset, we used established methods to determine the solvent-accessible surface area (ASA) of the residue side-chains in the PDB structure;[25,26] assigned "outside (o)" or "inside (i)" to side-chains that were solvent-accessible or inaccessible, respectively; and generated the o/i consensus sequence in **Figure 3.7**. A side-chain was deemed to be solvent-accessible, if its relative accessibility was >5%, as calculated using NACCESS.[26]

3.3.2. Synthesis and Characterization of Peptides

The OAllyl-Gly-(1*R*,2*S*)-CHDA-Val-OH and Alloc-Glu-Val-^{D}Pro-DADME-Fmoc linkers were synthesized via the routes described in **Appendix 1**.[27] Peptide **3.A** was synthesized via the route depicted in **Figure 3.2**, and all other peptides were synthesized by analogous

routes. Peptides were purified by reversed-phase high-performance liquid chromatography (HPLC), using a C18 preparatory column, and a mobile phase of CH_3CN:H_2O:CF_3COOH. Peptide identities were confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (**Figures 3.4**). MALDI-TOF MS showed the following m/z values for the $[M+H^+]^+$ species: peptide **3.A** calculated = 3488.93, observed = 3487.9; peptide **3.B** calculated = 3574.94, observed = 3574.7; peptide **3.C** calculated = 3460.90, observed = 3460.0; peptide **3.D** calculated = 3380.86, observed = 3380.2; peptide **3.E** calculated = 3251.77, observed = 3250.9; peptide **3.F** calculated = 3488.93, observed = 3488.3; peptide **3.G** calculated = 2722.50, observed = 2722.0. Purity was determined by analytical HPLC or ultra performance liquid chromatography (UPLC), and was >95% for all purified peptides (**Figures 3.3**).

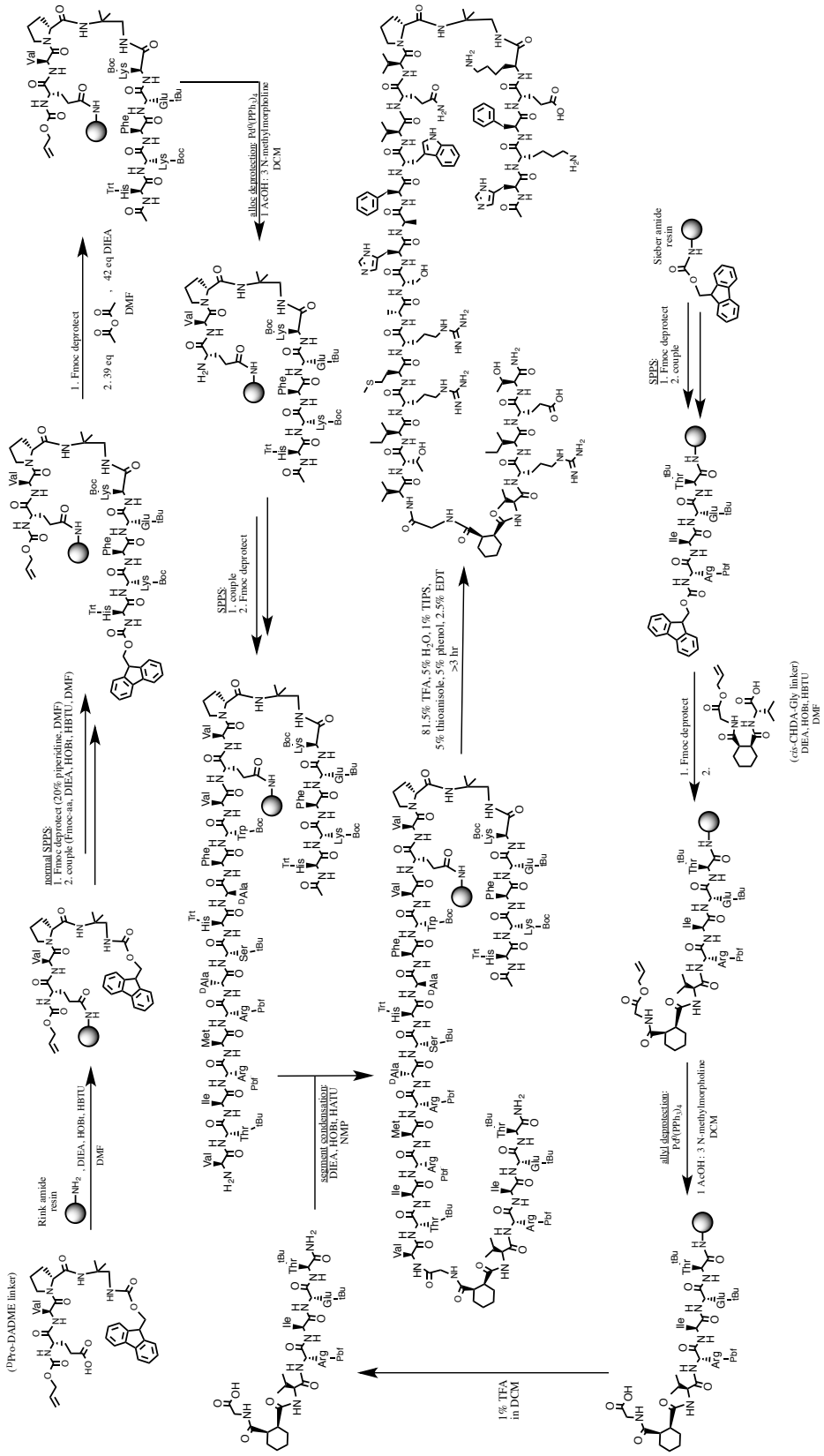

Figure 3.2. Synthetic route to peptide **3.A.** Analogous routes were employed to synthesize other peptides.

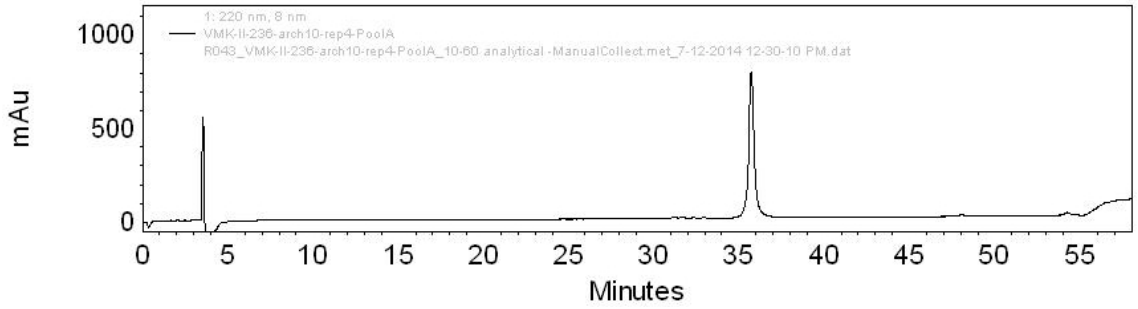

Figure 3.3A. Analytical HPLC trace for peptide **3.A.**

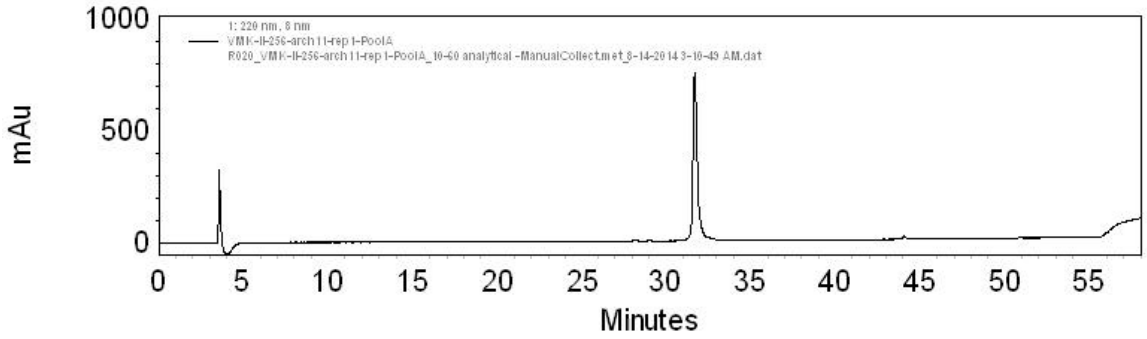

Figure 3.3B. Analytical HPLC trace for peptide **3.B**.

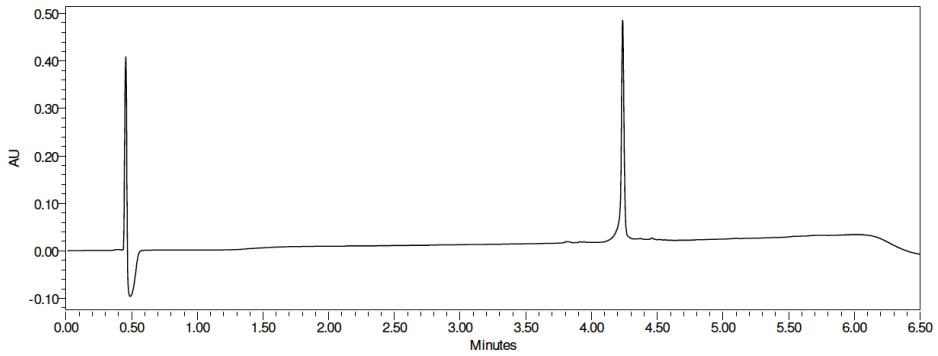

Figure 3.3C. UPLC trace for peptide **3.C**.

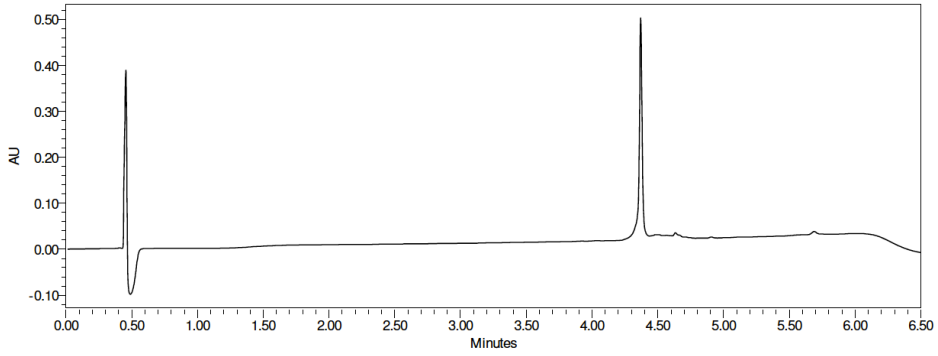

Figure 3.3D. UPLC trace for peptide **3.D**.

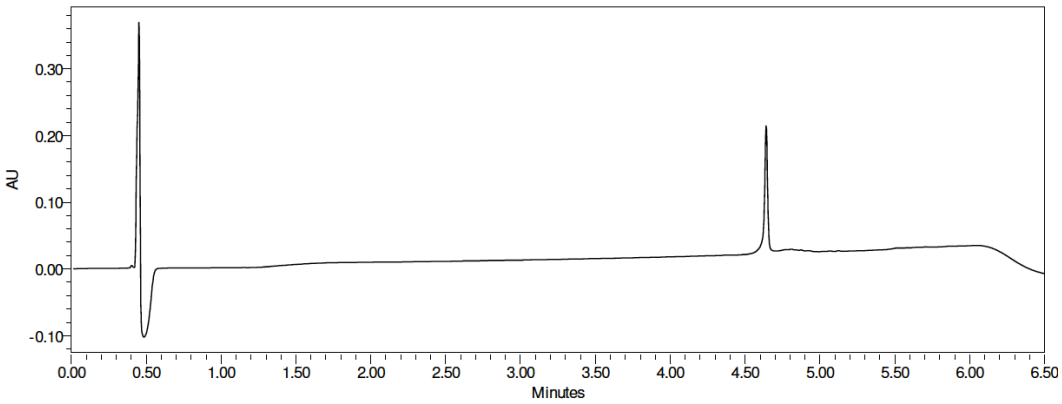

Figure 3.3E. UPLC trace for peptide **3.E**.

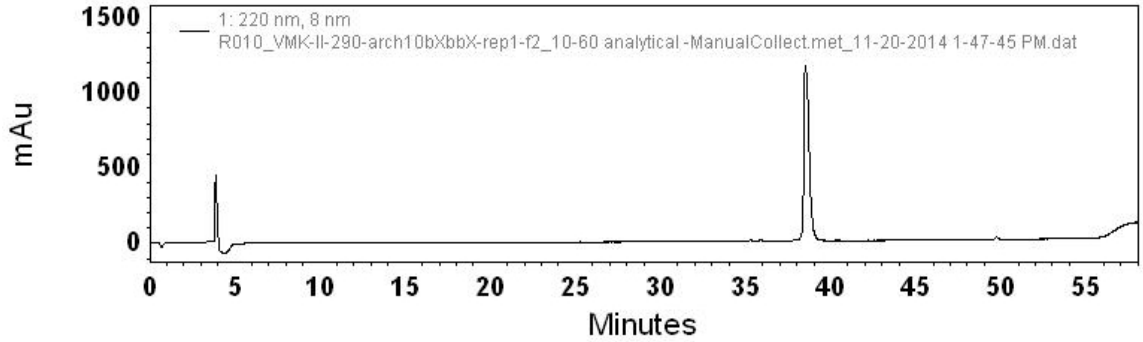

Figure 3.3F. Analytical HPLC trace for peptide **3.F**.

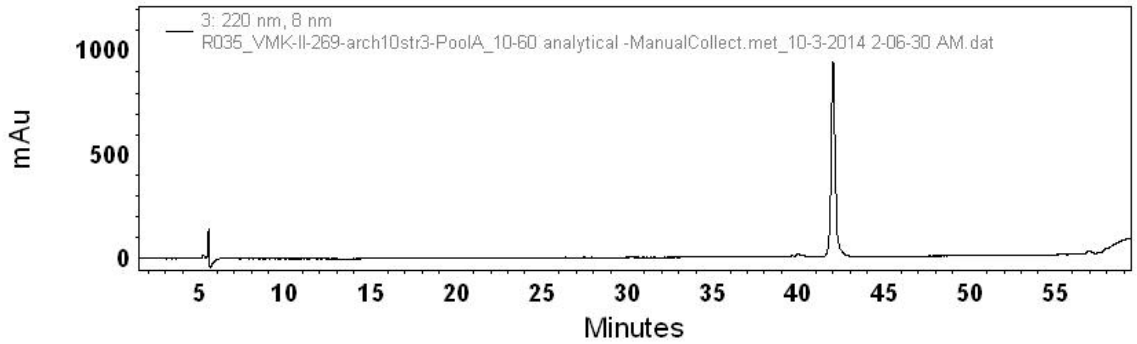

Figure 3.3G. Analytical HPLC trace for peptide **3.G**.

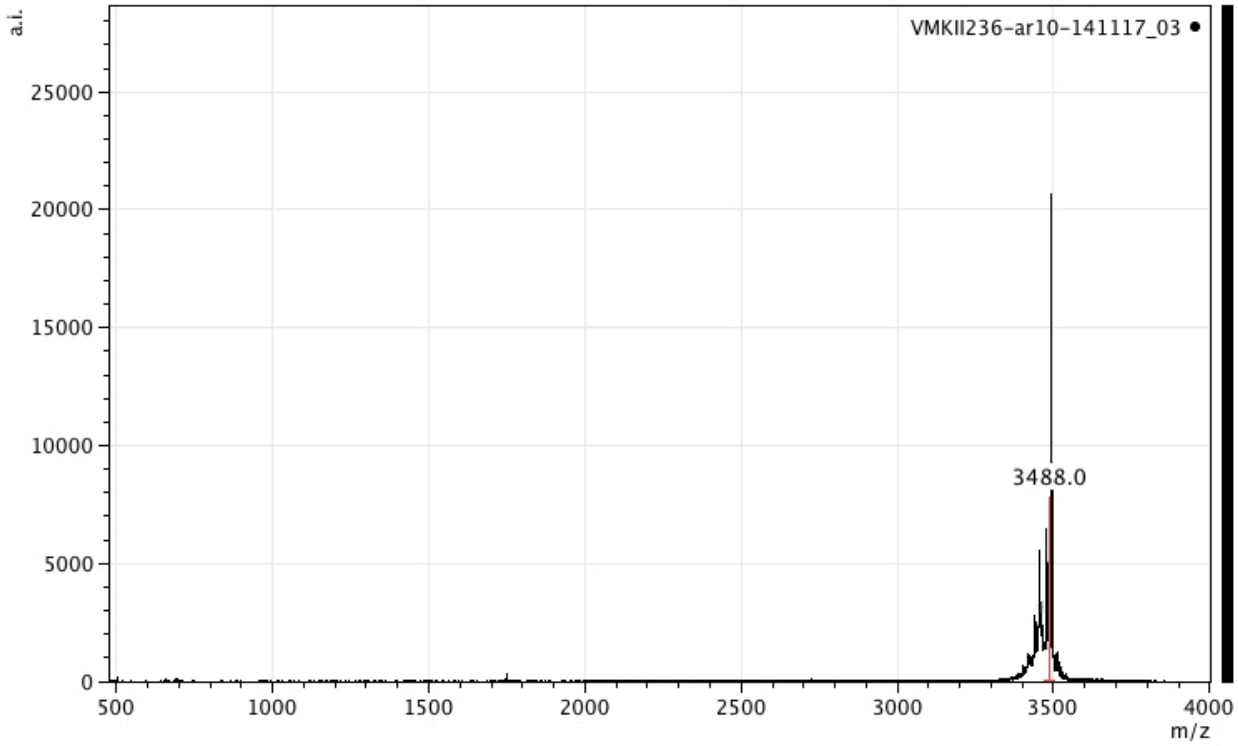

Figure 3.4A. MALDI-TOF of peptide **3.A**.

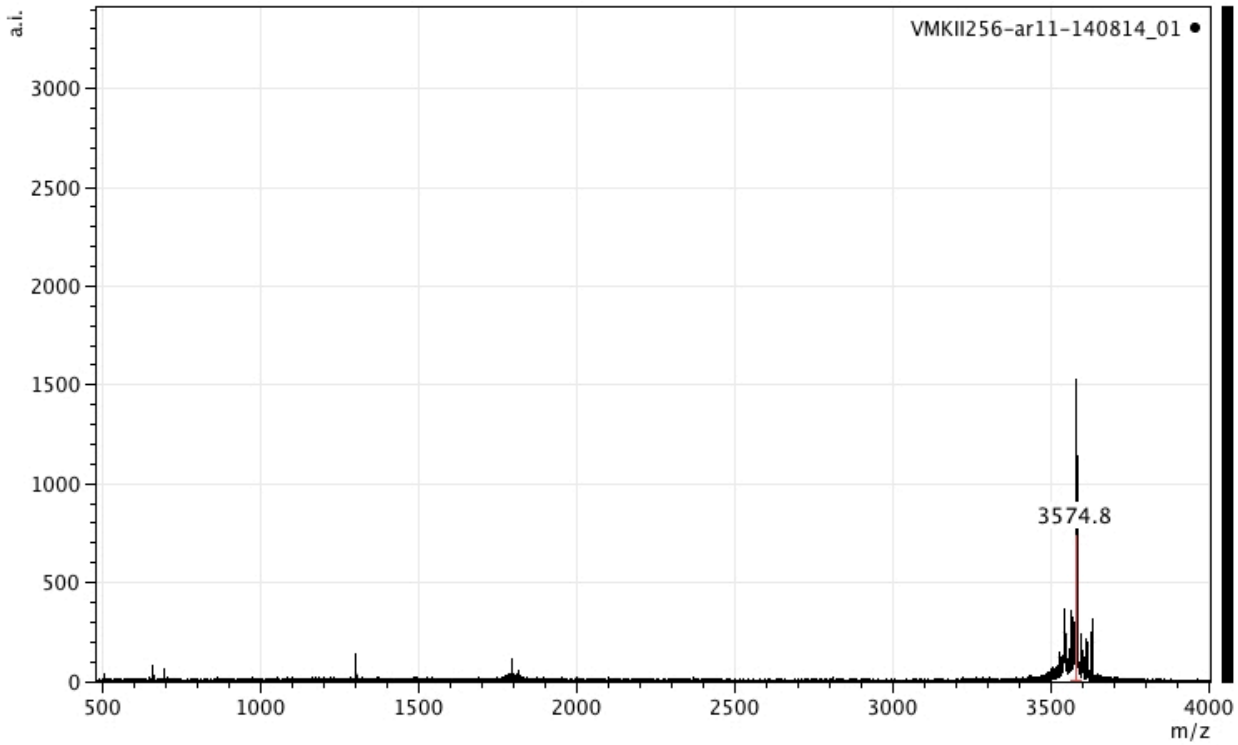

Figure 3.4B. MALDI-TOF of peptide **3.B**.

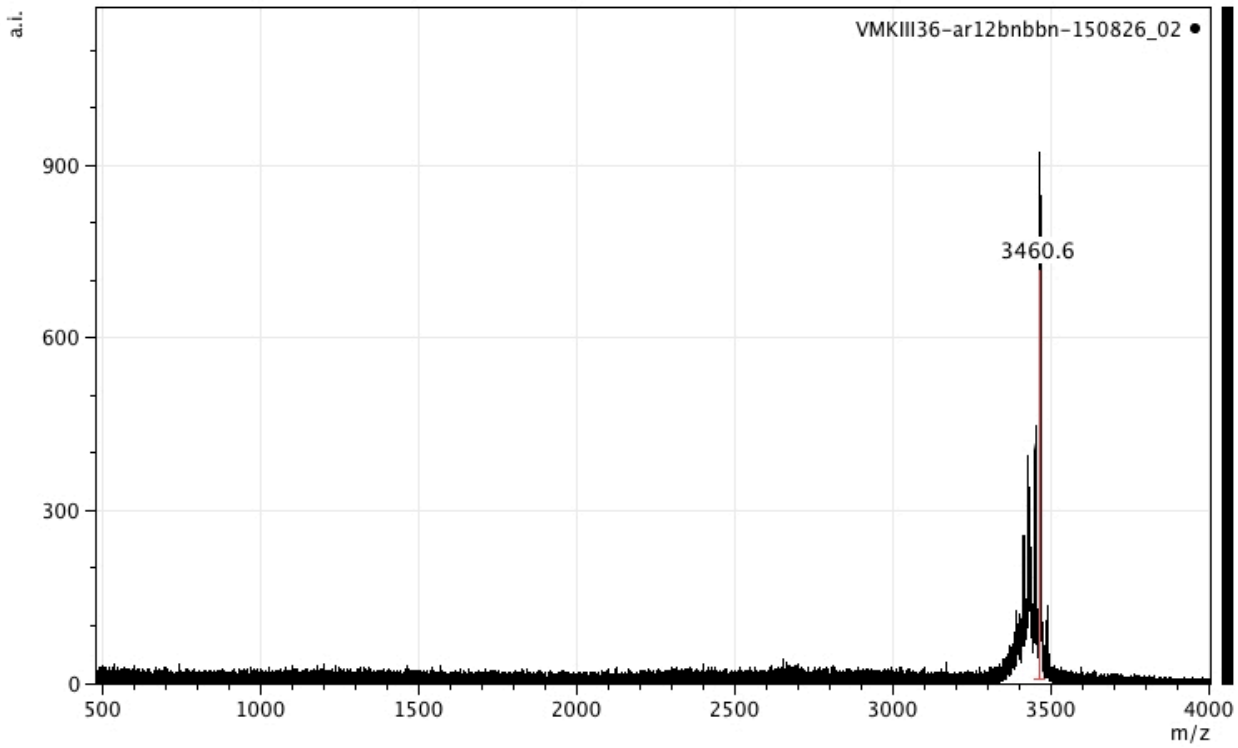

Figure 3.4C. MALDI-TOF of peptide **3.C**.

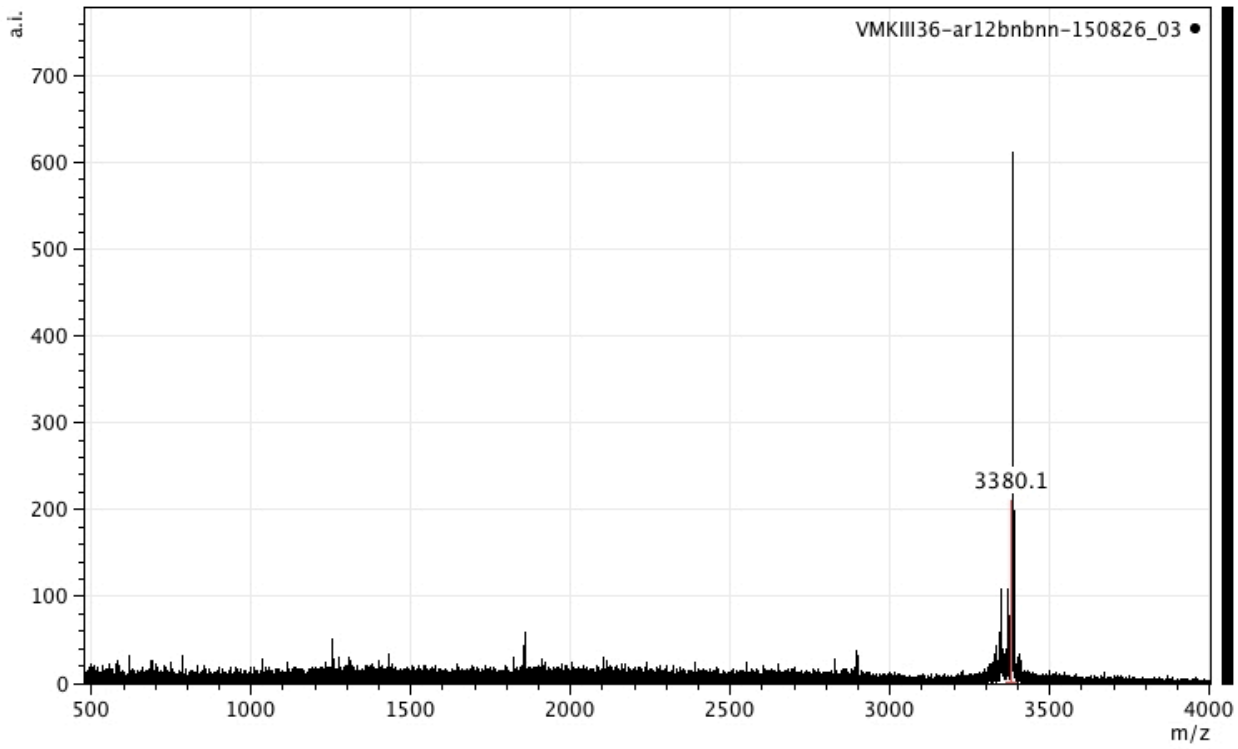

Figure 3.4D. MALDI-TOF of peptide **3.D**.

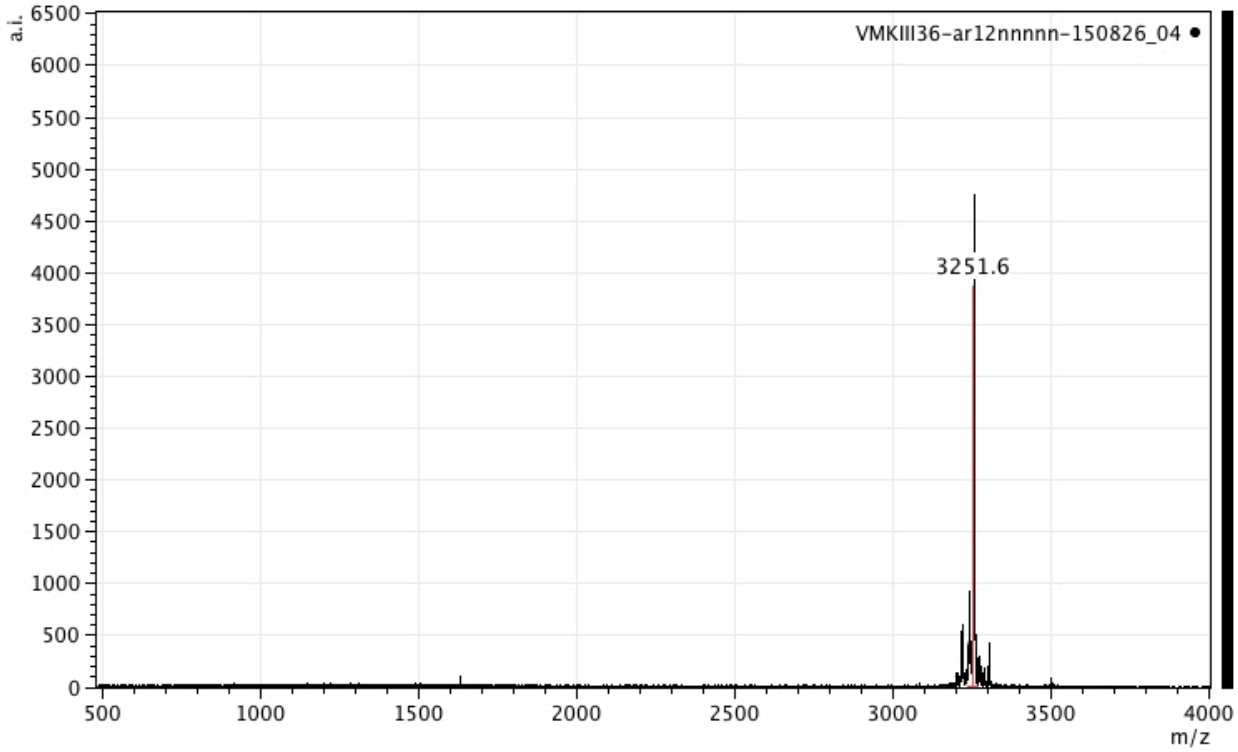

Figure 3.4E. MALDI-TOF of peptide **3.E**.

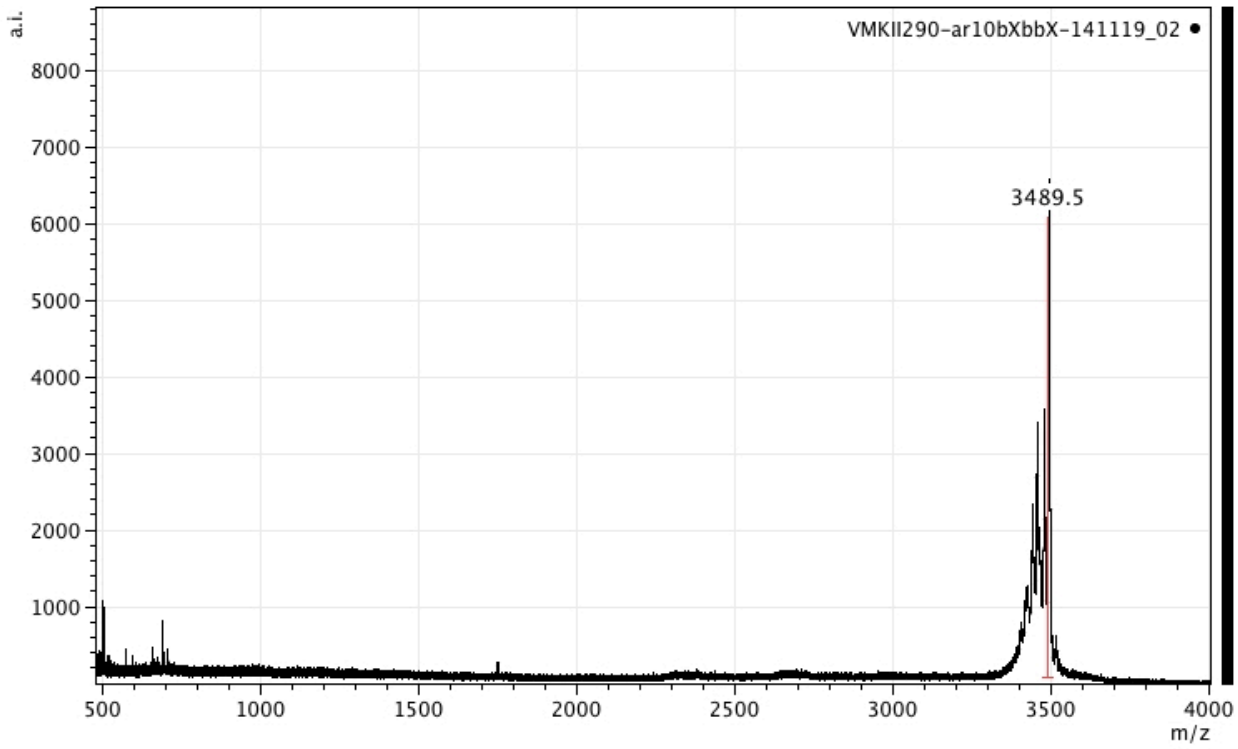

Figure 3.4F. MALDI-TOF of peptide **3.F**.

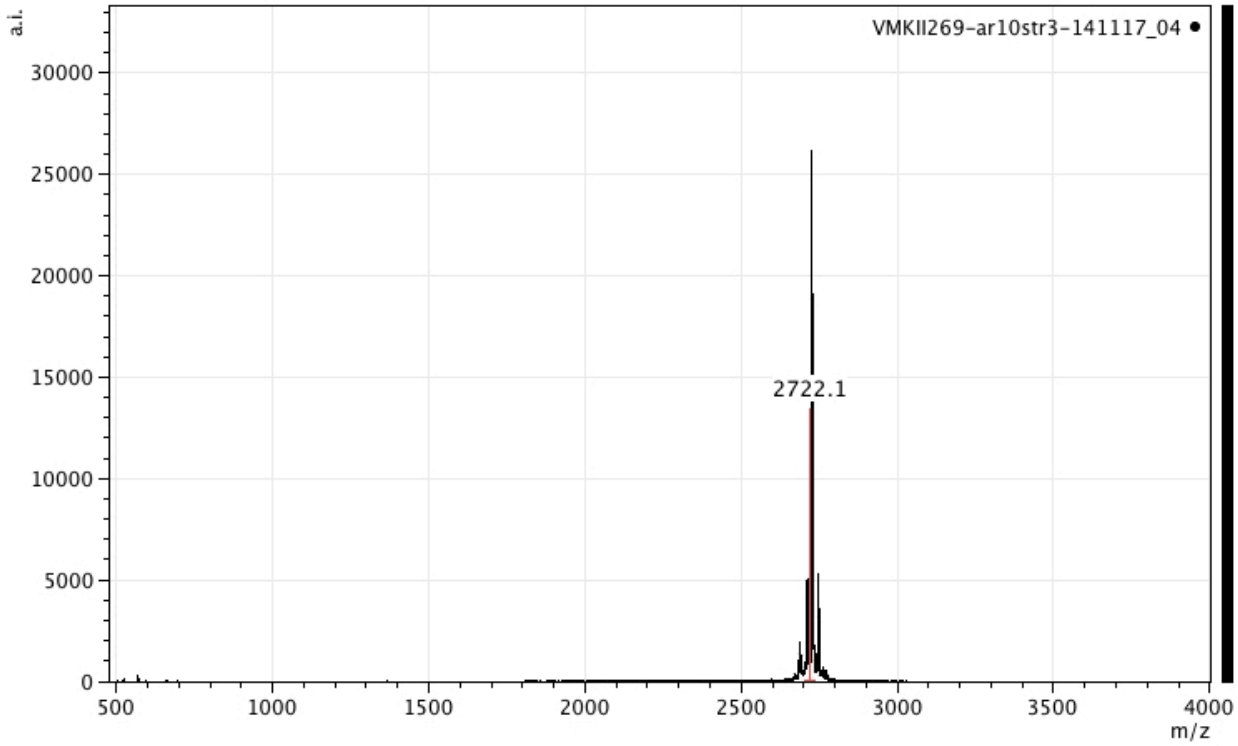

Figure 3.4G. MALDI-TOF of peptide **3.G**.

3.3.3. 2D-NMR data acquisition and NMR structure determination

Each NMR sample was prepared by dissolving lyophilized peptide to 0.2 mM concentration in 9:1 H_2O:D_2O, pH 3.8, 3 mM NaOAc-d_3 buffer, with trace amounts of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and/or dioxane as internal references. Peptide concentrations were determined by mass. Samples were prepared with total volumes of approximately 450 µL in 5 mm NMR tubes. Peptide samples were stable in solution for at least 1 week at 4-5 °C, showing no visible precipitation of peptide, NMR line broadening, or decrease in NMR signal intensity over the entire period of study. Low salt conditions were employed to facilitate observation of weaker NMR signals. For each peptide, the following sets of 1D ^{1}H spectra were superimposable: [1] before and after collection of 2D NMR spectra (COSY, TOCSY, NOESY), and [2] 2 mM peptide in low salt

buffer (9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8) and 2 mM peptide in high salt buffer (9:1 H_2O:D_2O, 100 mM NaOAc-d_3, pH 3.8).

Spectra were acquired at 5.0 °C on a Bruker Avance 600 MHz spectrometer, equipped with a 5 mm, z-axis gradient, triple resonance, cryogenic probe. The following standard Avance pulse programs were employed: 1D with water suppression using excitation sculpting (zgesgp),[28] phase-sensitive 2D COSY-DQF with WATERGATE using 3-9-19 (cosydfgpph19),[29-31] phase-sensitive 2D TOCSY with WATERGATE 3-9-19 and mixing using DIPSI-2 (dipsi2gpph19),[30,31] and phase-sensitive 2D NOESY with WATERGATE using water flip-back and 3-9-19 (noesyfpgpph19)[30-32]. Spectral widths of 8013-8418 Hz were used, with collection of 2048 points in the direct (f2) and 256 points in the indirect (f1) dimensions. TOCSY experiments used a mixing time of 100 ms. NOESY experiments used a mixing time of 250 ms. For peptide **3.A**, increasing the NOESY mixing time to 500 ms did not increase the number of NOEs observed. Note also that variable temperature studies suggested peptide unfolding at higher temperatures, with less peak dispersion (**Figure 3.11**), less downfield HN and Hα chemical shifts, and fewer NOEs at higher temperatures. Data were processed using NMRPipe.[33] Data were analyzed using Sparky,[34] with employment of sequential assignment procedures to assign chemical shifts of protons (**Tables 3.4**).[35] Inter-residue NOEs are diagramed in **Figures 3.14**.

NOE peak intensities were translated into a continuous distribution of interproton distance restraints, with a 35% distance uncertainty applied to take into account spin diffusion. For inferred hydrogen bonds, distance restraints of 1.9 Å and 2.9 Å were used for the H_N-O and N-O non-bonded distances, respectively.[35] To check for biases caused by hydrogen bond restraints, structure calculations were performed without these restraints; there were no major

changes in the resulting strand registers or in the backbone conformations in the residues for which hydrogen bond inferences had been made. The secondary chemical shift, or difference between an observed chemical shift and random coil values, was calculated for α-protons (**Figures 3.12**). (Note that only Hα secondary chemical shifts of peptides **3.A** and **3.G** are numerically reported here, but comparisons of Hα and HN chemical shifts can be made for all peptides using the overlaid 2D TOCSY spectra in **Figure 3.13**.) Calculated chemical shifts for peptides in random coil conformations (in H_2O, pH 3.80, 277.15 K) were obtained from the CSDb[2] program, which takes into account nearest-neighbor and other effects.[36] Compared to random coil values, amino acid residues experience a mean α-proton downfield shift of 0.37 ppm when placed in a β-strand or extended configuration (and a mean α-proton upfield shift of 0.39 ppm when placed in a helix).[37] A gross pattern of downfield α-proton chemical shifts can be a sensitive indicator of β-strand backbone dihedral angles.[38-40] β-Strand backbone dihedral angle ($\Phi \in [-180, -110]$, $\Psi \in [90, 180]$)[41] restraints were applied for residues in regions with at least three consecutive residues with secondary chemical shift > 0.1 ppm.[42] To check for biases caused by backbone dihedral angle restraints, structure calculations were performed without these restraints; there were no major changes in the resulting backbone conformations.

Three-dimensional structure calculations and refinements, with the described structural restraints, were carried out using the torsion angle molecular dynamics and the internal variable dynamics modules of Xplor-NIH (v. 2.35), with patches for the non-natural residues.[43] The target function minimized was comprised of the experimental NMR restraints (NOE-derived interproton distances, inferred hydrogen bonds and torsion angles), a repulsive van der Waals potential for the non-bonded contacts,[44] a torsion angle database potential of mean force,[45] and a gyration volume potential[46]. The initial topology and parameter files for non-natural residues

(DADME, CHDA) were generated previously.[27]

An ensemble of the 10 lowest energy (out of 100) structures for peptide **3.A** is shown in **Figure 3.15**, and the average of this ensemble is shown in **Figure 3.17**. There were no (0) consistent violations of restraints (NOE-derived distance, hydrogen bond, dihedral angle). Structural statistics for all (not just ordered) residues were generated using the Protein Structure Validation Suite (PSVS) version 1.5,[47] and are reported in **Table 3.5**. The structures of these peptides in solution are more dynamic than is typical of well-ordered globular proteins, such that the NOE intensities represent averages over multiple conformers, but we make the simplifying assumption of there being a single set of structures. One result of this assumption is that reported clash scores are poorer than those typical of standard PDB structures.

3.3.4. Circular Dichroism (CD) Spectroscopy

Circular dichroism can be used to assess protein folding in solution, with the association of α-helix (208 nm, 222 nm) and antiparallel β-sheet (218 nm) secondary structures with characteristic minima in CD spectra.[48] The CD signature of parallel β-sheets is not as well characterized,[49,50] due to the difficulty of parallel β-sheet isolation. We have previously reported CD wavelength scans for parallel β-sheets of defined lengths in aqueous solution.[27] Here we report a CD wavelength scan (**Figure 3.19** red) for a minimal folding unit of amyloid-like structure, containing two two-stranded parallel β-sheets interacting via their hydrophobic side-chains (**Figure 3.17**). CD measurements were taken with an AVIV Model 420 Circular Dichroism Spectrometer, using 1.00 nm bandwidths and 10.000 sec averaging times, in 1.00 mm quartz cuvettes. For each sample measurement, the corresponding buffer blank measurement was subtracted, and the baseline CD signal at 320 nm was set to zero. Peptide solutions were

prepared in aqueous solution with low (3 mM NaOAc, pH 3.8) or high (3 mM NaOAc, 100 mM NaCl, pH 3.8) concentrations of a charge-shielding salt. Peptide concentrations were determined by mass, and verified with UV absorption measurements. CD spectra for peptide samples at 20.0 °C (293.2 K) are reported (**Figures 3.19**, **3.29**). For thermal denaturation experiments, measurements were taken every 2 °C in the 2.0-98.0 °C (275.2-371.2 K) temperature range. CD signals were also measured every 2 °C as solutions were cooled, and heating and cooling transitions were found to be superimposable, indicating that unfolding reactions were reversible. Addition of an alcohol co-solvent can enhance CD signals attributable to secondary structures, and reduce CD signals attributable to tertiary or higher order structures.[51]

3.3.5. Transmission Electron Microscopy (TEM)

Peptide samples were prepared at 0.08 mM concentration peptide in aqueous solution (3 mM NaOAc, pH 3.8), and incubated at 4 °C for 0-6 days. Samples were negative stained with Nano-W (methylamine tungstate, from Nanoprobes in Yaphank, NY), using the two-step method.[52] A 2 μL droplet of peptide sample was placed on a Pioloform (Ted Pella, Redding, CA) coated, 300 mesh copper Thin-Bar grid (Electron Microscopy Sciences, Hatfield, PA), coating side down. The excess was wicked with filter paper and allowed to barely dry. Then a 2 μL droplet of Nano-W was applied, wicked with filter paper, and allowed to dry. Samples were viewed on a Philips CM120 transmission electron microscope at 80 kV, and documented with a SIS (Olympus Soft Imaging Solutions, Munster, Germany) MegaView III digital camera. For peptides **3.A**-**3.G**, representative images are reported in **Figures 3.21-3.27**.

3.3.6. Sedimentation Equilibrium Analytical Ultracentrifugation (AUC)

Our AUC experimental design and data analysis methods were reported in greater detail elsewhere.[27] Sedimentation equilibrium data were collected using a Beckman Coulter Model XL-A Analytical Ultracentrifuge at 4 °C, for peptide **3.A**. Peptide solutions were prepared by simple dissolution in buffer (3 mM NaOAc, pH 3.8, with peptide concentrations of 0.09, 0.06, and 0.03 mM) and used without further manipulation. Initial peptide concentrations were determined from UV absorption spectra of samples diluted into 6 M GdmCl, using an extinction coefficient (ε) of 5690 M^{-1} cm^{-1}.[53] Gradients were monitored at the wavelength of maximal absorbance (276 nm for peptide **3.A**). Data were collected at rotor speeds of 10k, 15k, 23k, 32k, 42k. For each rotor speed, we assumed samples had reached equilibrium when gradients collected at least 3 hours apart were superimposable. After data collection at the highest rotor speed, gradients at one of the lower speeds were measured again and found to be superimposable with original measurements, indicating that there had been no irreversible loss of material during the course of the experiment. The equilibrium data were analyzed using software written by Darrell R. McCaslin, in a manner analogous to previously-described analysis methods.[54] In the analysis, the solvent density (ρ) at 4 °C was computed using density increment functions to be 1.001 g/mL.[55] The partial specific volume ($\bar{v}$) of peptide **3.A** (0.736 cm^3 g^{-1}) was calculated using consensus values for the natural amino acid residues, and corrections for acetyl, carboxamide, DADME, and CHDA groups.[56,57] The molecular weight (M_S) of peptide **3.A** (3489.2 Da) was calculated based on the peptide sequence, including non-natural linkers and termini. Plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center of rotation were linear, at all speeds and for all loading concentrations (**Figure 3.10**). Linearity in such plots is indicative of the presence of a single molecular weight

entity. The final estimate of the weight-average molecular weight (M_p, or MW_{obs}), reported in **Table 3.2**, was obtained from global fits of the data at the varied loading concentrations and speeds to the distribution of a single macromolecular species. Inclusion of a term for a second macromolecular species worsened fits.

AUC studies were also performed for peptide **3.A** (0.09, 0.06, and 0.03 mM) in buffer containing 100 mM neutral salt (NaCl). Data collection and analysis were performed as above. Note that the density of the "with salt" buffer (3 mM acetate, 100 mM NaCl, pH 3.8) at 4 °C was computed to be 1.005 g/mL.[55] For the samples with salt, data were collected at rotor speeds of 10k, 15k, 23k, 32k, 37k. With salt, plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center of rotation were nonlinear, suggesting the presence of multiple molecular weight entities and a self-associating system. Data could not be fit to one macromolecular species (with the reduced molecular weight of the monomer or dimer) or two macromolecular species (with the reduced molecular weights of the monomer and dimer). Partially reversible peptide precipitation was observed at higher rotor speeds.

3.3.7. NMR diffusion experiments

NMR experiments measuring peptide diffusion were performed. Samples were prepared by dissolving lyophilized peptide to ~2 mM concentration, with 0.6 mM dioxane, in aqueous buffer (9:1 H_2O:D_2O, 3 mM d_3-acetate, pH 3.8). Serial dilutions were made of these peptide solutions into buffer, to yield 0.2 mM and 0.02 mM peptide solutions. Spectra were acquired on a Bruker Avance 600 MHz spectrometer, equipped with a 5 mm, z-axis gradient, triple resonance, cryogenic probe. Self-diffusion rates were measured using a pulsed field gradient stimulated spin echo (PFGSTE) technique.[58] Spectra were acquired at 5.0 °C, with 75 pulsed field gradient (PFG) strengths in the 5-60 G/cm range, and acquisition of 16-240 transients per

gradient experiment. Gradient pulses were applied for 5 ms, with a diffusion delay of 100 ms.

Data were analyzed using the variable gradient least square fitting routines in NMRPipe,[33] and peak intensities were fit with a single exponential decay function to determine translational self-diffusion constants. For each condition studied, the self-diffusion constant of the peptide represented an average of values obtained from the intensities of at least 4 separate peaks, including peaks in the aromatic and aliphatic regions. The hydrodynamic radius (R_H) of the peptide was extracted by comparison of peptide diffusion data with that of an internal standard, 1,4-dioxane (R_H = 2.12 Å).[59]

3.4. Results

3.4.1. β-Arc Consensus Sequences

The β-arch is a protein motif containing a turn called a β-arc, which orients flanking β-strands such that they interact via their side-chains (**Figure 3.1** inset).[12,60] A previous bioinformatics study of β-arcs identified common backbone dihedral angles in these turns.[18] The RCSB Protein Data Bank (PDB)[19] has grown in the intervening decade, and we mined the updated PDB for structures containing β-arc backbone dihedral angles, using DSSP to assign dihedral angles.[20,21] We report updated sequence logos for β-arcs that are three to six residues in length and create 180° turns (**Figure 3.5**).

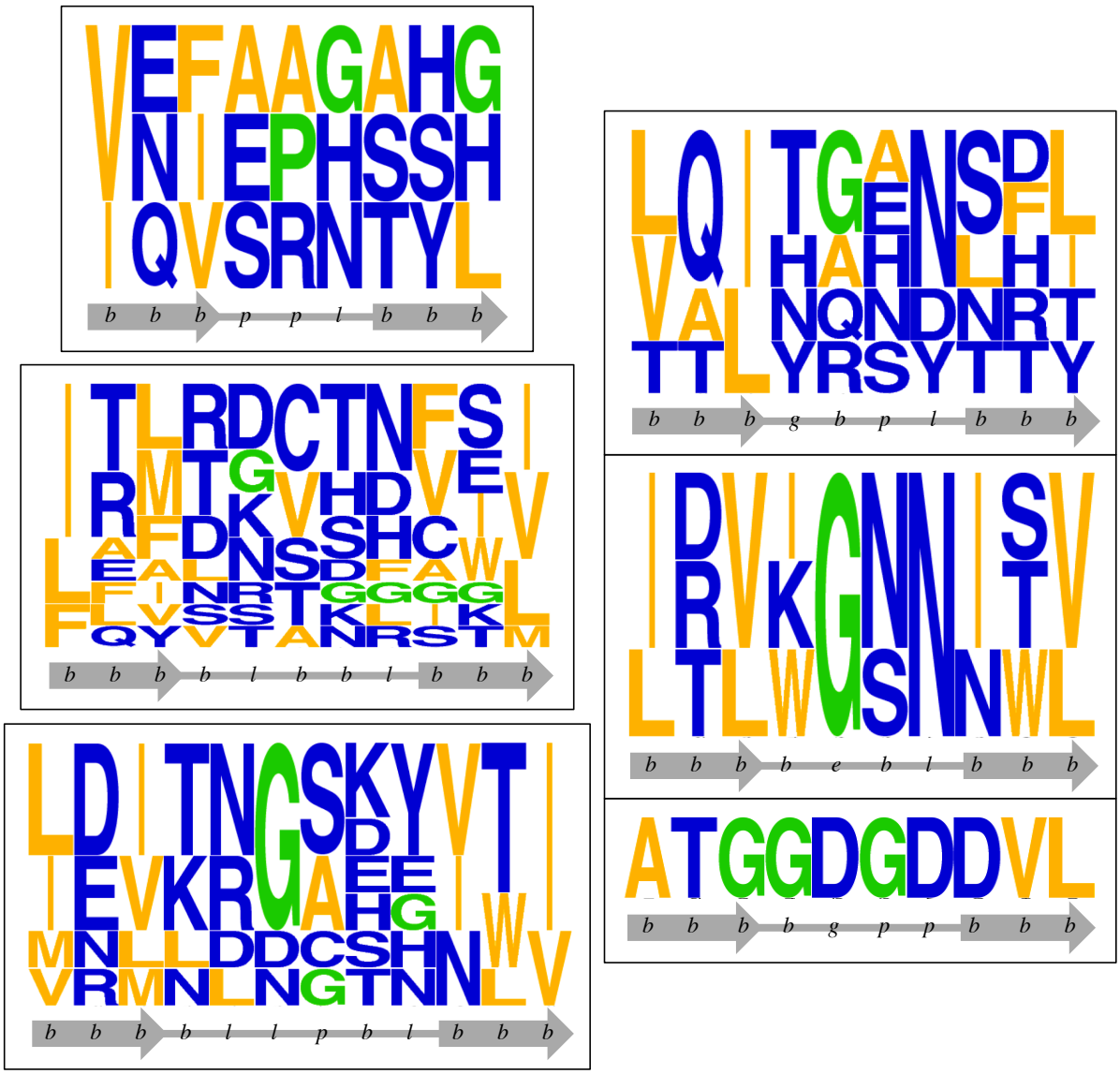

Figure 3.5. Sequence logos (residue frequency vs. position) for *ppl*, *gbpl*, *bebl*, *bgpp*, *blbbl*, and *bllpbl* β-arcs that produce 180° turns. A similar sequence logo for the *blbbl* β-arc (with a slightly relaxed definition of *b* dihedral angles for the flanking β-strands) appears as **Figure 3.6.**

We also perform more detailed analyses on the largest dataset: five-residue β-arcs (**Table 3.1**). The five residues of these β-arcs have backbone dihedral angles of *blbbl*, where *b* indicates the β-strand region, and *l* indicates the left-handed α-helix region of the Ramachandran plot. We obtain a sequence logo for five-residue β-arcs (**Figure 3.6**), with normalization by amino acid content in UniProt.[24] The consensus distribution of polar

(blue) and nonpolar (orange) residues (**Figure 3.6**) closely corresponds to that of water-accessible (o) and buried (i) residues, respectively (**Figure 3.7**).

Table 3.1. The redundant set (culled set in red font) of *blbbl*-containing sequences in December 30, 2012 PDB. The sequences that had a *blbbl* backbone conformation, but that were not 180° β-arcs on visual inspection, are in gray boxes.

PDB code	A1 res#	*bbb*	***blbbl***	*bbb*
1a9x	4171	---	SFTMG	---
1ag8	453	TVW	VNCYD	VFG
1bhe	106	ITA	VSTTN	SGI
"	157	IQI	NKSKN	FTL
"	207	---	MSSKN	---
1clw	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1cs0	171	---	SFTMG	---
1czf	185	FDV	GNSVG	VNI
1edq	452	---	NGYQN	---
1ehn	452	---	NGYQN	---
1eib	452	---	NGYQN	---
1ezg	29	---	TNSQH	---
"	41	---	TGSTD	---
"	53	---	TNSKD	---
"	65	---	TDSTN	---
"	77	---	TNSSG	---
1ffr	452	---	NGYQN	---
1go7	67	---	TNVFG	---
1go8	67	---	TNVFG	---
1gv4	578	---	WNVFN	---
1h54	490	---	NNVDN	---
1hg8	157	---	TGSSQ	---
"	196	FDI	SSSDH	VTL
1ia5	110	---	HSLTN	---
"	133	FSV	AGSDY	LTL
"	164	FDI	GTSTY	VTI
1idk	159	ITL	DDCDL	VWI
1iru	156	---	KNMQN	---
1ja9	170	---	TGIPN	---
1k4z	1404, 2404	IFI	GKCSQ	VLV
"	1423, 2423	ISL	SETES	CSV
"	1442, 2442	MDV	IKSNK	FGI
"	1482, 2482	IYT	SCSTA	INV
1k7g	67	---	TNVFG	---
1k7i	67	---	TNVFG	---
1k7q	67	---	TNVFG	---
1k8f	354	---	YKCVN	---
"	373	ITV	DNCKK	LGL
"	392	VEI	INSKD	VKV
"	432	IVS	AKSSE	MNV
1k9t	452	---	NGYQN	---
1kee	171	---	SFTMG	---
1kq5	1404, 2404	---	GKCSQ	---
"	1423, 2423	ISL	SETES	CSV
"	1442, 2442	MDV	IKSNK	FGI
"	1482, 2482	IYT	SCSTA	INV
1m6i	579	---	WNIFN	---
1mhs	818	---	FWSSI	---
1nh6	452	---	NGYQN	---
1nhc	138	---	HDVED	---
"	191	FDI	SESTG	VYI
1o04	453	TVW	VNCYD	VFG
1o88	258	---	EKAIN	---
1o9j	453	---	VNCYL	---
1p4k	80, 380	---	ACMEH	---
1pxz	111	LFM	RKVSH	VIL
"	156	ITM	RNVTN	AWI
1q1l	65	---	SGVRF	---
1qa1	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qa2	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qa3	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qq1	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qrb	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qrc	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1qxo	60	---	SGVRH	---
1r53	67	---	SGTEF	---
1rmg	96	IAV	TDTTD	FEL
"	133	LRL	TDVTH	FSV
"	156	FTM	DTCSD	GEV
1tyv	265	LEI	RECIG	VEV
"	285	FLF	RGCHF	CKM
1u79	104	---	AGCKG	---
1um0	67	---	SGVFE	---
1uun	152	---	TGAAG	---
1vbl	153	---	KNVDN	---
"	195	ISI	EGSSH	IWI
1vz2	169	---	ERVKF	---
1vz3	169	---	ERVKF	---
1whu	356	---	EYKRC	---
1x6l	452	---	NGYQN	---
1y4z	394	---	WGEKG	---
1y5i	394	---	WGEKG	---
1yo8	1155	---	FSQEX	---
1zum	453	TVW	VNCYD	VFG
2a8j	131	---	GALSG	---
2a8l	131	---	GALSG	---
2a8m	131	---	GALSG	---
2b0r	55	---	ENHVN	---
"	97	IFI	ENCVG	CIF
"	116	IEI	VNCDD	IKL
"	135	ISL	DKSNK	VNI
"	156	VYS	SKSSE	MNL
"	78	VSV	RCSQN	SRI
2b39	1355	---	VSCKK	---
2bx6	65	FLI	QDCEN	CNI
"	84	VTI	DDCTN	CII
"	103	VFF	RNCRD	CKC
"	120	FRV	RDCRK	LEV
"	138	---	ESSSN	---
2c1d	(118)	---	DGFVR	---
2cu2	284	HVA	LDTFG	CVV
"	301	VAT	LGVSG	LVV
2ewe	258	---	EKAIN	---
2f8o	30	---	FHASD	---
2g85	61	---	SGIRH	---
2gaw	80	---	ACMEH	---
2gl9	80, 380	---	ACMEH	---
2gp1	305	---	SNIMW	---
2inu	277	---	ENHEG	---
"	300	IEF	TGCNR	CSV
2iq7	129	---	HSLKS	---
"	152	FSI	NSATT	LGV
"	183	FDV	GSSTG	VYI
2o2p	860	---	INTYN	---
2onp	453	TVW	VNCYD	VFG
2pyg	178	---	DYLVD	---
"	255	VLL	KMTSD	ITL
"	278	VRV	YGAQD	VQI
2qhf	61	---	SGIRH	---
2qx3	92	---	KNKSD	---
"	270	---	ENIKN	---
2qxz	92	---	KNKSD	---
"	270	---	ENIKN	---
2qy1	92	---	KNKSD	---
"	270	---	ENIKN	---
2rhp	1155	LFV	FSQEM	VYF
2uvf	360	IMN	LENHN	VVA
"	386	IEF	GNSQN	VMV
"	570	---	SDLRD	---
"	594	WHF	SEVKN	VKV
2v5i	310	LDI	RSASR	VEI
"	330	YLF	RSCHH	CKV
2vbk	514	LRY	LRSGG	VSV
2vfm	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfn	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfo	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfp	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vfq	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2vle	453	TVW	VNCYD	VFG
2wk2	452	---	NGYQN	---
2wme	441	---	INTWG	---
2wox	441	---	INTWG	---
2x65	283	VIL	VDSDR	VFV
"	300	IAV	VGLSD	VIV
2xc1	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
2yuh	14	---	SNLES	---
"	36	VLL	TELSN	CTV
"	55	LRL	TKAHS	CKL
"	74	VFL	EDCSD	CVL
"	91	---	HSTKD	---
"	109	---	EDCSG	---
2zah	367	---	GTVVA	---
3bh6	(43)	---	SGLKD	---
"	(65)	FLI	QDCEN	CNI
"	(84)	VTI	DDCTN	CII
"	(103)	VFF	RNCRD	CKC
"	(120)	FRV	RDCRK	LEV
"	(138)	---	ESSSN	---
3bh7	(65)	FLI	QDCEN	CNI
"	(84)	VTI	DDCTN	CII
"	(103)	VFF	RNCRD	CKC
"	(120)	FRV	RDCRK	LEV
"	(138)	---	ESSSN	---
3ddu	169	---	ERVKF	---
3egw	394	---	WGEKG	---
3eqn	620	YFL	KNAAN	HYI
"	669	MTV	QNSHG	ILV
3fby	727	---	FSQEN	---
3gd3	578	---	WNVFN	---
3gd4	578	---	WNVFN	---
3gq8	222	LSI	RACHN	VYI
"	285	ITT	HHSQY	INI
"	587	SGV	LTSSS	SKT
"	602	SLI	AGSST	SEA
"	617	NTI	LGSLG	AVA
"	632	---	ISASQ	---
"	647	NLI	LSSYG	INT
3hb2	67	---	TNVFG	---
3hbu	67	---	TNVFG	---
3hbv	67	---	TNVFG	---
3hwc	3189	---	AVGTG	---
3inl	453	---	VNCYD	---
3ir5	394	---	WGEKG	---
3ir6	394	---	WGEKG	---
3ir7	394	---	WGEKG	---
3it5	80	---	YHMDQ	---
3iuj	170	---	KDVKF	---
3iur	170	---	KDVKF	---
3iwk	447	---	INCAQ	---
3jqy	48	---	VNSKE	---
3jur	120	VYA	LDCEN	VAI
"	196	VQF	YRCRN	VLV
"	219	---	VLSEN	---
"	244	---	ESCKY	---
3k5i	149	---	TMAYD	---
3km5	1315	---	YNCSD	---
3lj5	45	---	LSQYT	---
3ljq	80	---	ACMEH	---
"	380	---	ACMEH	---
3m1h	1584	---	FGCTD	---
3n80	453	TVW	VNCYD	VFG
3n81	453	TVW	VNCYD	VFG
3ogl	146	---	DKCSG	---
3qcw	627	---	ENKAG	---
3rhj	860	---	INTYN	---
3rhl	860	---	INTYN	---
3rho	860	---	INTYN	---
3rhq	860	---	INTYN	---
3riq	281	---	NSANN	---
"	301	---	TLCNH	---
3s2k	871	---	GHLDY	---
3s94	48	---	GGLED	---
"	357	---	EDIRH	---
3suc	222	LSI	RACHN	VYI
"	285	ITT	HHSQY	INI
"	571	---	MGSTS	---
"	587	SGV	LTSSS	SKT
"	617	NTI	LGSLG	AVA
"	632	---	ISASQ	---
"	647	NLI	LSSYG	INT
3th0	265	---	RECIG	---
"	285	FLF	RGCHF	CKM
3ven	11	---	RDFHD	---
3ver	11	---	RDFHD	---
3vex	11	---	RDFND	---
3vez	11	---	RDFHD	---
3vf2	11	---	RDFHD	---
3vf4	11	---	RDFHD	---
3vmv	100	IRL	SNAHN	III
3zqa	441	---	INTWG	---
3zsc	88	LVI	KDAQN	VII
"	122	INV	ENSHH	IWI
4a0p	745	---	KDLDS	---
4a2a	43	---	RGLDE	---
4a2b	43	---	RGLDE	---
4ap6	97	---	IHQVR	---
4dg6	116	---	QELDQ	---
"	160	---	SEIYW	---
4ecd	61	---	TGVRH	---
4fdc	579	---	WNIFN	---
4fr8	453	TVW	VNCYD	VFG
4gdv	93	---	SAVQC	---
4h7n	415	---	INDAA	---
9gaa	80, 380	---	ACMEH	---
9gac	80	---	ACMEH	---
"	380	---	ACMEH	---
9gaf	80	---	ACMEH	---
"	380	---	ACMEH	---

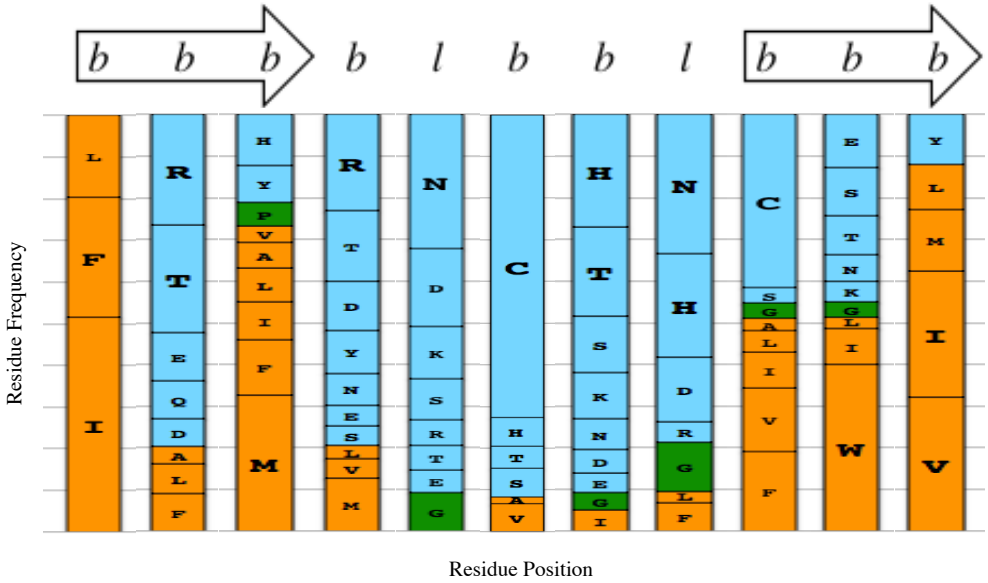

Figure 3.6. Sequence logo for *blbbl* β-arcs, where *b* = β-strand, and *l* = left-handed α-helix backbone dihedral angles. Residues of the β-arc are typically hydrophilic (blue), while residues of flanking β-strands are typically hydrophobic (orange).

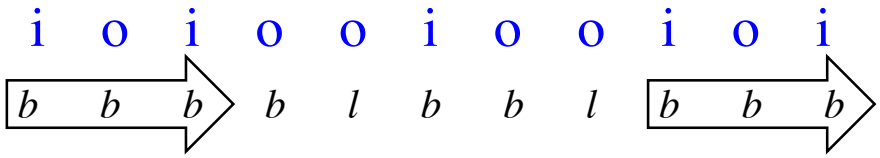

Figure 3.7. Solvent accessibility of residue side-chains in the *blbbl* β-arc. Outside (o) residues are solvent-accessible; inside (i) residues are solvent-inaccessible.

Of note in our *blbbl* β-arc consensus sequence, is the occurrence of L-asparagine (Asn, N), and less commonly L-aspartate (Asp, D), at the two positions that adopt backbone dihedral angles in the left-handed α-helix (*l*) region of the Ramachandran plot. This result is consistent with the observation that asparaginyl residues adopt conformations in partially-allowed regions of the Ramachandran plot more readily than any other non-glycyl amino acids.[61] One theory is that the *l* conformation of an asparaginyl residue is stabilized by non-covalent attractive forces between the dipoles of the asparaginyl residue's side-chain carbonyl, and the backbone carbonyl of either that residue or the previous residue.[62,63]

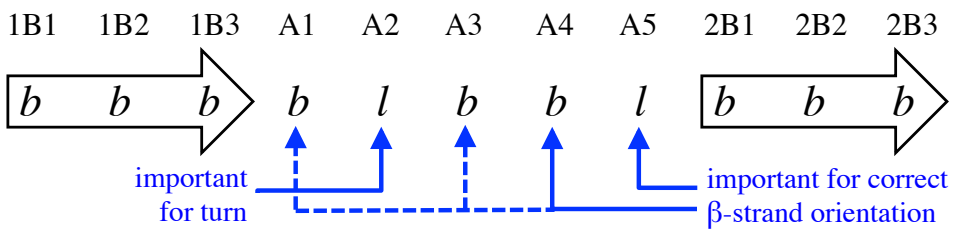

Figure 3.8. Contributions of backbone dihedral angles to *blbbl* β-arc.

Examination of *blbbl* sequences containing no dihedral angle constraints at any one position, *n*, revealed possible contributions of backbone angles to the β-arc (**Figure 3.8**). Many *bnbbl* sequences were parts of extended β-strands, which suggests that the *l* at position "A2" is important for creating a turn. Both *blbnl* sequences were β-turns, and many *blbbn* sequences created turns with flanking β-strands in orientations other than that of a β-arch; these results suggest that the *b* at position "A4" and the *l* at position "A5" are important for creating the distinctive β-strand orientation in a β-arch, which supports interactions between hydrophobic side-chains on facing β-strands. With the importance of the *l* angles at positions "A2" and "A5" established, the culled PDB was searched for *nlnnl* sequences. Not counting *blbbl* sequences, the *nlnnl* search returned 20 sequences, 8/20 of which were 180° β-arcs. The *nlnnl* sequences that were not 180° β-arcs were often β-turns, or β-arcs with a < 180° change of chain direction; this suggests that the *b* angles at positions "A1," "A3," and "A4" contribute to the orientation of the β-strands in a 180° β-arch. However, the collective contribution of the *b* positions of *blbbl* to the orientation of the β-strands is probably not as important as the contribution of the "A5" *l* angle, given that 8/20 *nlnnl* sequences are still 180° β-arcs, while only 3/24 *blbbn* sequences are 180° β-arcs. Our set of 12 *blbbl* β-arches produced a consensus sequence with asparaginyl residues in both of the *l* positions. In contrast, in the consensus sequence for the 12 *blbbl*,

1 *bnbbl*, 3 *blbbn*, and 8 *nlnnl* (including 1 *nlbbl*, 1 *blnbl*, and 1 *blbnl*) β-arches, the most common amino acid at the *l* positions was glycine (Gly).

3.4.2. Peptide Design

The sequence of residues 10-20 of our designed peptide, **3.A** (**Figure 3.9**), is based on our consensus sequence for the *blbbl* β-arch, IleArgMetArgAsnCysHisAsnCysTrpVal (**Figure 3.6**), with the intended β-arc at positions 13-17. At positions 14 and 17, we have promoted *l* dihedral angles with the non-natural residue D-alanine (^{D}Ala). We chose ^{D}Ala because its enantiomer, Ala, is known to favor backbone dihedral angles opposite to *l* (i.e., right-handed α-helical dihedral angles, *a*).[61,64] At position 18, we have employed Phe instead of the consensus Cys, because the residue at this position is usually nonpolar (thus participating in the interdigitation of hydrophobic side-chains in the β-arch), Phe is the second most common amino acid found at this position, and Phe is useful for assignment of ^{1}H NMR spectra. Similarly, Trp was used for position 19, despite the fact that a polar residue usually occupies this position, because Trp is useful for assigning NMR spectra and is the most commonly occurring amino acid at this position. At position 15, we have employed Ser, which is structurally similar to the consensus Cys, to avoid oxidation reactions that can complicate syntheses and characterizations.

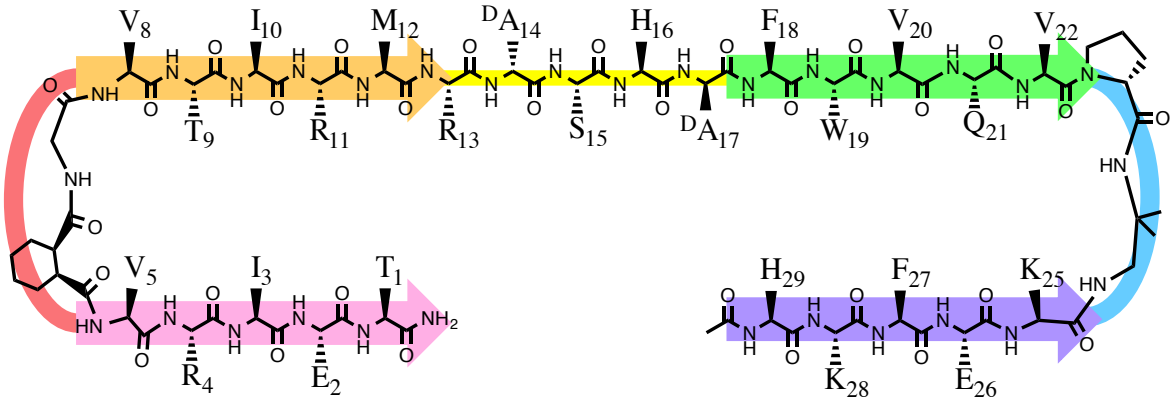

Figure 3.9. Sequence of peptide **3.A.** To facilitate visualization, residues 1-5 (pink), 8-12 (orange), 18-22 (green), and 25-29 (purple) are overlaid with cartoon arrows representing peptide chain direction, while residues 6-7 (red), 13-17 (yellow), and 23-24 (blue) are overlaid with cartoon loops. Control peptides were the following variants of peptide **3.A**: ^{D}A14N, ^{D}A17N (peptide **3.B**); ^{D}A14G, ^{D}A17G (peptide **3.C**); ^{D}A14G, H16G, ^{D}A17G (peptide **3.D**); R13G, ^{D}A14G, S15G, H16G, ^{D}A17G (peptide **3.E**); ^{D}A14A, ^{D}A17A (peptide **3.F**); and T1_G7del, acetyl-capped V8 (peptide **3.G**).

The identities of the remaining residues were chosen with several considerations in mind. First, they were chosen to promote parallel β-sheet folding, based on intrinsic propensities of amino acids to assume β-strand conformations[65,66] and amino acid pairing preferences between adjacent strands of parallel β-sheets[67]. We also used the following non-natural linkers to promote parallel β-sheet interactions: ^{D}Pro-DADME (**Figure 3.9** blue), a diamine linker derived from D-proline (^{D}Pro) and 2-methylpropane-1,2-diamine (a.k.a. 1,2-diamino-1,1-dimethylethane, or DADME), and (1*R*,2*S*)-CHDA-Gly (**Figure 3.9** red), a diacid linker derived from (1*R*,2*S*)-1,2-cyclohexanedicarboxylic acid ((1*R*,2*S*)-CHDA) and glycine (Gly).[68-72] Second, charged residues were incorporated, such that each peptide carried a net positive charge at the conditions studied, to promote water-solubility and prevent the aggregation that makes β-sheets notoriously difficult to isolate and study.[73] A positive net charge was chosen, because positively charged side-chains have slightly higher β-strand propensities than do negatively-charged side-chains.[66] Third, the peptide included as much residue diversity as was allowed by the aforementioned considerations,

to facilitate NMR interpretation. Finally, given the difficulty of synthesis of parallel β-sheet model peptides of >2 strands (**Figure 3.2**), amino acids that are known to add difficulty to peptide syntheses (e.g., Asp, Cys)[74] were avoided, if this avoidance did not significantly affect other design considerations.

We also designed variants of peptide **3.A** (**Figure 3.9**), to assess for the contributions of various residues to the fold of this peptide. To assess the contribution of the non-natural β-arc linker to the peptide fold, we designed the following peptides: **3.B** (^{D}A14N, ^{D}A17N), **3.C** (^{D}A14G, ^{D}A17G), **3.D** (^{D}A14G, H16G, ^{D}A17G), **3.E** (R13G, ^{D}A14G, S15G, H16G, ^{D}A17G), and **3.F** (^{D}A14A, ^{D}A17A). Peptides **3.B-3.F** substitute out residues of the designed β-arc, two to five residues at a time, and substitute in more flexible residues (Gly, Asn) or residues of the opposite stereochemistry (^{D}Ala to Ala). To assess the contributions of hydrophobic residues to the peptide fold, we designed peptide **3.G** (T1_G7del, acetyl-capped V8). Peptide **3.G** deletes a strand of peptide **3.A** that contains two aliphatic, hydrophobic residues that appear to be buried by the tertiary fold of peptide **3.A** (**Figure 3.17** pink strand, discussed in **section 3.4.3**).

3.4.3. Characterization of Peptide Monomers in Aqueous Solution

Peptide **3.A** behaved like a monomeric species in aqueous solution at low salt concentrations (0.02-0.2 mM peptide, 9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8), based on NMR diffusion experiments (**Table 3.3**) and sedimentation equilibrium analytical ultracentrifugation (AUC) experiments (**Figure 3.10**, **Table 3.2**).

AUC experiments were designed to examine the association state of peptide **3.A**, at low-salt buffer conditions similar to those employed for NMR structural studies. Plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center

of rotation were linear, at all speeds and for all loading concentrations (**Figure 3.10**), indicating the presence of a single molecular weight entity. For peptide **3.A**, the weight-average molecular weight (MW_{obs}) was much smaller than the molecular weight based on the sequence (M_s). Furthermore, the apparent weight decreased with increasing concentration. These results suggest that at the low-salt conditions studied, the peptide is monomeric, and not forming dimers or larger molecular weight aggregates. We conclude that the observed $MW_{obs}/MW_{calc} \leq 1$ (**Table 3.2**) indicates a monomeric state, with electrostatic nonideality accounting for $MW_{obs}/MW_{calc} << 1$.

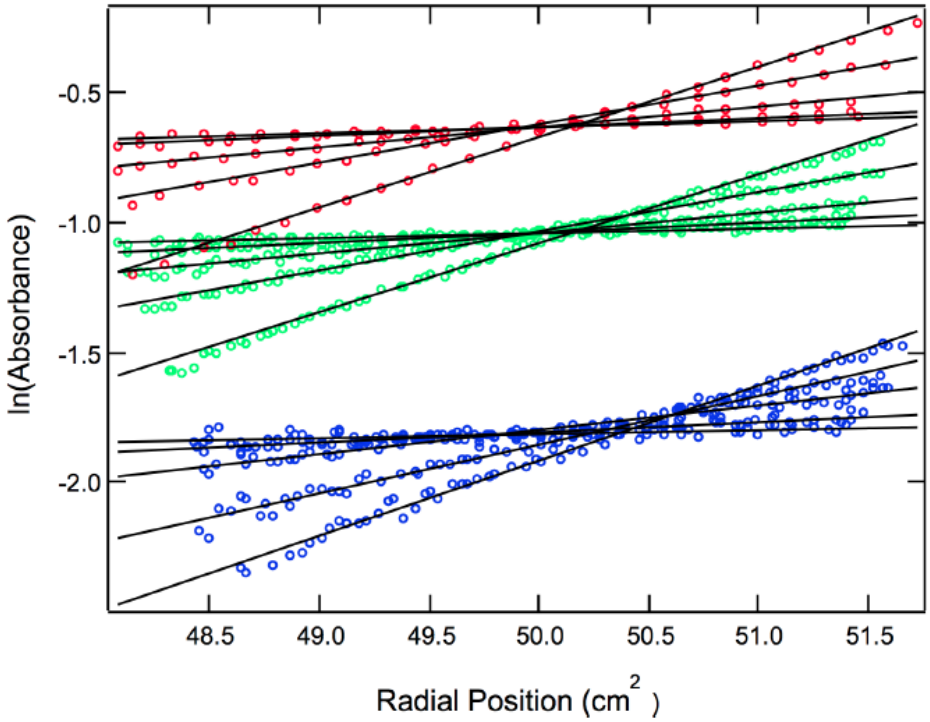

Figure 3.10. Equilibrium AUC data for peptide **3.A** in 3 mM NaOAc, pH 3.8. The data shown are for speeds of 10k, 15k, 23k, 32k, 42k, at initial loading concentrations of 0.03 mM (blue), 0.06 mM (green), and 0.09 mM (red).

Table 3.2. Sedimentation equilibrium results for peptide **3.A**. Weight-average molecular weight (MW_{obs}) from fits to a single-species model, reported with least-squares fitting error in the fitted parameter.

peptide	**buffer**	**MW_{obs}/M_S**	**MW_{obs}** (Da)
3.A	3 mM NaOAc, pH 3.8	0.693 ± 0.004	2417

The hydrodynamic radii derived from NMR diffusion experiments (obs R_H) for peptides **3.A-3.G** are reported in **Table 3.3**. The predicted hydrodynamic radii of the monomeric peptides were calculated using the atomic-level (a = 2.84 Å) shell-model of HYDROPRO 10,[75] and are in good agreement with the hydrodynamic radii derived from NMR diffusion data at low salt conditions (**Table 3.3**). Furthermore, variable concentration NMR diffusion analysis of the peptides at low salt conditions shows little to no increase in R_H values as peptide concentrations are increased from 0.02 mM to 0.2 mM. Also, no changes in chemical shifts, peak shapes, or splitting were observed on 1D ^{1}H NMR at low salt conditions, as peptide concentrations were increased from 0.02 mM to 0.2 mM. These results support the conclusion that peptides are monomeric (no peptide aggregation occurs) at the low salt conditions used to obtain NMR structures of the peptides.

Table 3.3. Hydrodynamic radii of peptides, derived from NMR diffusion experiments (obs R_H) or HYDROPRO 10 calculations for monomeric peptides (calc R_H).

peptide	buffer	obs R_H (Å) at 0.2 mM peptide	obs R_H (Å) at 0.02 mM peptide	calc R_H (Å)
3.A	3 mM NaOAc, pH 3.8	9.3 ± 0.6	8.3 ± 1.7	13.5
	3 mM NaOAc, 100 mM NaCl, pH 3.8	25.4 ± 1.5	---	---
3.B	3 mM NaOAc, pH 3.8	14.0 ± 1.1	---	---
3.C	3 mM NaOAc, pH 3.8	11.4 ± 0.4	8.4 ± 1.3	---
3.D	3 mM NaOAc, pH 3.8	8.6 ± 0.4	12.8 ± 1.4	---
3.E	3 mM NaOAc, pH 3.8	12.6 ± 0.5	7.3 ± 1.7	---
3.F	3 mM NaOAc, pH 3.8	17.6 ± 0.4	---	---
3.G	3 mM NaOAc, pH 3.8	9.8 ± 0.4	9.5 ± 0.8	15.2

The structure of monomeric peptide **3.A** was examined using NMR and circular dichroism (CD) spectroscopy. Variable temperature NMR studies suggested peptide unfolding as temperature was increased, with less peak dispersion (**Figure 3.11**), less downfield HN and Hα chemical shifts, and fewer NOEs observed at higher temperatures. We therefore performed NMR studies at 5 °C. Assignment of almost all proton NMR signals (**Table 3.4A**) was achieved with two-dimensional COSY, TOCSY, and NOESY spectra

obtained in aqueous solution (0.2 mM peptide, 9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8). A downfield pattern was observed for secondary chemical shifts of α-protons ($\Delta\delta_{\alpha H}$), consistent with β-strand structure (**Figure 3.12A**, **Figure 3.13** red).[37,42] Likewise on CD spectroscopy at these low salt conditions, a modest-intensity minimum was observed at ~216 nm, consistent with parallel β-sheet structure (**Figure 3.19** red).[27] The observed NOEs (**Figure 3.14A**) support an NMR structure for the monomer with interdigitation of hydrophobic residues in a β-arch-like fold (**Figure 3.17**).

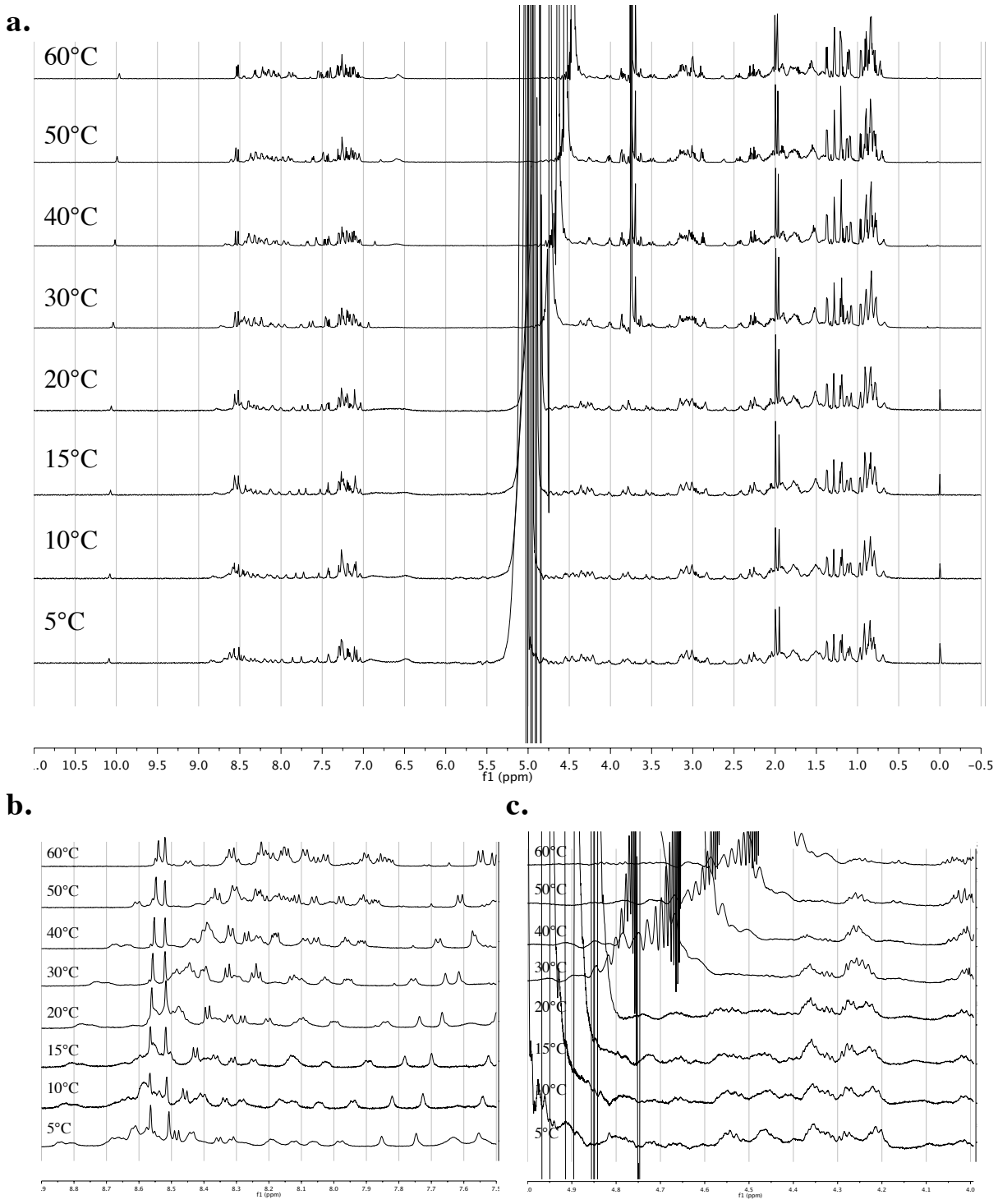

Figure 3.11. Increased ^{1}H NMR peak dispersion as temperature is decreased from 60 °C to 5 °C, for 0.2 mM peptide **3.A** in aqueous solution (3 mM NaOAc-d_3, pH 3.8, ± DSS or dioxane). Panels (b) and (c) show the Hα and HN regions, respectively.

The NMR and CD structural data for monomeric peptides **3.B-3.F** (**Figures 3.13, 3.19**) and **3.G** (**Table 3.4G**, **Figures 3.12G, 3.13, 3.14G, 3.19**) are presented here, and discussed in **sections 3.4.5** and **3.4.6**, respectively.

Table 3.4. Proton resonances (ppm) for peptides **3.A** and **3.G**.

A.

		NH	αH	βH		γH		δH		εH		others				
1	**Thr**	8.41	4.35	3.81		1.20										
2	**Glu**	8.51	4.34	1.65	1.37	2.82	2.82									
3	**Ile**	8.62	4.53	2.03		1.28	1.09					0.89 (γ^2H)				
4	**Arg**	8.69	4.62	1.75	1.75	1.55	1.55	3.09	3.09			7.21 (εH)				
5	**Val**	8.38	4.37			0.87	0.87									
6	**CHDA**	3.07 (αH), 2.93 (βH), 2.62 (αH), 2.07 (βH), 1.93 (βH), 1.83 (βH), 1.80 (γH), 1.62 (γH), 1.39 (γH), 1.27 (γH)														
7	**Gly**	8.20	4.84,4.04													
8	**Val**	8.12	4.22	2.19		0.96	0.96									
9	**Thr**	8.68	4.80	4.02		1.09										
10	**Ile**	8.82	4.36	1.79								0.81 (γ^2H)				
11	**Arg**	8.55	4.42	1.77	1.77	1.54	1.54	3.13	3.13			7.25 (εH)	7.11 (εH)			
12	**Met**	8.85	4.56	2.09	1.89	2.32	2.32									
13	**Arg**	8.36	4.20	1.17	1.17	1.45	1.45	2.84	2.84			7.54 (εH)				
14	**D-Ala**	8.45	4.38	1.21												
15	**Ser**	8.41	4.32	3.79	3.79											
16	**His**	8.66	4.67	3.31	3.16			7.27 (δ^2H)				8.56 (ε^1H)				
17	**D-Ala**	8.19	4.23	1.12												
18	**Phe**	8.32	4.78	3.09	2.88			7.29		7.75		7.61 (zH)				
19	**Trp**	8.45	4.72	3.12	2.94			7.11 (δ^1H)				7.56 (ε^3H)	7.17 (z^3H)	7.42 (h^2H)	7.05 (z^2H)	10.09 (ε^1H)
20	**Val**	8.24	4.10	1.91		0.87	0.87									
21	**Gln**	8.58	4.33	2.25	1.93	3.53	3.53			7.21	7.86					
22	**Val**	7.98	4.41	1.95		0.84	0.84									
23	**D-Pro**		4.28	2.24	2.24	2.02	2.02	3.79	3.79							
24	**DADME**	8.06 (NH), 3.56 (αH), 3.52 (αH), 1.31 (CH_3)														
25	**Lys**	8.62	4.49	2.42	2.02	1.53	1.53	1.78	1.78	3.02	3.02	7.63 (zH)				
26	**Glu**	8.56	4.51													
27	**Phe**	8.49	4.55	2.98	2.98			6.50		7.25		7.07 (zH)				
28	**Lys**	8.58	4.46	1.84	1.84	1.41	1.41	1.73	1.53	3.02	3.02	7.63 (zH)				
29	**His**	8.44	4.90	3.10	2.95			7.19 (δ^2H)				8.51 (ε^1H)				
	Ac		1.78													

G.

		NH	αH	βH		γH		δH		εH		others				
	Ac		1.52													
8	**Val**	8.58	4.51	2.03		0.93	0.93									
9	**Thr**	8.33	4.24	3.78		1.16										
10	**Ile**	8.39	4.14	1.82		1.17						0.84 (γ^2H)				
11	**Arg**	8.33	4.34	1.18	1.18	1.48	1.48	2.83	2.83			7.54 (εH)				
12	**Met**	8.65	4.66	3.12	3.12	3.25	3.25									
13	**Arg**	8.55	4.56	1.85	1.85	1.61	1.61	2.93	2.93			7.58 (εH)				
14	**D-Ala**	8.37	4.40	1.17												
15	**Ser**	8.27	4.35	3.83	3.83											
16	**His**	8.35	4.69	3.12	3.12			7.25 (δ^2H)				8.55 (ε^1H)				
17	**D-Ala**	8.29	4.23	1.12												
18	**Phe**	8.45	4.56	3.02	3.02			7.11		7.25						
19	**Trp**	8.30	4.71	3.07	2.87			7.11 (δ^1H)						7.15 (h^2H)	7.43 (z^2H)	10.10 (ε^1H)
20	**Val**	8.18	4.08	1.90		0.83	0.83									
21	**Gln**	8.52	4.26	2.31	2.31	1.30	1.19			7.84	7.84					
22	**Val**	8.32	4.13	2.05		0.95	0.95									
23	**D-Pro**		4.37	2.24	2.24	2.05	1.93	3.78	3.78							
24	**DADME**	8.06 (NH), 3.52 (αH)														
25	**Lys**	8.56	4.47	2.24	2.24	1.53	1.43	1.80	1.73	3.01	3.01	7.64 (zH)				
26	**Glu**	8.53	4.45	1.50	1.50	1.80	1.71									
27	**Phe**	8.44	4.90	2.98	2.98			7.04		7.41		7.54 (zH)				
28	**Lys**	8.47	4.25	2.44	2.44	1.76	1.76	1.55	1.55	3.10	3.10	7.19 (zH)				
29	**His**	8.45	4.43	3.07	2.95			7.09 (δ^2H)				8.50 (ε^1H)				
	Ac		1.36													

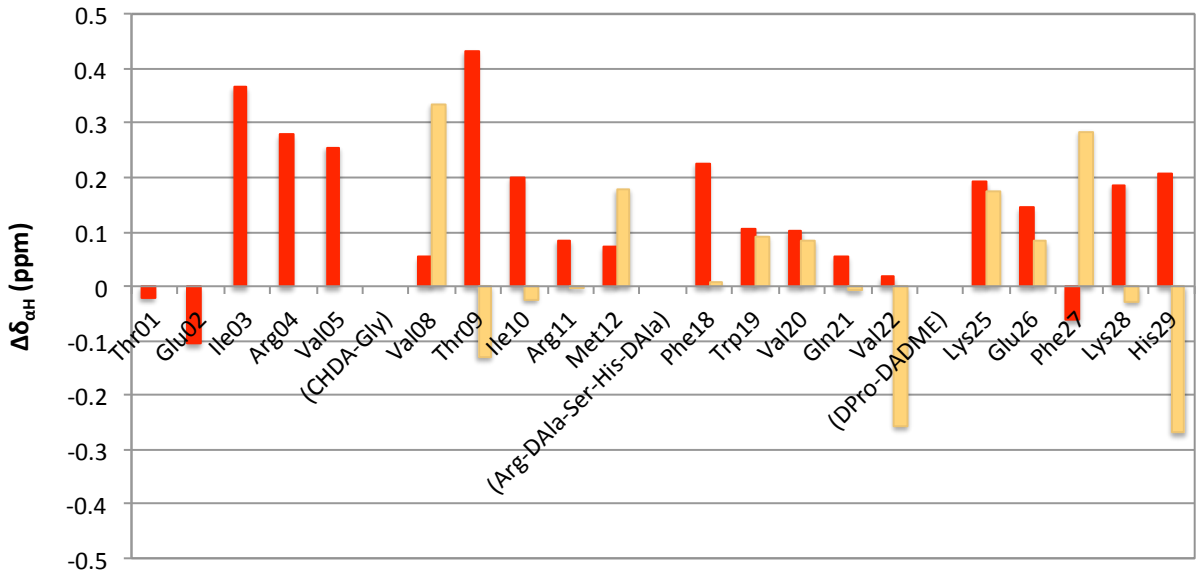

Figure 3.12. Secondary α-proton chemical shifts of peptides **3.A** (red) and **3.G** (orange). NMR spectra were taken at 5.0 °C, with 0.2 mM peptide, in 9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8. Random coil values were calculated using CSDb[2].

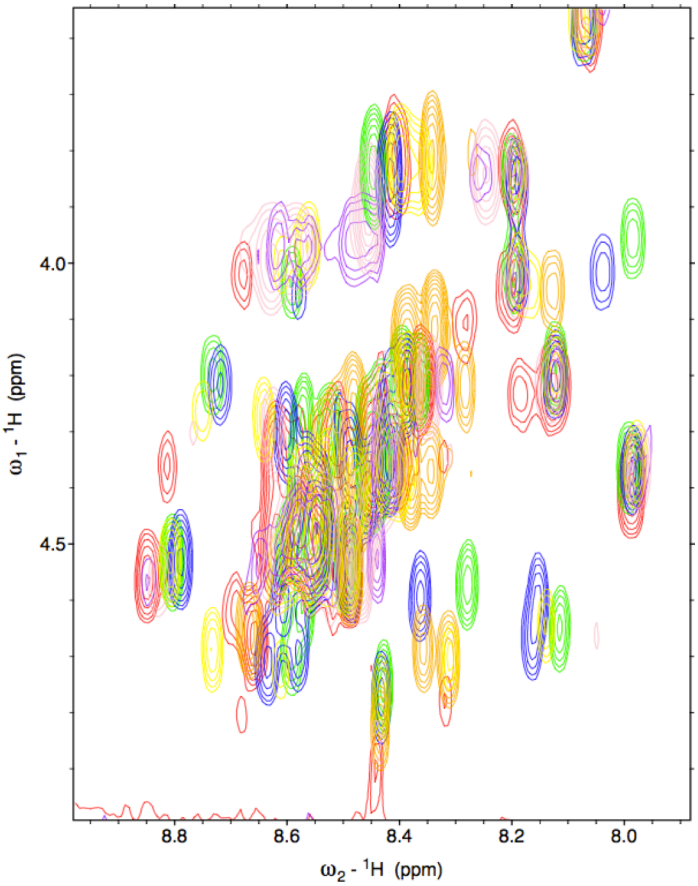

Figure 3.13. Overlay of Hα-HN regions of 2-D TOCSY spectra, for peptides **3.A** (red), **3.D** (pink), **3.E** (purple), **3.B** (blue), **3.C** (yellow), **3.F** (green), and **3.G** (orange). More peak dispersion and more downfield chemical shifts are observed for Hα and HN of **3.A,** than of other peptides. Spectra were taken at 5.0 °C, with 0.2 mM peptide, in 9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8.

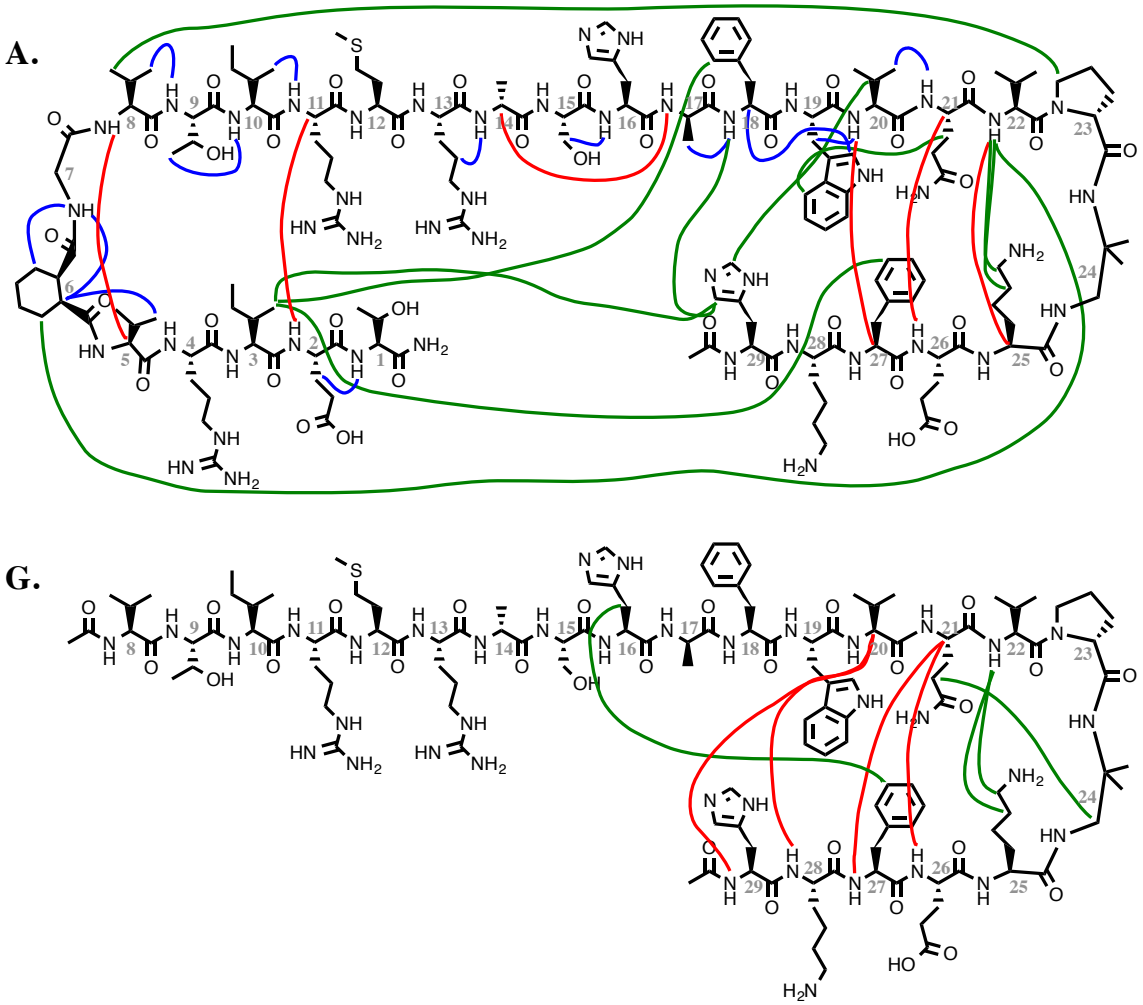

Figure 3.14. NOE's observed in peptides **3.A** and **3.G**, in backbone (red), to side-chains of sequential residues (blue), and to side-chains of non-sequential residues (green) of peptide. For clarity, intra-residue and sequential backbone NOE's are not depicted.

Table 3.5. Structural statistics for peptide **3.A,** 10 (of 100 calculated) structures with the overall lowest energies. Reported statistics are for all (not just ordered) residues.

Conformationally restricting restraints	Experimental ^{1}H-^{1}H distance restraints	
	Intraresidue (i = j)	154
	Sequential (|i –j| = 1)	35
	Medium range (1 < |i –j| < 5)	6
	Long range |i –j| ≥ 5	10
	Hydrogen-bond restraints	10
	Dihedral angle restraints	12
Residual restraint violations	Ave no. of NOE restraint violations per structure (>0.5 Å)	1.8
	Ave no. of H-bond violations per structure (>0.5 Å)	0.2
	Ave no. of dihedral angle violations per structure (>10º)	0
Model quality	RMSD backbone atoms (Å)	2.1
	RMSD heavy atoms (Å)	2.9
	RMSD bond lengths (Å)	0.012
	RMSD bond angles (º)	1.4
MolProbity Ramachandran statistics	Most favored regions (%)	61.0
	Allowed regions (%)	22.9
	Disallowed regions (%)	16.2
Global quality scores (mean/Z score)	Verify3D	-0.01/ -7.54
	ProsaII	-1.21/ -7.69
	PROCHECK (**ϕ-ψ**)	-1.70/ -6.37
	PROCHECK (all)	-1.11/ -6.56
	MolProbity clash score	82.67/-12.66
Model contents	BMRB accession number	25982
	PDB ID code	2NBL

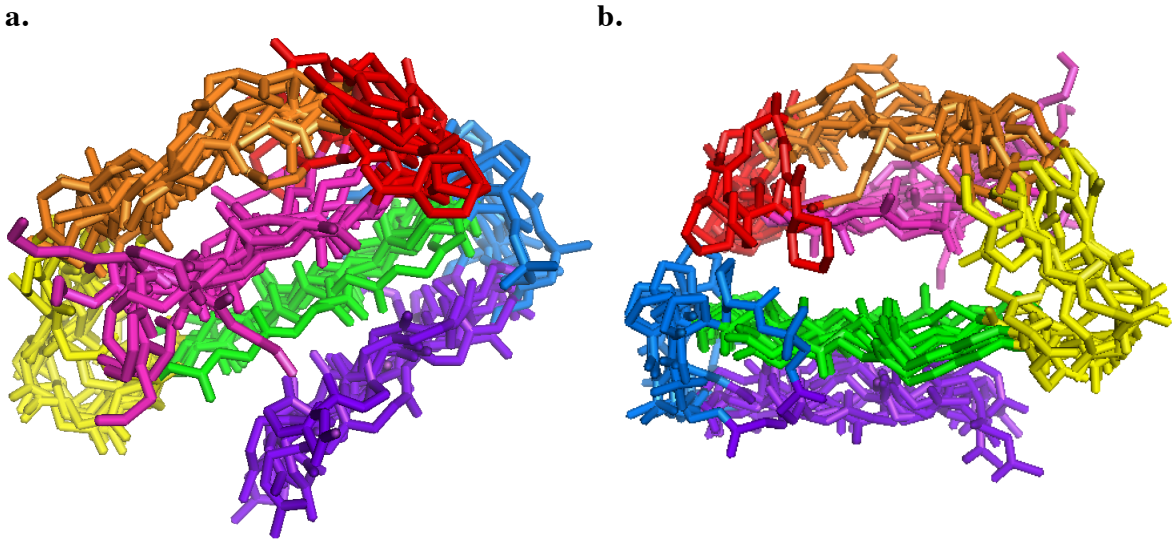

Figure 3.15. (a) Top view and (b) back side view, of structure ensemble of 10 lowest energy (out of 100 calculated) NMR structures, for peptide **3.A**. For clarity, only backbone atoms are shown, and the peptide is color-coded with residues 1-5 in pink, 6-7 in red, 8-12 in orange, 13-17 in yellow, 18-22 in green, 23-24 in blue, and 25-29 in purple.

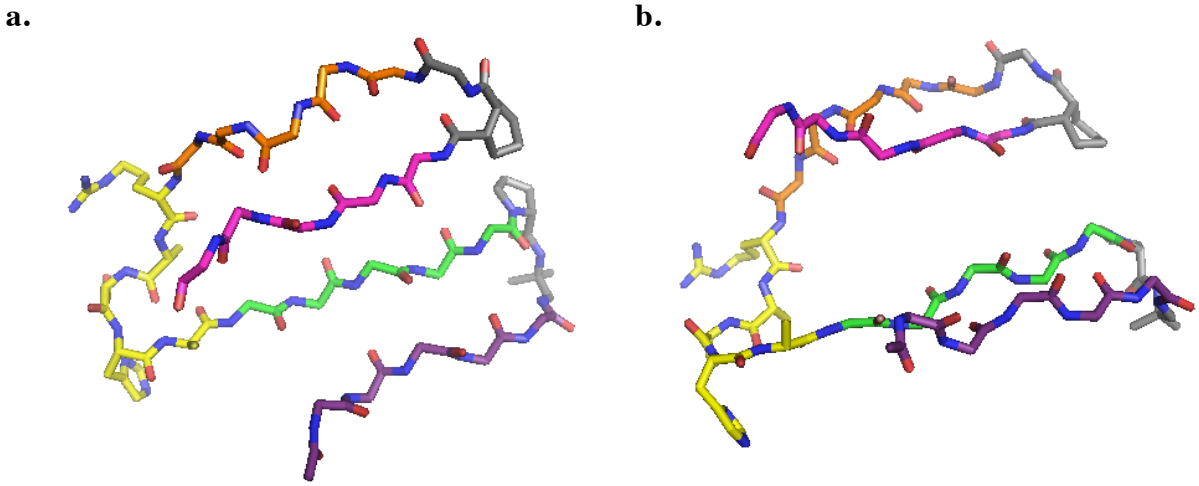

Figure 3.16. (**a**) Top view and (**b**) side view of average of 10 lowest energy (out of 100 calculated) NMR structures, for peptide **3.A**. For clarity, only backbone atoms are shown for the following residues, which are color-coded: residues 1-5 in pink, 6-7 in red, 8-12 in orange, 13-17 in yellow, 18-22 in green, 23-24 in blue, and 25-29 in purple.

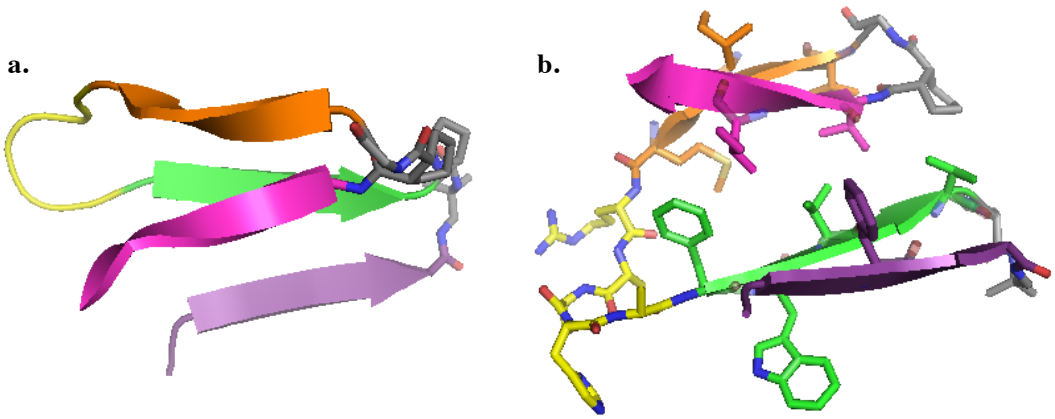

Figure 3.17. (**a**) Top view and (**b**) side view of average of 10 lowest energy (out of 100 calculated) NMR structures, for peptide **3.A**. To facilitate visualization, residues 1-5 (pink), 8-12 (orange), 18-22 (green), and 25-29 (purple) are depicted as cartoon arrows representing backbone atom positions, but not necessarily canonical β-strand structures. In (b), side-chains of hydrophobic residues and linker regions (β-arc in yellow, diacid and diamine linkers in gray) are shown.

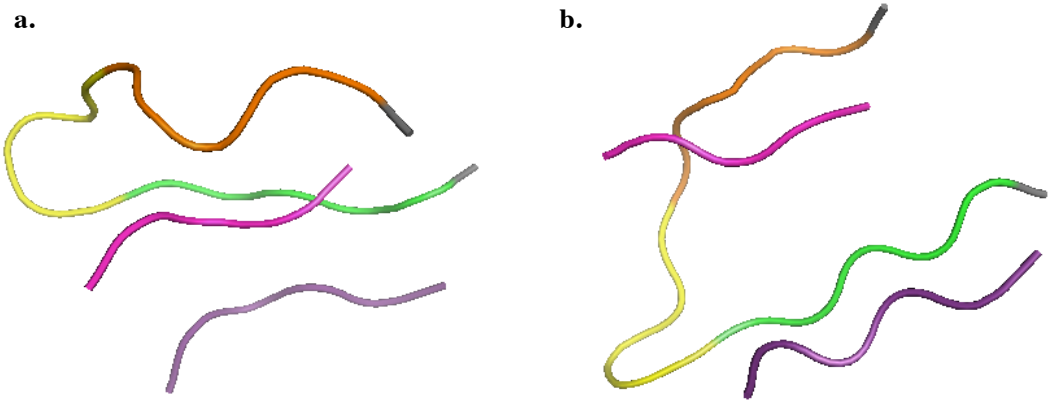

Figure 3.18. (**a**) Top view and (**b**) side view of average of 10 lowest energy (out of 100 calculated) NMR structures, for peptide **3.A**, using so-called "tube" depictions of backbone atom positions. Non-natural residues DADME and CHDA are not shown. For clarity, residues are color-coded: residues 1-5 in pink, 6-7 in red, 8-12 in orange, 13-17 in yellow, 18-22 in green, 23-24 in blue, and 25-29 in purple.

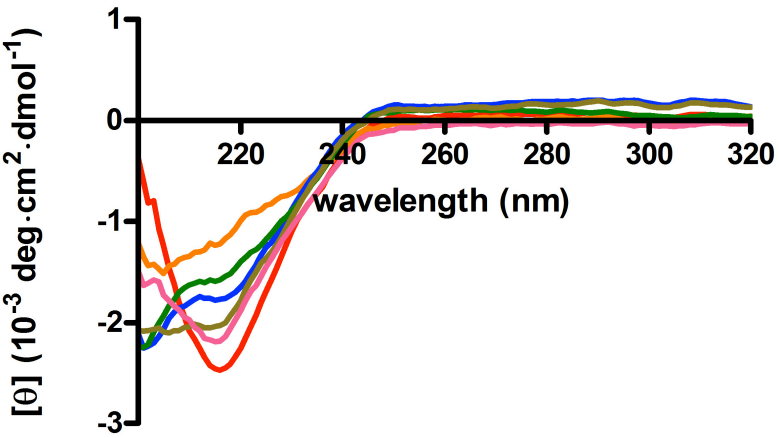

Figure 3.19. Mean residue ellipticity of peptides **3.A** (red), **3.D** (pink), **3.B** (blue), **3.C** (yellow), **3.F** (green), and **3.G** (orange), at 20.0 °C (293.2 K). Spectra are reported for 0.2 mM peptide samples in 3 mM NaOAc, pH 3.8.

3.4.4. Characterization of Peptide Oligomers and Protofibrils

Imaging of peptide **3.A** by transmission electron microscopy (TEM) with negative staining, after incubation at 0.08 mM concentration in aqueous solution with low salt content (9:1 H_2O:D_2O, 3 mM NaOAc-d_3, pH 3.8) at 4 °C for 1-3 hours, revealed that peptide **3.A** can assume oligomeric structures **Figure 3.21**), similar in appearance to early stage amyloid oligomers and protofibrils.[11,15,76,77] Images suggest that the protofibrils grew out of the spherical oligomers (**Figure 3.21a**). For control peptides **3.B**-**3.G**, unstructured aggregates and/or spherical oligomers were observed, while protofibrils were infrequently observed (**Figures 3.22**-**3.27**).

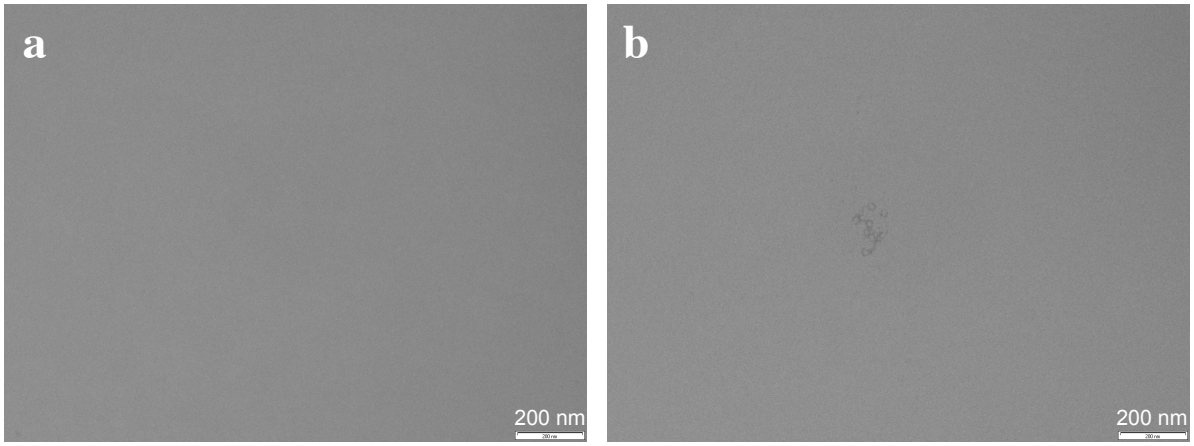

Figure 3.20. TEM of buffer (3 mM NaOAc, pH 3.8) showed (b) some salt artifacts, but otherwise (a, b) clean samples, without the structures attributed to peptide in **Figures 3.21-3.27**.

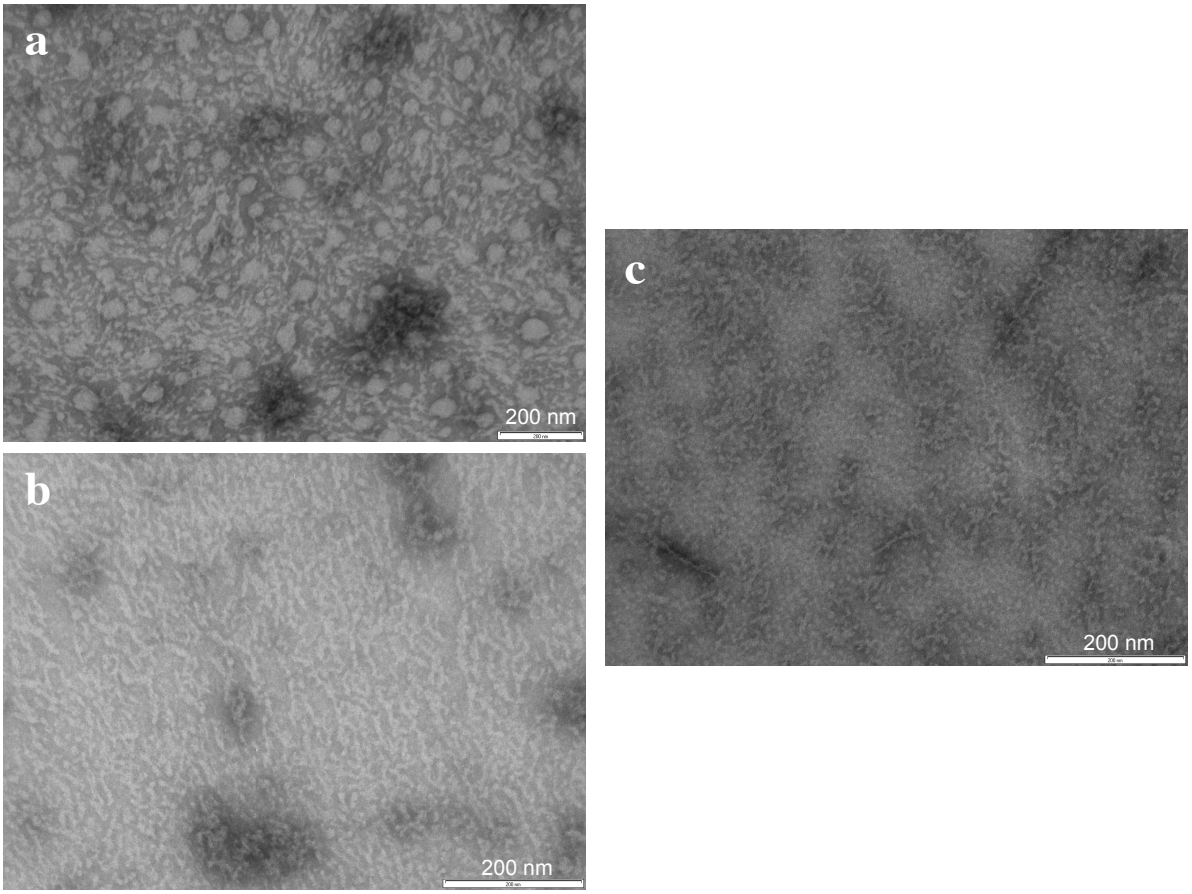

Figure 3.21. TEM of peptide **3.A,** after (a, b) 1-3 hour and (c) 1 day incubations in aqueous solution. Appearance after 2 and 3 day incubations (not shown) did not differ significantly. Spherical oligomers (a) had diameters of ~30-50 nm, and shrank with the appearance of protofibrils. Protofibrils (a, b, c) had widths of ~10 nm, and lengths of ~50-120 nm.

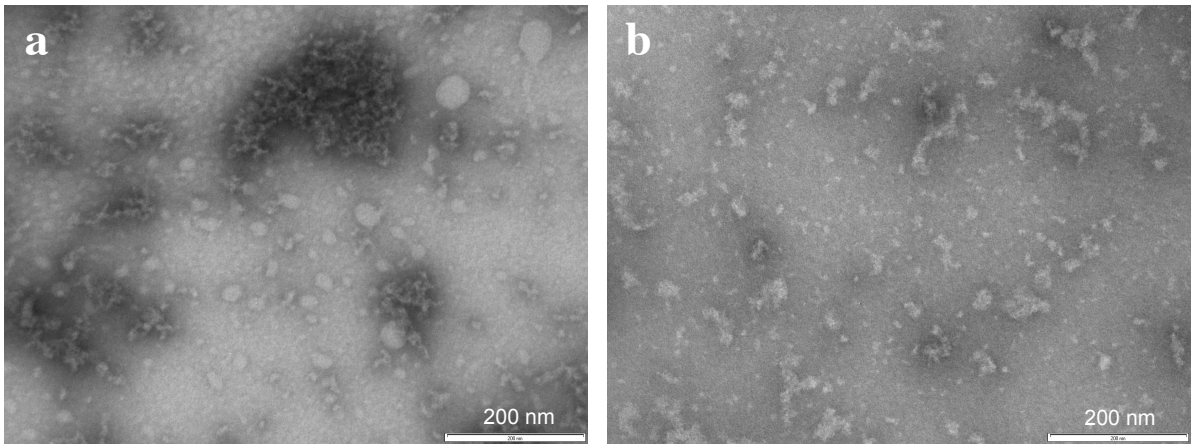

Figure 3.22. TEM of peptide **3.B**, after (a) 1-3 hour, and (b) 2 day incubations in aqueous solution. In general, only unstructured aggregates (a, b) were observed. Few spherical oligomers (a), of diameter ~30 nm, were also observed.

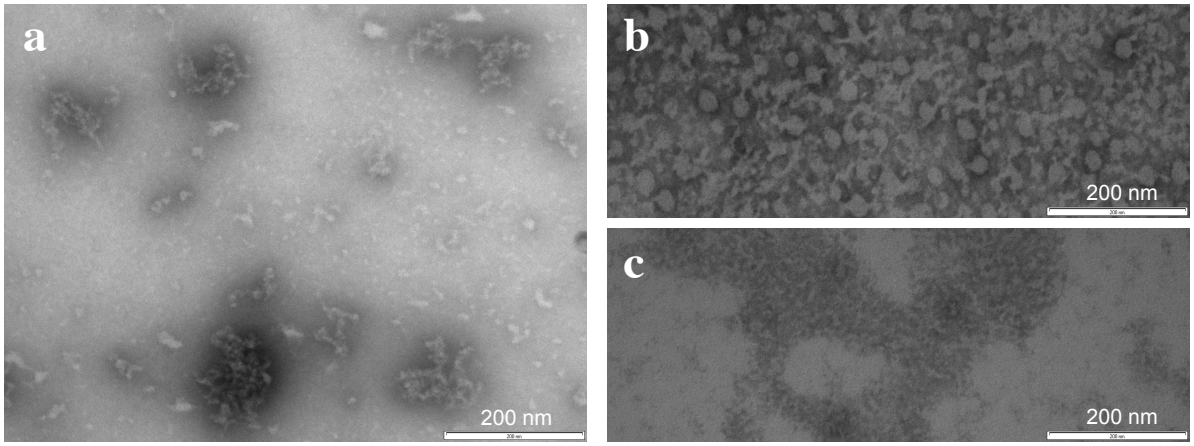

Figure 3.23. TEM of peptide **3.C**, after (a) 1-3 hour, and (b, c) 4 day incubations in aqueous solution. In general, only unstructured aggregates (a, c) were observed. The structures most resembling spherical oligomers (b) had diameter ~20-30 nm.

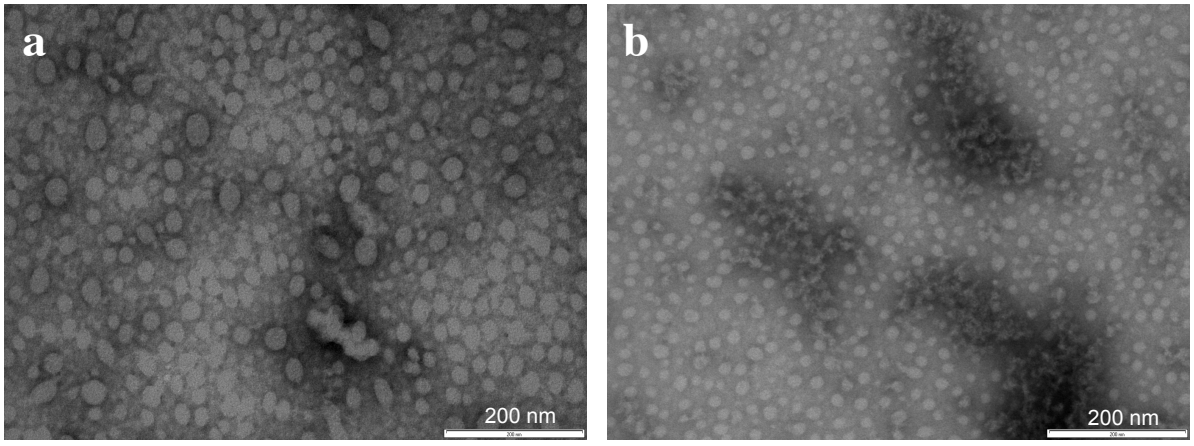

Figure 3.24. TEM of peptide **3.D**, after (a) 1-3 hour, and (b) 4 day incubations in aqueous solution. Small spherical oligomers (a, b), of ~20 nm diameter, were observed.

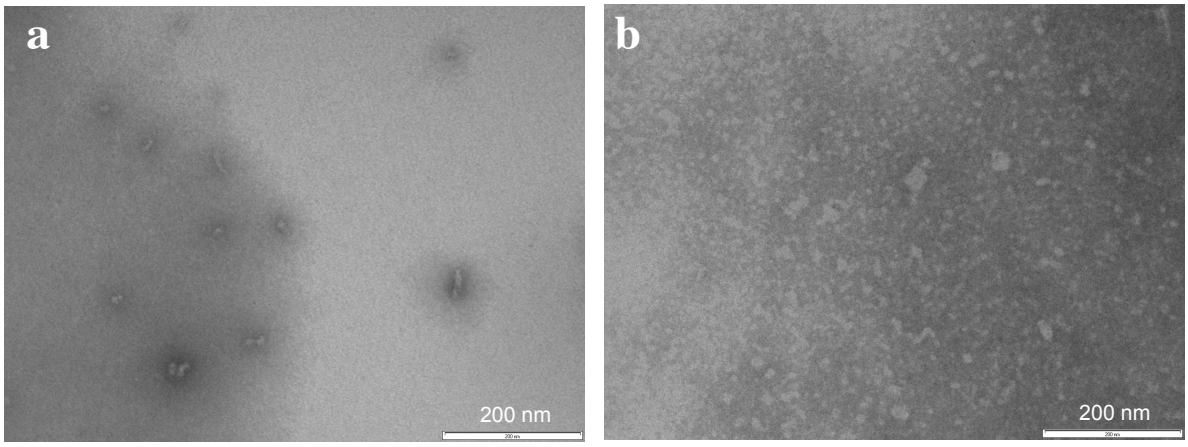

Figure 3.25. TEM of peptide **3.E**, after (a) 1-3 hour, and (b) 6 day incubations in aqueous solution. In general, only sparse, unstructured aggregates were observed (a, b).

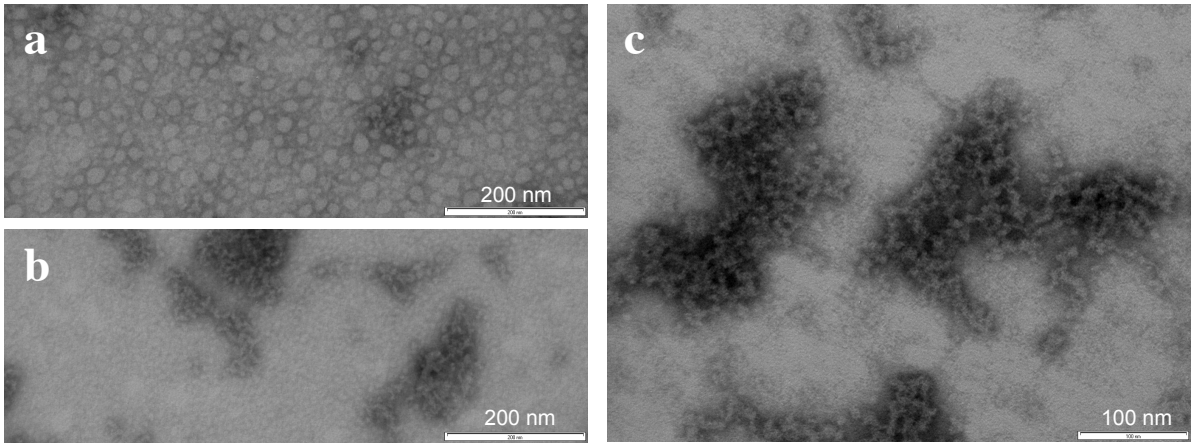

Figure 3.26. TEM of peptide **3.F**, after (a, b) 1-3 hour, and (c) 2 day incubations in aqueous solution. In general, only unstructured aggregates (b, c) were observed. Some spherical oligomers (a) were also observed, though they were small (diameter ~20 nm).

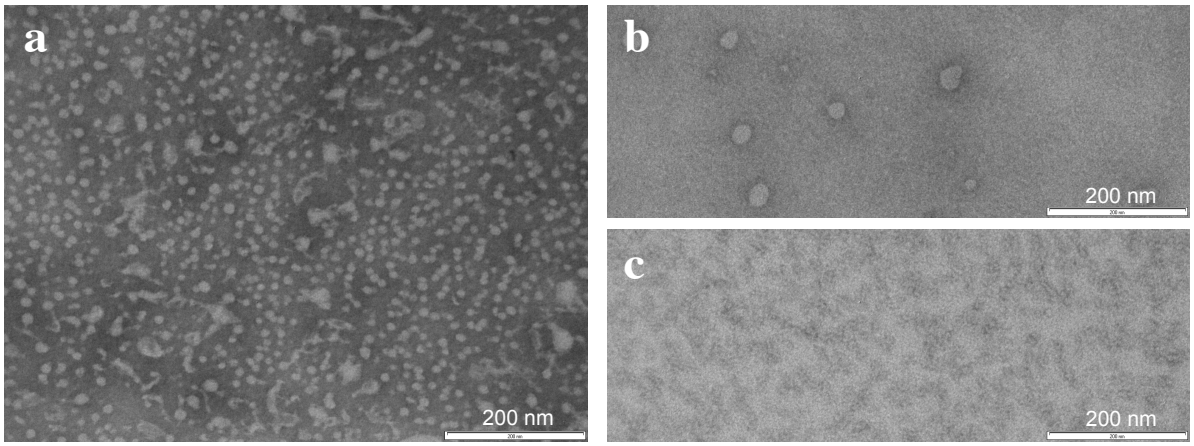

Figure 3.27. TEM of peptide **3.G**, after (a) 1-3 hour, and (b, c) 2 day incubations in aqueous solution. In general, only unstructured aggregates (c) were observed. The structures most resembling spherical oligomers (a, b) were small (diameter ~10-20 nm) and few.

Since large oligomer populations were not detectable in aqueous solution with low salt concentrations (**Figure 3.10**, **Table 3.3**), it is likely that oligomers observed by TEM were artifacts of TEM sample preparation (e.g., increased local concentrations of peptide on drying). To study the oligomeric structures in aqueous solution, we induced oligomerization using a charge-shielding salt, NaCl (9:1 H_2O:D_2O, 3 mM NaOAc-d_3, 100 mM NaCl, pH 3.8). Peptide oligomerization at high salt concentrations was confirmed with NMR diffusion and AUC experiments. At high salt (100 mM NaCl) conditions, large R_H values were observed for peptide **3.A** (**Table 3.3**), suggesting oligomerization of the peptide. Similarly, AUC studies were performed on peptide **3.A** in buffer containing 100 mM neutral salt (NaCl). The resulting plots of the logarithm of the measured absorbance as a function of the squared radial distance from the center of rotation were nonlinear, suggesting the presence of multiple molecular weight entities and a self-associating system. Data could not be fit to one macromolecular species (with the reduced molecular weight of the monomer or dimer) or two macromolecular species (with the reduced molecular weights of the monomer and dimer), supporting the presence of larger oligomers. The presence of large oligomers was also supported

by the observation of peptide precipitation at higher rotor speeds.

Compared to the minimum in the CD wavelength scan observed for monomeric peptide **3.A** (**Figure 3.19** red), the CD minimum of oligomeric peptide **3.A** (**Figure 3.29** red) was shifted to the slightly higher wavelength of ~220 nm, and had a greater intensity (**Figure 3.28**). A strong minimum at ~220 nm is consistent with the CD signature of amyloids.[78-80] Thermal denaturation of the oligomeric structure was monitored by CD, and showed a reversible sigmoidal transition, which is evidence for a cooperative process (**Figure 3.30**). Likewise for peptides **3.B-3.F** in the presence of 100 mM NaCl, single minima in the CD signal were observed near 220 nm (**Figure 3.29**), and the CD signal at 220 nm was monitored as a function of temperature for thermal denaturation experiments (**Figure 3.30**).

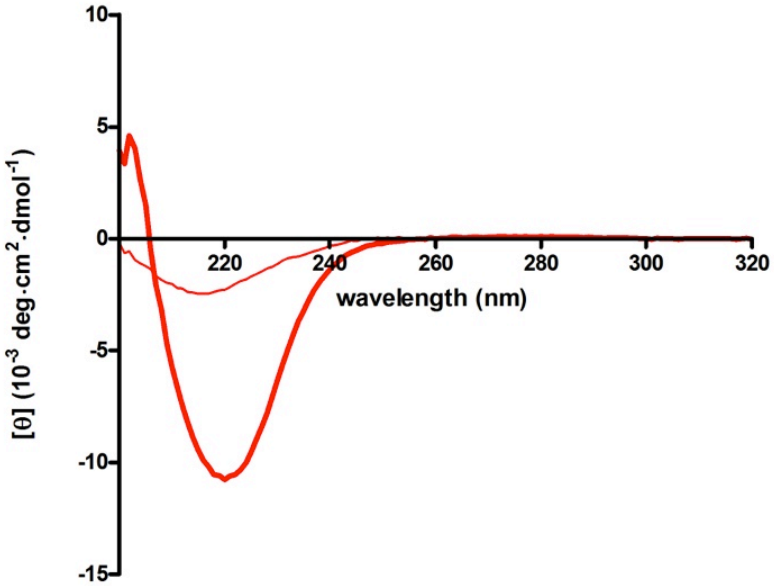

Figure 3.28. Mean residue ellipticity of 0.2 mM peptide **3.A**, at 20.0 °C (293.2 K), in 3 mM NaOAc, pH 3.8 (thin line), or 3 mM NaOAc, 100 mM NaCl, pH 3.8 (thick line).

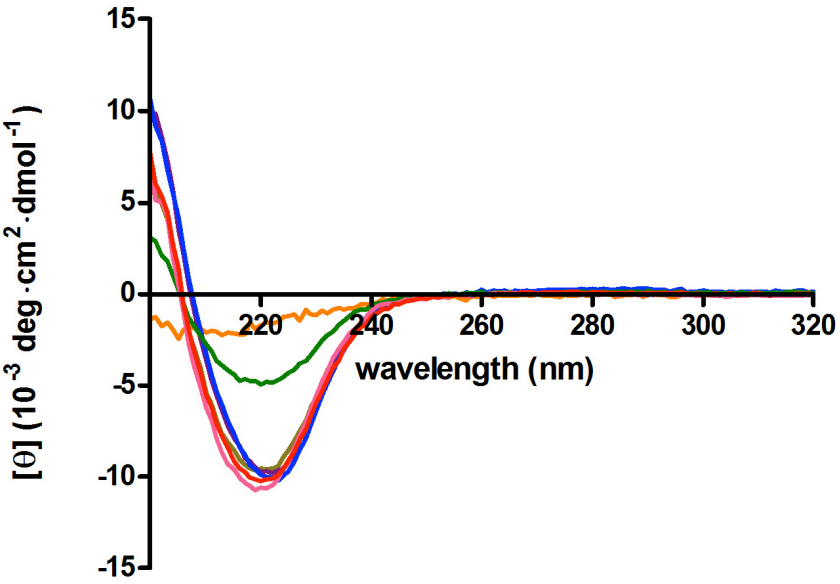

Figure 3.29. Mean residue ellipticity of peptides **3.A** (red), **3.D** (pink), **3.E** (purple), **3.B** (blue), **3.C** (yellow), **3.F** (green), and **3.G** (orange), at 20.0 °C (293.2 K). Spectra are reported for 0.1 mM peptide samples in 3 mM NaOAc, 100 mM NaCl, pH 3.8.

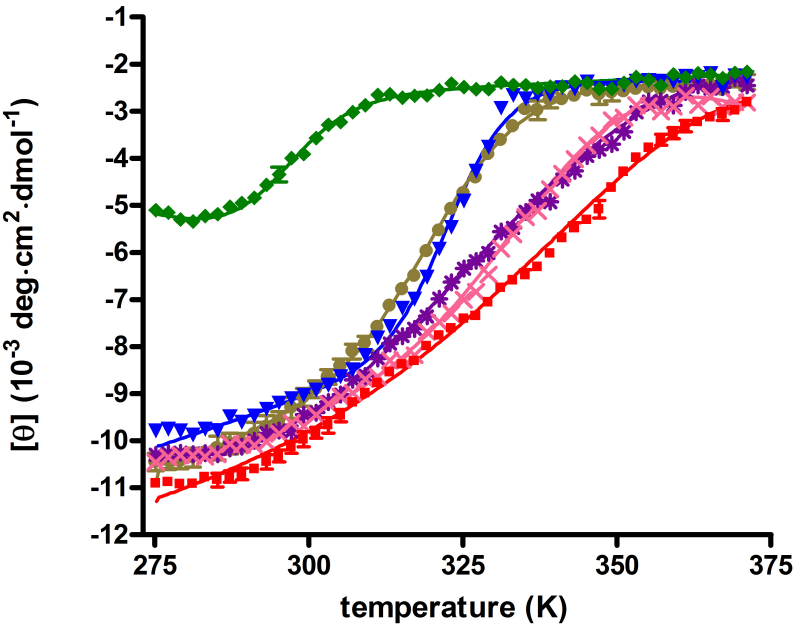

Figure 3.30. Mean residue ellipticity at 220 nm, of peptides **3.A** (red square), **3.D** (pink X), **3.E** (purple star), **3.B** (blue triangle), **3.C** (yellow circle), and **3.F** (green diamond) as a function of temperature. Solutions consisted of 0.1 mM peptide in aqueous solution (3 mM NaOAc, 100 mM NaCl, pH 3.8). Points represent mean values (n=3 for **3.A,** n=3 for **3.C**, n=2 for **3.F**, n=1 for all other peptides), with accompanying standard error of the mean.

3.4.5. Contribution of β-Arc to Peptide Structure

To test the hypothesis that the ^{D}Ala residues promote the *l* dihedral angles of a β-arc turn, substitutions were made at these positions with Asn (peptide **3.B**, with ^{D}Ala14Asn and ^{D}Ala17Asn) or Gly (peptide **3.C**, with ^{D}Ala14Gly and ^{D}Ala17Gly), residues that allow for *l* dihedral angles, but are also more promiscuous in the Ramachandran space that they can occupy.[61] On ^{1}H NMR of monomeric **3.B** and **3.C**, compared to **3.A**, less peak dispersion and less downfield chemical shifts were observed for both alpha and amide protons (**Figure 3.13**), which would be consistent with some loss of β-strand structure. Both monomeric **3.B** and monomeric **3.C** did not have the characteristic CD minimum observed for monomeric **3.A** (**Figure 3.19**). Both oligomeric **3.B** and oligomeric **3.C** retained the characteristic CD signature also observed for oligomeric **3.A** (**Figure 3.29**), but thermal denaturation of these peptides showed a leftward shift in the denaturation curve when Asn or Gly was introduced (**Figure 3.30**). Together these findings suggest that Asn or Gly in the β-arc allow for retention of this structure, but introduce a destabilizing flexibility. Similar results were obtained when additional Gly substitutions were made in the β-arc, for peptides **3.D** (with ^{D}Ala14Gly, His16Gly, and ^{D}Ala17Gly) and **3.E** (with Arg13Gly, ^{D}Ala14Gly, Ser15Gly, His16Gly, and ^{D}Ala17Gly), as shown in **Figures 3.13**, **3.19**, **3.29**, and **3.30**. In contrast to employing residues to promote flexibility at the *l* positions, the peptide **3.F** variant employed Ala, the enantiomer of ^{D}Ala, to promote dihedral angles opposite to *l* (i.e., right-handed α-helical dihedral angles, *a*) at these positions. Spectroscopic examination of peptide **3.F** showed a more pronounced destabilization of monomeric and oligomeric structures, with less downfield chemical shift of alpha and amide protons by NMR (**Figure 3.13**), loss of characteristic CD minima (**Figures 3.19**, **3.29**), and decreased stability

against thermal denaturation (**Figure 3.30**). Thus, the *l* dihedral angles are important for formation of the *blbbl* β-arc, and using Ala to promote the opposite (*a*) dihedral angles almost eliminates β-arch formation.

3.4.6. Hydrophobic Effect

Peptide **3.A** in aqueous solution likely contains interdigitation of hydrophobic side-chains, for both monomeric (**Figure 3.17b**, **Figure 3.14A**) and oligomeric structures. One signature of the dehydration of nonpolar surfaces is a large negative change in the standard heat capacity.[81] The energetic contribution of the hydrophobic effect to the folding of the peptide **3.A** monomer was determined as follows. The solvent-accessible surface areas (A) of peptide **3.A** were determined with Surface Racer, with settings of van der Waals radii from Richards 1977 and a probe radius of 1.4 Å, and were based on NMR structures of the monomer.[82] There was a burial of 814 $Å^2$ of nonpolar surface ($-\Delta A_{np}$) and 257 $Å^2$ of polar surface ($-\Delta A_p$), on folding of the peptide **3.A** monomer. From these values, we estimate the change in standard heat capacity with peptide folding, $\Delta C_P° = -225$ cal·mol^{-1}·K^{-1} (with dehydration of hydrophobic residues alone, $\Delta C_{P,np}° = -261$ cal·mol^{-1}·K^{-1}), and the change in standard free energy with hydrophobic dehydration, $\Delta G°_{np} = -21$ kcal·mol^{-1}, at room temperature, using the liquid-liquid transfer model of the hydrophobic effect.[83,84] Use of the gas-liquid transfer model of the hydrocarbon effect would yield a less negative value for ΔG_{np} (-9 kcal·mol^{-1}).[85] Thermal denaturation of the peptide **3.A** monomer could not be monitored by CD spectroscopy, because the intensity of the CD signal was too weak. Thermal denaturation of the peptide **3.A** oligomer was monitored by CD spectroscopy (**Figure 3.30** red), and a one-step, sigmoidal transition was observed. A one-step unfolding curve does not prove a two-state unfolding mechanism; however, for simplicity, a two-state

folding mechanism was assumed during fitting of these data. Van't Hoff analyses were performed,[86] using software written by Darrell R. McCaslin. This yielded a $\Delta C_P°_{oligomerization}$ = -686 cal·mol^{-1}·K^{-1}, which is 3-fold greater than the aforementioned $\Delta C_P°$ value for monomer folding. Since $\Delta C_P°$ is proportional to ΔA_{np}, this result suggests that there is even more burial of hydrophobic surfaces on oligomerization of peptide **3.A** than what is observed on folding of the monomer. This is because in the oligomers, hydrophobic surfaces may be buried via tertiary (intramolecular) or quaternary (inter-molecular) contacts.

Deletion of a strand containing two hydrophobic residues yielded peptide **3.G** (Thr1_Gly7del). NMR spectroscopy of monomeric peptide **3.G**, compared to peptide **3.A,** showed a loss of NOEs expected from β-arch tertiary structure (**Figure 3.14G**), as well as less peak dispersion and less downfield chemical shifts for both alpha and amide protons (**Figure 3.13** orange, **Figure 3.12**). Also, no minimum at ~220 nm was observed on CD spectroscopy of peptide **3.G** at high salt concentrations (**Figure 3.31** orange solid line). So, deletion of a strand containing two hydrophobic residues led to a loss of β-arch structure in the monomer and loss of oligomerization. This finding is consistent with the hypothesis that folding of peptide **3.A** is driven by the burial of hydrophobic side-chains.

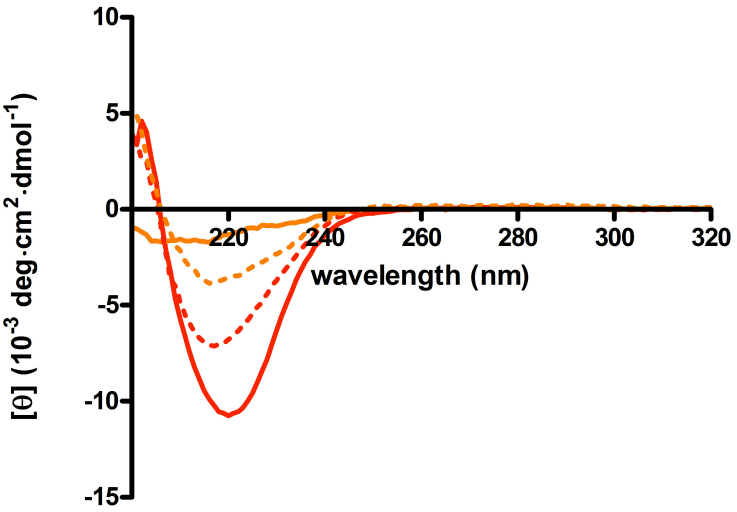

Figure 3.31. Mean residue ellipticity of peptides **3.A** (red) and **3.G** (orange), at 20.0 °C (293.2 K). Spectra are reported for 0.2 mM peptide samples in aqueous solution (3 mM NaOAc, 100 mM NaCl, pH 3.8), without (solid lines) or with (dashed lines) 50% v/v TFE.

Further evidence for the hydrophobic effect was provided by studies of peptide **3.A** in an alcohol co-solvent (50% v/v 2,2,2- trifluoroethanol, TFE) at high salt conditions. In the presence of the alcohol co-solvent, a loss in the CD signal of peptide **3.A** was observed (**Figure 3.31** red dashed line). This suggests less tertiary structure and/or less oligomerization in the less hydrophilic environment created by the alcohol, compared to in a purely aqueous solution. In juxtaposition, addition of the alcohol co-solvent to **3.G**, a peptide for which there was no evidence for tertiary structure in purely aqueous solution, led to an enhancement of the CD signal consistent with increased secondary structure (**Figure 3.31** orange dashed line).[27,51]

Finally, it appears that the hydrophobic effect, and not the artificial β-arc linker, drives peptide folding, as replacement of all five residues of the β-arc with glycine residues did not lead to abolition of the tertiary fold or oligomerization (**Figures 3.13**, **3.29**).

3.5. Conclusions

Diseases involving amyloid proteins debilitate tens of millions of people worldwide,[87] yet there are no medical tools available for early diagnosis, prevention, or treatment of amyloid formation, partly because amyloid structures and mechanisms of formation remain enigmatic. *In vitro* models of amyloid formation – including *de novo* designed peptides that form fibrils with antiparallel β-sheet structure,[88-90] and oligomer models stabilized with *in situ* chemical cross-linking[91] – have been influential in the spectroscopic characterization of amyloids and development of theories about the forces driving amyloid formation. Here we describe a new class of synthetic protein architecture, containing both parallel β-sheet structure and β-arch structure. The monomeric form of our designed peptide models the minimal folding unit of an amyloid polymorph thought to be medically relevant, and is stable in aqueous solution for days. The retention of this peptide fold when the designed β-arc is substituted with more flexible residues shows that conformational preferences of the designed β-arc do not drive the peptide folding. Rather, the peptide's tertiary fold and oligomerization are likely driven by the hydrophobic effect, a hypothesis supported by results of varying the hydrophobicity of the solvent or deleting hydrophobic residues. Furthermore, based on the NMR structure of peptide **3.A**, we calculate a burial of 814 $Å^2$ of nonpolar surface area on folding of the monomer, which corresponds to a 9-21 $kcal \cdot mol^{-1}$ stabilization of the structure. We demonstrate how varying salt content can control oligomerization in our model system, allowing for study of the monomeric or oligomeric structures, and report the characteristic CD spectra of the monomeric and oligomeric forms. Model systems such as ours may have many potential applications, including in studies of the forces driving amyloid oligomerization, as agents to

seed amyloid growth *in vitro*, as amyloid substitutes in screens for diagnostic and therapeutic tool design, or even as scaffolds of diagnostic or therapeutic agents.

3.6. Future Directions

We probed the effect of hydrophobic contacts on the folding of peptide **3.A**, with structural studies of peptides **3.A** and **3.G**, the latter of which was a peptide **3.A** variant with the deletion of a strand containing two aliphatic, hydrophobic residues (**section 3.4.6**). Further studies could be performed on the roles of hydrophobic side-chains in the folding of the peptide. We did not study a peptide **3.A** variant with the deletion of a strand containing aromatic, hydrophobic residues, because aromatic residues were needed for the assignment of proton signals on NMR. Similarly, we did not study the hydrophobic effect using less dramatic variants of peptide **3.A** (e.g., substituting out one hydrophobic residue at a time), because studies of peptides **3.B-3.F** determined that even peptides that were only marginally less folded than peptide **3.A** showed losses in NMR peak dispersion, which prohibited the unambiguous assignment of proton peaks. With access to higher sensitivity probes, natural abundance ^{13}C NMR spectra could be taken, to facilitate the assignment of spectra with less peak dispersion.

The aqueous solution structure of peptide **3.A** in this study, as well as structures of peptides from other studies (e.g., peptide **2.C** in **Chapter 2**), suggests that the diacid linker we employed may not be very effective at promoting parallel β-sheet structure. A next step in the optimization of the peptide models we are developing would be to redesign the diacid linker.

Our model system may be employed to quantify how known mutations in amyloid proteins effect changes in the kinetics and thermodynamics of amyloid folding. One could replace β-strands of peptide **3.A** with ones from specific amyloid structures. Single amino acid

substitutions have been identified as precipitating earlier onsets of certain amyloid diseases, and of conferring greater infectivity to certain prions. The effects of such substitutions on secondary, tertiary, and higher order folding could be studied in our model system, or model systems that ours may inspire.

3.7. References

(1) Merlini, G.; Bellotti, V. *New Engl J Med* **2003**, *349*, 583.

(2) Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.; Merlini, G.; Saraiva, M. J.; Westermark, P. *Amyloid* **2014**, *21*, 221.

(3) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. *Nature* **2012**, *491*, 222.

(4) Floudas, C. A.; Fung, H. K.; McAllister, S. R.; Monnigmann, M.; Rajgaria, R. *Chem Eng Sci* **2006**, *61*, 966.

(5) Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N. *Angew Chem Int Ed* **2013**, *52*, 5700.

(6) Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. *Annu Rev Biophys* **2013**, *42*, 315.

(7) Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. *Trends Biotechnol* **2014**, *32*, 99.

(8) Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W. *Curr Opin Struc Biol* **2015**, *33*, 16.

(9) Tycko, R. *Neuron* **2015**, *86*, 632.

(10) Zerovnik, E.; Stoka, V.; Mirtic, A.; Guncar, G.; Grdadolnik, J.; Staniforth, R. A.; Turk, D.; Turk, V. *The FEBS J* **2011**, *278*, 2263.

(11) Fandrich, M. *J Mol Biol* **2012**, *421*, 427.

(12) Kajava, A. V.; Baxa, U.; Steven, A. C. *FASEB J* **2010**, *24*, 1311.

(13) Klein, W. L. *Neurochem Int* **2002**, *41*, 345.

(14) Lesne, S. E.; Sherman, M. A.; Grant, M.; Kuskowski, M.; Schneider, J. A.; Bennett, D. A.; Ashe, K. H. *Brain* **2013**, *136*, 1383.

(15) Chimon, S.; Shaibat, M. A.; Jones, C. R.; Calero, D. C.; Aizezi, B.; Ishii, Y. *Nat Struct Mol Biol* **2007**, *14*, 1157.

(16) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773.

(17) Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. *J Mol Biol* **2010**, *395*, 717.

(18) Hennetin, J.; Jullian, B.; Steven, A. C.; Kajava, A. V. *J Mol Biol* **2006**, *358*, 1094.

(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* **2000**, *28*, 235.

(20) Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G. *Nucleic Acids Res* **2011**, *39*, D411.

(21) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.

(22) Wang, G. L.; Dunbrack, R. L. *Bioinformatics* **2003**, *19*, 1589.

(23) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res* **2004**, *14*, 1188.

(24) Magrane, M.; Consortium, U. *Database* **2011**, *2011*, bar009.

(25) Lee, B.; Richards, F. M. *J Mol Biol* **1971**, *55*, 379.

(26) Hubbard, S. J.; Campbell, S. F.; Thornton, J. M. *J Mol Biol* **1991**, *220*, 507.

(27) Kung, V. M.; Cornilescu, G.; Gellman, S. H. *Angew Chem Int Ed* **2015**, *54*, 14336.

(28) Hwang, T. L.; Shaka, A. J. *J Magn Reson Ser A* **1995**, *112*, 275.

(29) Derome, A. E.; Williamson, M. P. *J Magn Reson* **1990**, *88*, 177.

(30) Piotto, M.; Saudek, V.; Sklenar, V. *J Biomol NMR* **1992**, *2*, 661.

(31) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J Magn Reson Ser A* **1993**, *102*, 241.

(32) Lippens, G.; Dhalluin, C.; Wieruszeski, J. M. *J Biomol NMR* **1995**, *5*, 327.

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

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

(35) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(36) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J Am Chem Soc* **2004**, *126*, 7238.

(37) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J Mol Biol* **1991**, *222*, 311.

(38) Wishart, D. S. *Prog Nucl Mag Res Sp* **2011**, *58*, 62.

(39) Sharman, G. J.; Griffiths-Jones, S. R.; Jourdan, M.; Searle, M. S. *J Am Chem Soc* **2001**, *123*, 12318.

(40) Mielke, S. P.; Krishnan, V. V. *Prog Nucl Mag Res Sp* **2009**, *54*, 141.

(41) Hollingsworth, S. A.; Karplus, P. A. *Biomol Concepts* **2010**, *1*, 271.

(42) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.

(43) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J Magn Reson* **2003**, *160*, 65.

(44) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett* **1988**, *229*, 317.

(45) Clore, G. M.; Kuszewski, J. *J Am Chem Soc* **2002**, *124*, 2866.

(46) Schwieters, C. D.; Clore, G. M. *J Phys Chem B* **2008**, *112*, 6070.

(47) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins* **2007**, *66*, 778.

(48) Greenfield, N. J. *Nat Protoc* **2006**, *1*, 2876.

(49) Sieber, V.; Jurnak, F.; Moe, G. R. *Proteins* **1995**, *23*, 32.

(50) Woody, R. W. *Tetrahedron-Asymmetr* **1993**, *4*, 529.

(51) Venkatraman, J.; Shankaramma, S. C.; Balaram, P. *Chem Rev* **2001**, *101*, 3131.

(52) *Electron Microscopy: Principles and Techniques for Biologists*; Jones and Bartlett Publishers: Boston, MA, 1992.

(53) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.

(54) Laue, T. M. *Method Enzymol* **1995**, *259*, 427.

(55) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; The Royal Society of Chemistry: Cambridge, U.K., 1992, p 90.

(56) Perkins, S. J. *Eur J Biochem* **1986**, *157*, 169.

(57) Durchschlag, H.; Zipper, P. *Prog Colloid Polym Sci* **1994**, *94*, 20.

(58) Tanner, J. E. *J Chem Phys* **1970**, *52*, 2523.

(59) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424.

(60) Efimov, A. V. *FEBS Lett* **1987**, *224*, 372.

(61) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr D* **2002**, *58*, 768.

(62) Deane, C. M.; Allen, F. H.; Taylor, R.; Blundell, T. L. *Protein Eng* **1999**, *12*, 1025.

(63) Srinivasan, N.; Anuradha, V. S.; Ramakrishnan, C.; Sowdhamini, R.; Balaram, P. *Int J Pept Prot Res* **1994**, *44*, 112.

(64) *D-Amino Acids: A New Frontier in Amino Acid and Protein Research - Practical Methods and Protocols*; Nova Science Publishers, Incorporated, 2006.

(65) Munoz, V.; Serrano, L. *Proteins* **1994**, *20*, 301.

(66) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat Struct Biol* **1995**, *2*, 596.

(67) Fooks, H. M.; Martin, A. C. R.; Woolfson, D. N.; Sessions, R. B.; Hutchinson, E. G. *J Mol Biol* **2006**, *356*, 32.

(68) Fisk, J. D.; Gellman, S. H. *J Am Chem Soc* **2001**, *123*, 343.

(69) Fisk, J. D.; Schmitt, M. A.; Gellman, S. H. *J Am Chem Soc* **2006**, *128*, 7148.

(70) Freire, F.; Fisk, J. D.; Peoples, A. J.; Ivancic, M.; Guzei, I. A.; Gellman, S. H. *J Am Chem Soc* **2008**, *130*, 7839.

(71) Freire, F.; Gellman, S. H. *J Am Chem Soc* **2009**, *131*, 7970.

(72) Fisk, J. D.; Powell, D. R.; Gellman, S. H. *J Am Chem Soc* **2000**, *122*, 5443.

(73) Nesloney, C. L.; Kelly, J. W. *Bioorgan Med Chem* **1996**, *4*, 739.

(74) Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: New York, 2000.

(75) Ortega, A.; Amoros, D.; de la Torre, J. G. *Biophys J* **2011**, *101*, 892.

(76) Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. *Nat Struct Mol Biol* **2010**, *17*, 561.

(77) Stroud, J. C.; Liu, C.; Teng, P. K.; Eisenberg, D. *Proc Natl Acad Sci USA* **2012**, *109*, 7717.

(78) Micsonaia, A.; Wienb, F.; Kernyaa, L.; Leec, Y.-H.; Gotoc, Y.; Réfrégiersb, M.; Kardos, J. *Proc Natl Acad Sci USA* **2015**, *112*, E3095.

(79) Bouchard, M.; Zurdo, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V. *Protein Sci* **2000**, *9*, 1960.

(80) McPhie, P. *Biopolymers* **2004**, *75*, 140.

(81) Privalov, P. L.; Gill, S. J. *Adv Protein Chem* **1988**, *39*, 191.

(82) Tsodikov, O. V.; Record, M. T.; Sergeev, Y. V. *J Comput Chem* **2002**, *23*, 600.

(83) Spolar, R. S.; Ha, J. H.; Record, M. T., Jr. *Proc Natl Acad Sci USA* **1989**, *86*, 8382.

(84) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777.

(85) Baldwin, R. L. *Proc Natl Acad Sci USA* **2014**, *111*, 13052.

(86) Pace, C. N.; Shirley, B. A.; Thomson, J. A. In *Protein Structure: A Practical Approach*; Creighton, T. E., Ed.; IRL Press: Oxford, UK, 1990, p 311.

(87) Prince, M.; Bryce, R.; Albanese, E.; Wimo, A.; Ribeiro, W.; Ferri, C. P. *Alzheimers Dement* **2013**, *9*, 63.

(88) Lopez De La Paz, M.; Goldie, K.; Zurdo, J.; Lacroix, E.; Dobson, C. M.; Hoenger, A.; Serrano, L. *Proc Natl Acad Sci USA* **2002**, *99*, 16052.

(89) Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Muller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O. *Proc Natl Acad Sci USA* **2004**, *101*, 4435.

(90) Biancalana, M.; Makabe, K.; Koide, S. *Proc Natl Acad Sci USA* **2010**, *107*, 3469.

(91) Ono, K.; Condron, M. M.; Teplow, D. B. *Proc Natl Acad Sci USA* **2009**, *106*, 14745.

Appendix 1. Supplemental Information for Synthetic Protocols

A1.1. Development of Synthetic Route to Four-Stranded β-Arch Peptides

Previous work in the Samuel H. Gellman Laboratory indicated that the synthesis and purification of two-stranded parallel β-sheet peptides containing either one diacid linker or one diamine linker, was challenging and time-consuming. There was no record of a successful synthesis and purification of a model peptide containing at least three intended β-strands, a diamine linker, and a diacid linker, although attempts at these syntheses were made my multiple graduate students and post-doctoral researchers. We developed a serviceable synthetic route to a four-stranded β-arch peptide (**Chapter 3**), and an analogous route was employed to synthesize a three-stranded parallel β-sheet (**Chapter 2**).

The most challenging step of syntheses of peptides containing more than two parallel β-strands appeared to be the segment condensation step (e.g., **Figure A1.1**). We attempted segment condensations using conditions that prior work in the Gellman Laboratory had indicated might overcome the challenges of synthesizing peptides containing both a diamine and a diacid linker.[1] These conditions yielded unreacted starting materials ("amine" and "acid" segments, as drawn in **Figure A1.1**), and little to no product (**Figure A1.2**). It was apparent that these standard segment condensation conditions would not yield enough peptide for conducting NMR or other experiments.

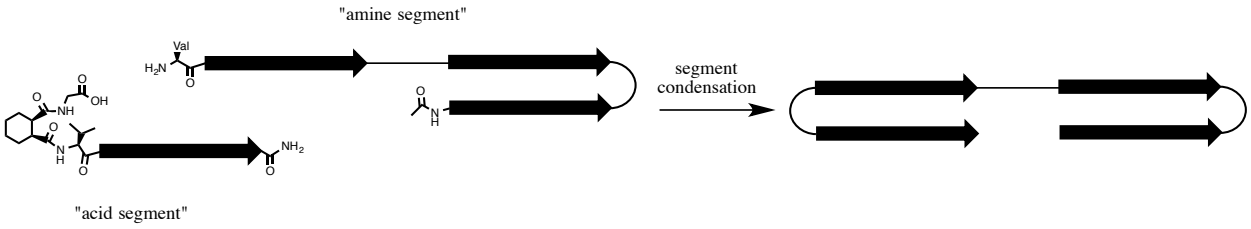

Figure A1.1. Segment condensation step for the synthesis of a 4-stranded β-arch peptide.

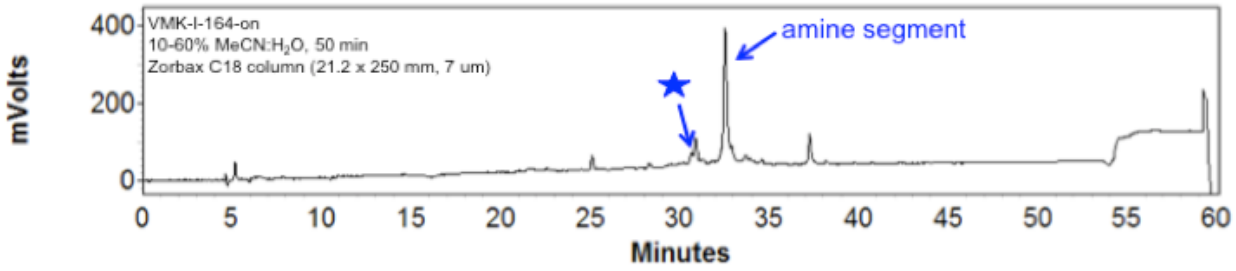

Figure A1.2. Representative HPLC trace of crude product of a segment condensation that employed standard segment condensation conditions (including PyBOP as the coupling reagent). Star indicates HPLC peak corresponding to peptide **A2.X**; main peak corresponds to unreacted starting material (amine segment).

We proceeded to optimize the segment condensation step for the synthesis of four-stranded β-arch peptides. When using PyBOP as the coupling reagent, we found that the following factors did not have a significant impact on the yield: the amount of coupling reagent (1-2 eq, per 1 eq of "acid segment"), the duration and temperature of the reaction (1-3 day reactions at 21°C, or 12-36 minute reactions at 70°C), and the addition of a nucleophilic catalyst (DMAP). The yield was improved by using excess "acid segment" (2 eq, per 1 eq of "amine segment"), but not by using excess "amine segment" (1-2 eq, per 1 eq of "acid segment"). This was not surprising, but was noteworthy because it would otherwise have been desirable to make the limiting reagent the "acid segment," since the diacid linker (**section A1.3.1**) is more time-consuming to synthesize than the diamine linker (**section A1.3.3**). We expected higher yields from segment condensations conducted with all reagents in solution (less steric congestion), as opposed to with the "amine segment" still attached to the resin (more steric congestion). (The alternate on-resin condition, with the "acid segment" on resin and "amine segment" in solution, was not tested because we expected the larger "amine segment" to have greater difficulty diffusing through solution than the "acid segment.") Contrary to expectation, reactions off of resin did not necessarily yield more product than reactions on resin. Side products specific to

off-resin segment condensations may be attributable to partial deprotection of the side-chains of the reactant peptide segments, during their cleavage from Sieber amide resin or subsequent manipulations. Similarly, incorporating a pseudoproline dipeptide unit in the middle of the peptide synthesis (at residues 15-16, or at residues 8-9) – a technique that can sometimes increase SPPS yields by preventing the aggregation of hydrophobic residues within the growing peptide chain[2] – did not have a significant effect on the peptide yield.

Using highly reactive peptide coupling reagents increased the yield of the segment condensation, as long as the reaction was only allowed to proceed for ~24 hours. PyBOP (the coupling reagent recommended by Gellman and coworkers after their screens of segment condensation reaction conditions) is a stronger coupling reagent than HBTU, the latter of which was at the time the coupling reagent of choice for normal SPPS in the Gellman Laboratory, but even stronger coupling reagents existed. We selected one strong coupling reagent from each class of coupling reagents to screen in β-arch peptide segment condensations: DIC (carbodiimide), HATU (aminium), PyBrOP (phosphonium), and TPTU (uronium). These were coupling reagents indicated in literature to have the fastest reactivity, while causing the least amount of amino acid racemization, and/or give the best results for peptide segment coupling or other difficult couplings.[3,4] Segment condensations employing DIC, PyBrOP, TPTU, or HATU yielded β-arch peptides in at least 5x greater yield than segment condensations employing PyBOP. However, if the segment condensation was allowed to proceed longer than 24 hours, or if the crude product was not immediately purified, degradation of the β-arch peptides was observed (**Figure A1.3**). This product degradation was particularly rapid when the coupling reagent was DIC or PyBrOP, with total disappearance of product by hour 156, so we recommend the employment of HATU or

TPTU. All side products were isolated by HPLC and characterized by MALDI-TOF MS. For the synthesis of peptide **A2.Y** (see **Appendix 2** for peptide design), the observed $[M+H^+]^+$ peaks of major side products were 2635.7 and 2447.0, for DIC-mediated segment condensations; and 2361.7 and 2605.9 (HATU values), for PyBrOP-, TPTU-, and HATU-mediated segment condensations. All four of these side product masses are within 1 Da of calculated masses for peptide fragments produced by the hydrolysis of amide bonds in peptide **A2.Y** (**Figure A1.4**). It is notable that amide bond hydrolysis may account for the observed peptide degradation, since amide bonds are typically stable under the reaction conditions discussed here, the most extreme of which involved only moderately basic conditions and heating (excess DIEA, 12 min at 70˚C). We recommend the "extended" heating method (12 min at 70˚C) in place of "double coupling" methods, when possible, as the latter require more time and reagents. For segment condensation reactions in the synthesis of β-arch peptides, we obtained similar yields for 24-hour reactions at room temperature and 12-min reactions at 70 ˚C; however, in theory the former (no heating) is preferable, to minimize amide hydrolysis. Finally we recommend N-methyl-2-pyrrolidone (NMP), rather than DMF (which can decompose to formic acid), as the solvent for segment condensation reactions, as a precautionary measure against cleavage of side-chain protecting groups.

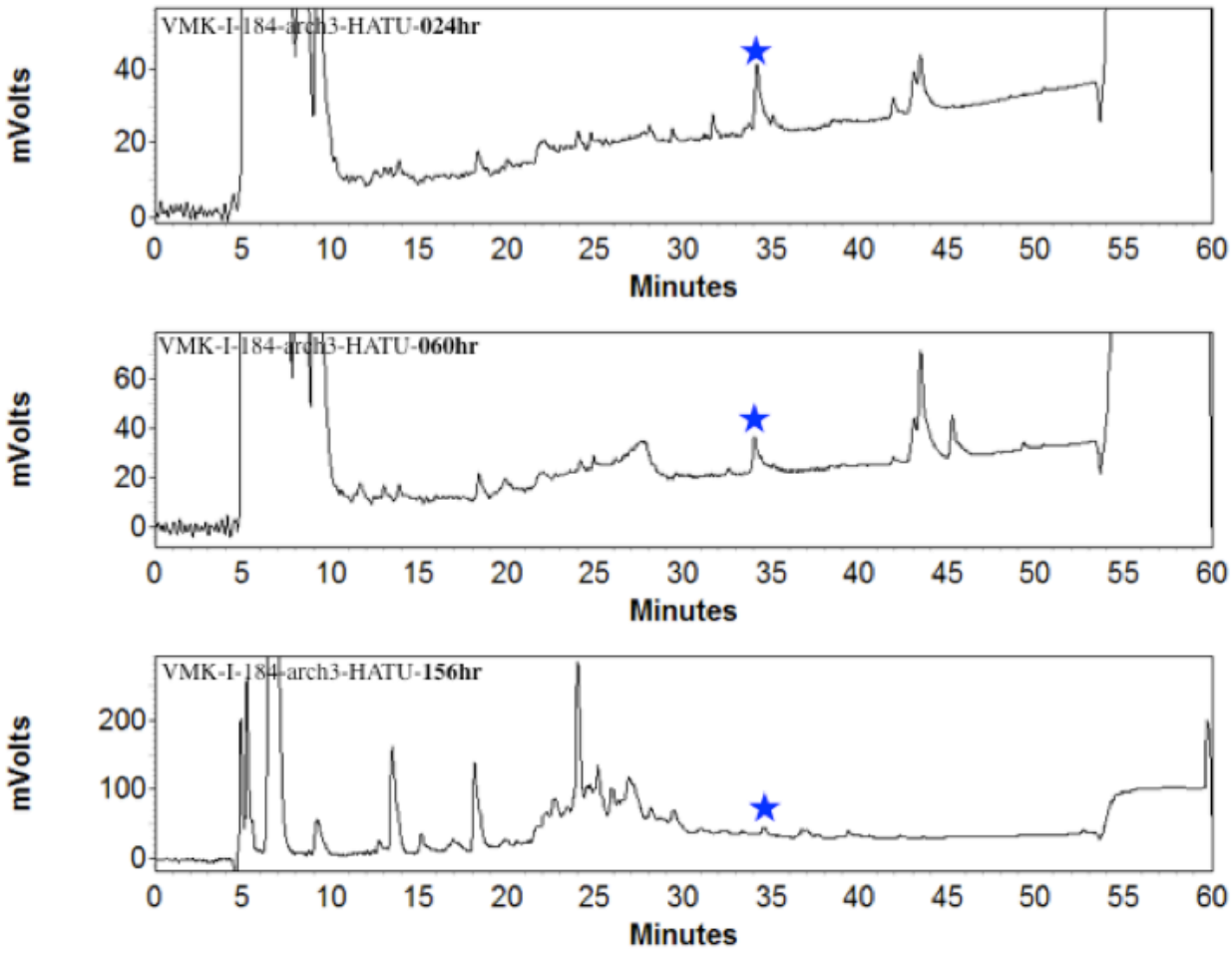

Figure A1.3. Representative HPLC traces showing degradation of peptide **A2.Y** over time.

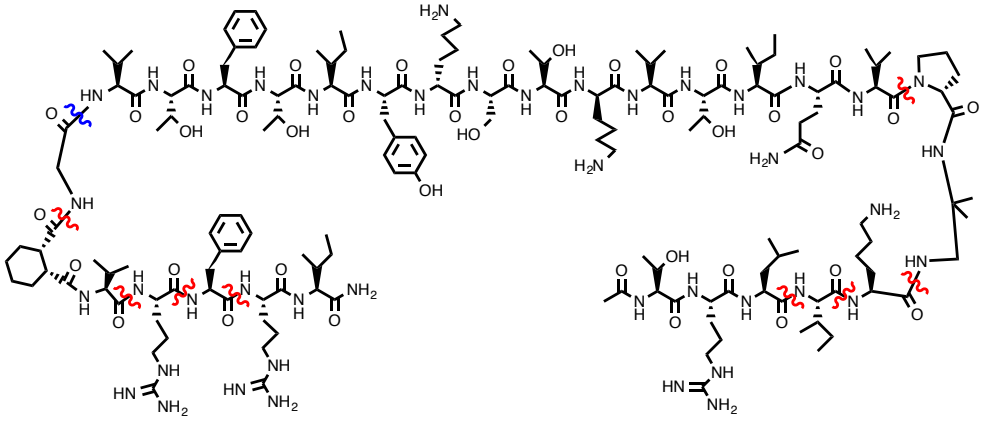

Figure A1.4. Sites of amide bond hydrolysis of peptide **A2.Y**.

The β-arch peptides contained multiple amino acids notorious for causing "difficult" solid phase peptide syntheses. We designed peptide **A1.L**, the truncation of peptide **A2.Z**

(see **Appendix 2** for peptide design) to residues 10-20 (with acetyl-capped N-terminus and amide at C-terminus), to test the feasibility of the synthesis and handling of a peptide containing Met, Cys, His, and Trp residues. We used established methods for minimizing the production of undesired side products when synthesizing and handling peptides containing these "difficult" amino acids. The coupling of Cys onto the growing peptide chain employed 2,4,6-collidine, a weaker base than DIEA, to minimize the racemization of Cys at the α-carbon.[5] The peptide was cleaved from resin in acid under $N_{2(g)}$, in a cleavage cocktail containing extra scavengers (TIPS, thioanisole, phenol, EDT), to suppress oxidation and alkylation of side-chains. Standard peptide purification procedures were used. These synthesis and purification conditions produced peptide **A1.L** in high yield. For purified peptide, no oxidation or other side products were observable by ^{1}H NMR, analytical HPLC, or MALDI-TOF MS, at the conditions used to obtain NMR spectra (0.02-2 mM peptide, 2-100 mM NaOAc-d_3, pH 3.8, 4-5 °C, 0-48 hours).

Crude peptides should be stored dry, under $N_{2(g)}$ at -20 °C, until purification, to minimize peptide degradation. Crude products (100 μmol scale syntheses) can be dissolved in 5-10 mL of 50% solvent B prior to purification, and an excess of tris(2-carboxyethyl)phosphine (TCEP) may be added to this solution to reduce any undesired disulfide bonds. Purification of the completed four-stranded β-arch peptide by reverse-phase HPLC can be accomplished using standard methods. Before each use, care should be taken to maintain the HPLC instrument (e.g., fix all leaks, flush the instrument and column thoroughly to remove impurities), such that instrument pressures do not fluctuate significantly during runs, and product peaks can be resolved from closely eluting impurities. Even after maintenance, old HPLC instruments may still not be capable of

resolving the desired product from impurities (e.g., fluctuations in instrument pressures will lead to fluctuations in product retention times, peak broadening, and other effects that prevent the successful separation of products from impurities using automated fraction collection methods). In such cases, fractions can be manually collected, as the human eye can be more effective than current HPLC software at identifying where fraction collections should begin and end (when retention times vary from trace to trace, or peaks are poorly resolved). Purifications of the four-stranded β-arch peptides reported in this thesis were accomplished with a combination of automated and manual fraction collections. The solid phase was a C18-functionalized preparatory or semi-preparatory column. The use of a semi-preparatory column may increase yields, if that column is well-maintained. Each peptide required a series of 1-3 purifications by HPLC, with mobile phases that were always linear gradients of "B solvent" (CH_3CN:CF_3COOH, 100:0.1 v/v) in "A solvent" (H_2O:CF_3COOH, 100:0.1 v/v). The chosen linear gradients depended on the peptides and the conditions of the HPLC instruments, but were among the following: 10-60% B over 50 min, x-z% B over 20 min, x-z% B over 50 min, and x-y% B over 20 min, where z = x+10, y = x+5, and x, y, z $\in$ [10, 60].

Finally, it is notable that a four-stranded β-arch peptide may not be detectable by MALDI-TOF MS using standard MALDI-TOF MS conditions, since the β-arch peptide is larger and less abundant than other peptide fragments in the crude product. α-Cyano-4-hydroxycinnamic acid (CHCA) is the preferred MALDI matrix for peptides in the mass range including peptide **A2.Y** (~3.5 kDa), while sinapinic acid (SA) is generally used for larger peptides/proteins.[6-8] After screening a few MALDI conditions, we determined that small amounts of our four-stranded β-arch peptide could be detected in crude product

mixtures, by using a 1:1 mixture of SA and CHCA dissolved in MeCN/H_2O, at low laser attenuation.

The above experiments resulted in the development of the synthetic route depicted in **Chapter 3** (**Figure 3.2**), which is also described in more detail below (**section A1.2**). With the employment of this route for a 100 μmol scale synthesis of a four-stranded β-arch peptide, the final yield of purified peptide was typically ~6 mg (~2 μmol).

A1.2. Peptide Synthesis, Purification, and Quantification Methods

Peptides were synthesized via the routes depicted within chapters and appendices. More details about these routes can be found below, and a description of the development of routes toward peptides containing >1 change in strand direction can be found above (**section A1.1**).

Standard Fmoc solid phase peptide synthesis (SPPS) methods were employed. Both peptide coupling and Fmoc-deprotection reactions were performed on 100-200 μmol scale, in 10 mL N,N-dimethylformamide (DMF) solvent, in 12-24 mL fritted syringes.

Peptides were synthesized on NovaPEG Rink amide resin, unless they were peptide fragments that were used for off-resin segment condensations, in which case Sieber amide resin was employed, because it can be cleaved by mildly acidic conditions that preserve side-chain protecting groups. All reactive side-chains were protected with protecting groups that were removable in strongly acidic conditions. For coupling reactions, 4 eq. Fmoc-protected amino acids (with acid-labile side-chain protecting groups) were activated in solution with 4 eq. O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU), 4 eq. hydroxybenzotriazole (HOBt), and 8 eq. N,N-diisopropylethylamine (DIEA) at room temperature for 2 min, and then combined with

the growing peptide chain and heated at 70 °C for 4 min (microwave-assisted). An extended coupling method (microwaved-assisted heating at 70 °C for 36 min), still employing HBTU as the coupling reagent, was used to couple diamine or diacid linkers to growing peptide chains. Fmoc deprotection reactions were accomplished using an excess of 20% v/v piperidine in DMF, heated at 80 °C for 2 min (microwave-assisted). After each coupling and deprotection reaction, the resin was washed 9x with DMF. The resin was dried on an aspirator prior to cleavage of the peptide from resin.

Acetyl capping of the C-termini of peptides was accomplished by treatment with an excess of acetic anhydride (0.4 mL acetic anhydride, 0.7 mL DIEA, 3 mL DMF) for 5-10 min at room temperature. Thorough washing with DMF followed.

Alloc and allyl deprotection reactions were accomplished with stoichiometric amounts of $Pd(PPh_3)_4$ [87 mg $Pd(PPh_3)_4$, 0.4 mL 1:3 (v/v) AcOH:N-methylmorpholine, 4 mL DCM]. Literature procedures using catalytic amounts of $Pd(PPh_3)_4$ will suffice, if one has access to $Pd(PPh_3)_4$ that has been appropriately stored. Washing steps after alloc/allyl-deprotection reactions should be thorough (e.g., DCM 3x, DMF 3x, 20% piperidine 1x, DMF 3x).

Side-chain protected peptide segments were cleaved from Sieber amide resin, by flowing 100 mL of 1% v/v TFA in dichloromethane (DCM) across swollen resin into a neutralizing solution of pyridine (1.056 mL) in DCM, and then dried on a high-vacuum rotary evaporator to an oil. The peptide segments were then precipitated in H_2O, immediately pelleted (4,500 rpm for 10 min), and dried. These side-chain protected peptide segments were subsequently condensed with other peptide segments. Purification of the side-chain protected peptide segments [by extraction (dissolve in EtOAc, extract with

saturated $NaHCO_{3(aq)}$ and then brine) or HPLC] prior to segment condensation reactions is helpful but not necessary for achieving high-yielding segment condensation reactions. Time and materials can be saved by adding the crude peptide segment to a segment condensation reaction with an excess of base (4-20 eq. DIEA).

Completed peptides were cleaved from Rink amide resin and globally side-chain deprotected, by treating the peptide-resin for 5 hours with 10 mL of a cocktail of 82.5% v/v trifluoroacetic acid (TFA), 5% v/v H_2O, 1% v/v triisopropylsilane (TIPS), 5% v/v thioanisole, 5% v/v phenol, and 2.5% v/v ethanedithiol (EDT).[9] Volumes of the cleavage solutions were subsequently reduced to < 5 mL under an $N_{2(g)}$ stream. The crude peptides were precipitated in 45 mL of anhydrous diethyl ether at -20 °C for 1 hour, pelleted (4,500 rpm for 10 min), and dried under nitrogen.

Peptides were purified by reverse-phase high-performance liquid chromatography (HPLC). The solid phase was a C18-functionalized preparatory (Agilent ZORBAX 300SB-C18 PrepHT column, 7µm, 2x250mm) or semi-preparatory (Supelco Discovery BIO Wide Pore C18 column, 5µm, 10x250mm) column. The mobile phase was a linear gradient of B solvent (CH_3CN:CF_3COOH, 100:0.1 v/v) in A solvent (H_2O:CF_3COOH, 100:0.1 v/v).

The purified peptide was dried by lyophilization. Peptide identities were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), using a matrix of α-cyano-4-hydroxycinnamic acid (CHCA). Peptide purity was assessed to be >95%, by integration of peak areas on analytical HPLC (Waters xBridge BEH130 C18 column, 5µm, 6x250mm; 10-60% B over 50 min) or ultra performance liquid chromatography (UPLC) (ACQUITY UPLC peptide BEH C18 column, 130Å, 1.7µm, 2.1x100mm; 10-60% B over 5 min).

Peptide stock solutions were prepared by dissolving lyophilized peptide in aqueous solution (Millipore quality H_2O). A small sample of this solution was used to make peptide stock solution in 6 M guanidinium chloride (GdmCl). Stock concentrations in 6 M GdmCl were determined by UV spectrophotometry in a quartz cell (l = 1 cm). The absorbance (A) and calculated extinction coefficient (ε) for the peptide,

$$\varepsilon_{280nm} = (\#Trp)\cdot\left(5690 M^{-1}cm^{-1}\right) + (\#Tyr)\cdot\left(1280 M^{-1}cm^{-1}\right) + (\#Cys)\cdot\left(0120 M^{-1}cm^{-1}\right),$$[10]

were used to determine the concentration of peptide in the stock solutions, according to Beer's Law, $A = \varepsilon \cdot c \cdot l$. Relative quantities of peptides can subsequently be determined by UPLC, given a well-maintained UPLC instrument and column.

A1.3. Synthesis and Characterization of Non-Natural Residues.

Select characterization data for linker syntheses are reported below, as these and other information were missing from or erroneous in all previous reports of these syntheses.

A1.3.1. OAllyl-Gly-(1S,2R)-CHDA-Val-OH Linker.

The OAllyl-Gly-(1S,2R)-CHDA-Val-OH linker was synthesized via the route depicted in **Figure A1.5**, which is similar to a previously reported synthetic route.[11] **Figures A1.9-A1.15** are ^1H NMR spectra from this synthesis, taken in $CDCl_3$, with an internal standard of TMS, on a 400 MHz spectrometer.

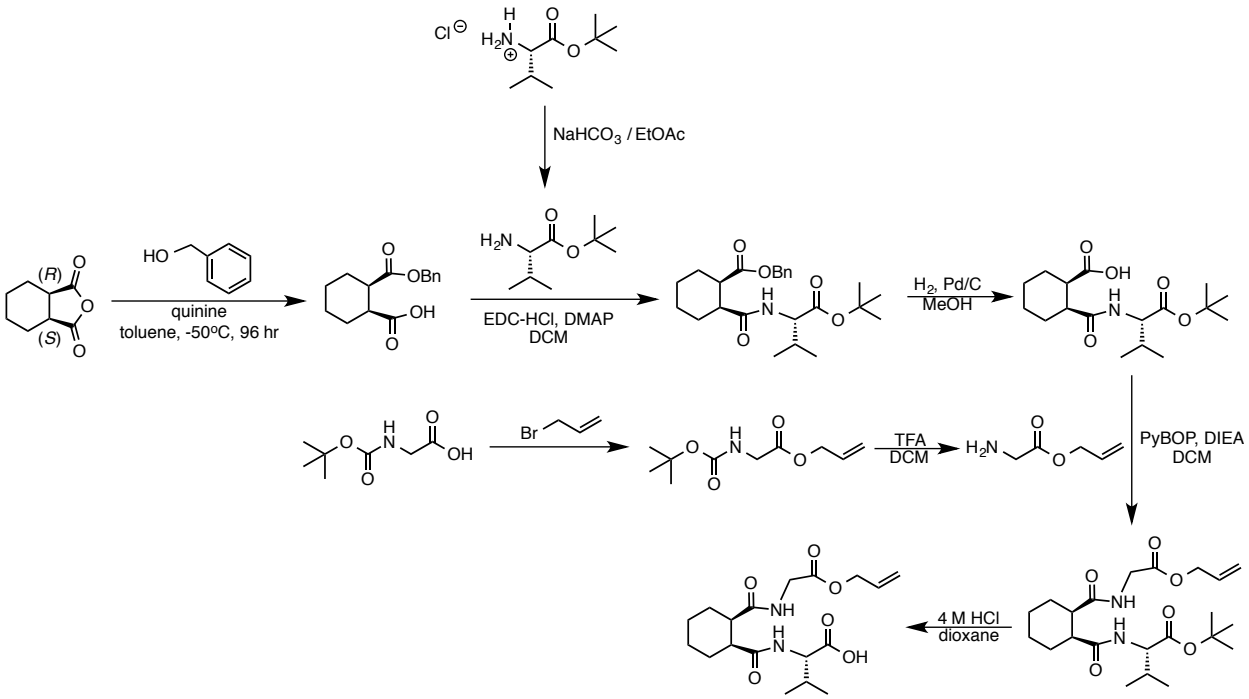

Figure A1.5. Synthetic route to OAllyl-Gly-(1S,2R)-CHDA-Val-OH linker.

A1.3.2. OAllyl-Gly-(1R,2*S*)-CHDA-Val-OH Linker.

The OAllyl-Gly-(1R,2*S*)-CHDA-Val-OH linker was synthesized via the route depicted in **Figure A1.5**, except using quinidine instead of quinine, for the asymmetric alcoholysis of the meso-anhydride. **Figures A1.16-A1.20** are ^{1}H NMR spectra from this synthesis, taken in $CDCl_3$, with an internal standard of TMS, on 300 MHz or 400 MHz spectrometers.

BnO-(1*R*,2*S*)-CHDA-OH. Alkaloid-induced asymmetric alcoholysis of meso-anhydrides (**Figure A1.5**) is discussed in more detail elsewhere.[12,13] Cyclohexane-*cis*-1,2-dicarboxylic anhydride (6.6 g, 42.8 mmol) and quinidine (15.2 g, 46.9 mmol) were dissolved in 500 mL toluene at -50 °C. Benzyl alcohol was added (13.4 mL, 129.5 mmol), and the reaction was stirred for 5 days at -50 °C. The crude product was dried under reduced pressure (35 °C, 33 mbar), to an oil. The crude product was dissolved in 250 mL diethyl ether, and washed with 2 N $HCl_{(aq)}$ (3 x 60 mL). The aqueous layer was back-extracted with diethyl ether (5 x 100 mL). The combined organic layers were extracted

with saturated $NaHCO_{3(aq)}$ (5 x 150 mL). The aqueous layers were combined and washed with diethyl ether (1 x 200 mL). The resulting aqueous layer was acidified with concentrated $HCl_{(aq)}$, and then extracted with DCM (5 x 200 mL). The organic layer was dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a pale yellow oil (4.1 g, 36% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.39-7.31 (m, 5H), δ 5.13 (ABq, J_{AB} = 12.4 Hz, 2H), δ 2.95-2.83 (m, 2H), δ 2.12-1.96 (m, 2H), δ 1.87-1.70 (m, 2H), δ 1.61-1.32 (m, 4H). ESI-MS $C_{15}H_{18}O_4$: $[M+H^+]^+$ calculated = 263.13, observed = 263.1.

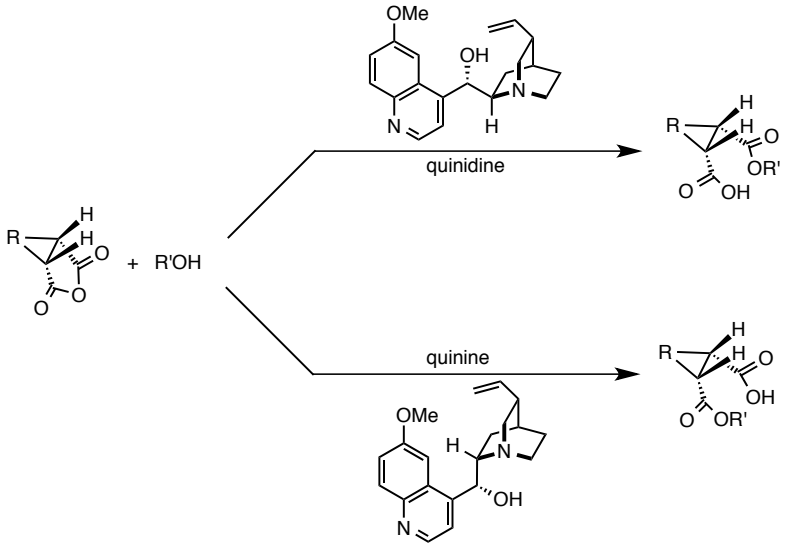

Figure A1.6. Alkaloid-induced asymmetric alcoholysis of meso-anhydrides.

H-Val-O*t*Bu. H-Val-O*t*Bu•HCl (6 g, 28.6 mmol) was dissolved in 80 mL EtOAc, and washed with saturated $NaHCO_{3(aq)}$ (3 x 80 mL). The organic layer was dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a brown oil (3.1 g, 62% yield).

BnO-(1*S*,2*R*)-CHDA-Val-O*t*Bu. BnO-(1*R*,2*S*)-CHDA-OH (4.1 g, 15.6 mmol), H-Val-O*t*Bu (3.1 g, 17.8 mmol), EDC-HCl (4.2 g, 21.8 mmol), and DMAP (cat.) were dissolved in 150 mL DCM, and stirred for 2 days. The solution was washed with 0.1 M HCl (3 x 500 mL), and every aqueous layer was back-extracted with DCM. Then the organic layer

was washed with saturated $NaHCO_{3(aq)}$ (2 x 500 mL), followed by brine (1 x 500 mL). The organic layer was then dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a yellow oil with white crystals. This was purified by column chromatography on silica gel (4:6 EtOAc:hexane, R_f = 0.71), to yield a yellow oil (3.8 g, 59% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.33 (m, 5H), δ 6.19 (d, J = 8.5 Hz, 1H), δ 5.15, 5.07 (ABq, J_{AB} = 12.5 Hz, 2H), δ 4.43 (dd, J = 8.5, 4.4 Hz, 1H), δ 2.90-2.73 (m, 2H), δ 2.10 (m, 3H), δ 1.87-1.32 (m, 6H), δ 1.46 (s, 9H). ESI-MS $C_{24}H_{35}NO_5$: $[M+H^+]^+$ calculated = 418.26, observed = 418.2.

HO-(1*S*,2*R*)-CHDA-Val-OtBu. BnO-(1*S*,2*R*)-CHDA-Val-O*t*Bu (3.8 g, 9.1 mmol) was dissolved in 120 mL MeOH. Pd/C Degussa (2 g) was added, the flask was purged with $N_{2(g)}$, and then two $H_{2(g)}$ balloons were attached. The reaction was stirred for 12 hours, and then filtered through celite. The product was dried to yield a yellow oil (2.5 g, 83% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.49 (d, J = 8.6 Hz, 1H), δ 4.43 (dd, J = 8.6, 4.4 Hz, 1H), δ 2.86 (m, 2H), δ 2.71 (m, 2H), δ 2.26-2.07 (m, 2H), δ 2.07-1.90 (m, 1H), δ 1.83-1.55 (m, 4H), δ 1.47 (s, 9H), δ 0.93 (d, J = 6.6 Hz, 3H), 0.90 (d, J = 6.8 Hz, 3H). ESI-MS $C_{17}H_{29}NO_5$: $[M+H^+]^+$ calculated = 328.21, observed = 328.1.

Boc-Gly-OAllyl. Boc-Gly-OH (8 g, 45.7 mmol) and DIEA (16 mL, 91.9 mmol) were dissolved in an excess of allyl bromide (40 mL, 462.2 mmol), and heated at reflux (70 °C) with stirring for 1 hour. The solution changed colors from white to red over this time. The red solution (which contained an aqueous-soluble white precipitate) was then cooled to room temperature, and mixed with 400 mL EtOAc. The organic layer was then washed with 0.1 M HCl (2 x 200 mL), saturated $NaHCO_{3(aq)}$ (2 x 200 mL), and brine (2 x 200 mL). The organic layer was then dried over anhydrous Na_2SO_4, filtered, and concentrated under

reduced pressure, to yield a red oil. The product was purified by column chromatography on silica gel (2:1 hexane:EtOAc, R_f = 0.67), to yield a yellow oil (10.4 g, quantitative yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 5.92 (ddt, J = 17.2, 10.3, 5.8 Hz, 1H), δ 5.34 (dq, J = 17.2, 1.5 Hz, 1H), δ 5.27 (dq, J = 10.4, 1.3 Hz, 1H), δ 5.02 (broad s, 1H), δ 4.65 (dt, J = 5.6, 1.4 Hz, 2H), δ 3.95 (d, J = 5.6 Hz, 2H), δ 1.46 (s, 9H). ESI-MS $C_{10}H_{17}NO_4$: $[M+H^+]^+$ calculated = 216.13, observed = 216.1.

TFA•H_2N-Gly-OAllyl. Boc-Gly-OAllyl (10.4 g, 48.1 mmol) was dissolved in 50 mL of 1:1 v/v TFA:DCM, and stirred for 20 min. The crude product was dried under reduced pressure, and residual TFA was azeotropically removed with $CHCl_3$ (3 x 150 mL), to yield a yellow oil (21.3 g, quantitative yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 5.88 (ddt, J = 17.2, 10.3, 6.0 Hz, 1H), δ 5.35 (dq, J = 17.1, 1.4 Hz, 1H), δ 5.31 (dq, J = 10.3, 1.1 Hz, 1H), δ 4.70 (dt, J = 6.0, 1.3 Hz, 2H), δ 3.90 (s, 2H).

OAllyl-Gly-(1R,2S)-CHDA-Val-OtBu. HO-(1*S*,2*R*)-CHDA-Val-OtBu (2.5 g, 7.5 mmol), H_2N-Gly-OAllyl (1.7 g, 15.0 mmol), PyBOP (7.8 g, 15.0 mmol), and DIEA (6.5 mL, 37.4 mmol) were dissolved in 150 mL DCM. The reaction was stirred for 1-2 days. The solution was washed with 0.1 M HCl (3 x 400 mL), saturated $NaHCO_{3(aq)}$ (3 x 400 mL), and brine (1 x 400 mL). The organic layer was then dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a runny, bright yellow oil. The product was purified by column chromatography on silica gel (2:1 EtOAc:hexane, R_f = 0.51), to yield a white solid (2.3 g, 72% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.61 (t, J = 4.7 Hz, 1H), δ 6.34 (d, J = 8.2 Hz, 1H), δ 5.91 (ddt, J = 16.9, 10.6, 5.8 Hz, 1H), δ 5.32 (d, J = 17.2 Hz, 1H), δ 5.26 (d, J = 10.5 Hz, 1H), δ 4.63 (broad d, J = 5.8 Hz, 2H), δ 4.37 (dd, J = 8.4, 4.5 Hz, 1H), δ 4.04 (AB part of ABX, 2H), δ 2.75 (m, 2H), δ 2.12 (m, 3H), δ 1.74 (m, 4H), δ 1.57 (m, 2H), δ 1.46

(s, 9H), δ 0.90 (d, J = 6.8 Hz, 3H), δ 0.89 (d, J = 6.9 Hz, 3H). ESI-MS $C_{22}H_{36}N_2O_6$: $[M+H^+]^+$ calculated = 425.27, observed = 425.2.

OAllyl-Gly-(1R,2S)-CHDA-Val-OH. OAllyl-Gly-(1R,2S)-CHDA-Val-OtBu (2.3 g, 5.4 mmol) was dissolved in 150 mL of 4 M HCl in dioxane (anhydrous), and stirred for 2 hours. The solution was concentrated under reduced pressure (10 mbar, 45 °C). The product was dried and resuspended in DCM 3x, and then dried to a light yellow solid (~2 g, quantitative yield). This crude product may be used without further purification in peptide syntheses. The product was purified by column chromatography on silica gel (0.95:0.05 v/v MeCN:AcOH, R_f = 0.39), to yield a light yellow solid (>1.2 g, >60% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.57 (d, J = 7.3 Hz, 1H), δ 6.52 (t, J = 4.9 Hz, 1H), δ 5.91 (ddt, J = 17.2, 10.5, 5.8 Hz, 1H), δ 5.33 (dd, J = 17.2, 1.4 Hz, 1H), δ 5.27 (dd, J = 10.4, 1.1 Hz, 1H), δ 4.64 (d, J = 5.8 Hz, 2H), δ 4.41 (dd, J = 8.0, 5.0 Hz, 1H), δ 4.03 (d, J = 5.2 Hz, 2H), δ 2.92 (m (ddd), 1H), δ 2.70 (m (ddd), 1H), δ 2.25 (m (dqq), 1H), δ 1.92-1.63 (m, 4H), δ 1.56-1.34 (m, 4H), δ 0.98 (d, J = 7.1 Hz, 3H), δ 0.96 (d, J = 7.0 Hz, 3H). ESI-MS $C_{18}H_{28}N_2O_6$: $[M+H^+]^+$ calculated = 369.20, observed = 369.1.

A1.3.3. Alloc-Glu-Val-^{D}Pro-DADME-Fmoc Linker.

The Alloc-Glu-Val-^{D}Pro-DADME-Fmoc linker was synthesized via the route depicted in **Figure A1.7**, which is similar to a previously reported synthetic route.[14] **Figures A1.21-A1.28** are ^{1}H NMR spectra from this synthesis, taken in $CDCl_3$, with an internal standard of TMS, on 300 MHz or 400 MHz spectrometers.

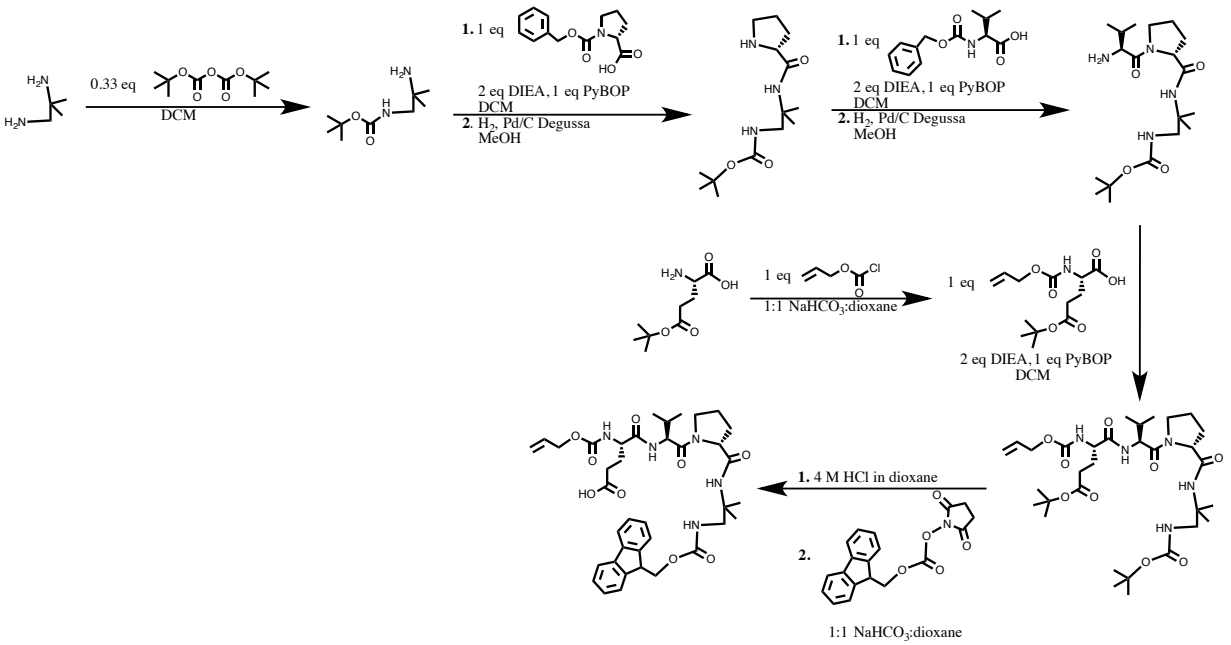

Figure A1.7. Synthetic route to alloc-Glu-Val-^{D}Pro-DADME-Fmoc linker.

H-DADME-Boc. 2-Methylpropane-1,2-diamine (a.k.a. 1,2-diamino-1,1-dimethylethane (DADME)) (5 mL, 48 mmol) was dissolved in 240 mL DCM. Boc_2O (3.7 g, 17 mmol) was dissolved in 150 mL DCM, and added dropwise (~100 drops/30 sec) to the stirring solution of DADME, and then the mixture was stirred for 12 hours. The crude product was transferred to a separatory funnel, and washed with saturated $NaHCO_{3(aq)}$. The resulting aqueous layer was back-extracted with DCM (2x). Organic layers were combined and dried to yield a white, crystalline solid (2.1 g, 66% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 4.93 (broad s, 1H), δ 3.00 (d, J = 6.3 Hz, 2H), δ 1.47 (s, 9H), δ 1.09 (s, 6H). ESI-MS $C_9H_{20}N_2O_2$: $[M+H^+]^+$ calculated = 189.16, observed = 189.1.

Cbz-^{D}Pro-DADME-Boc. Cbz-^{D}Pro-OH (2.8 g, 11.2 mmol) was activated for 2 min with PyBOP (5.8 g, 11.2 mmol) and DIEA (3.9 mL, 22.4 mmol), in 112 mL DCM. H-Dadme-Boc (2.1 g, 11.2 mmol) was added, and the reaction was stirred for 1-4 days. The product was purified by column chromatography on silica gel (3:1 EtOAc:hexane, R_f = 0.50) to yield a colorless oil (6.0 g, quantitative yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.35 (m, 5H),

δ 6.40 (broad s, 1H), δ 5.56 (broad s, 1H), δ 5.17 (m, 2H), δ 4.13 (m, 1H), δ 3.54 (m, 2H), δ 3.40-3.10 (m, 2H), δ 2.23-1.80 (m, 4H), δ 1.43 (s, 9H), δ 1.19 (m, 6H). ESI-MS $C_{22}H_{33}N_3O_5$: $[M+H^+]^+$ calculated = 420.25, observed = 420.2.

H-^{D}Pro-DADME-Boc. Cbz-^{D}Pro-DADME-Boc (5.9 g, 14.1 mmol) was dissolved in 180 mL MeOH. Pd/C Degussa (3 g) was added, the flask was purged with $N_{2(g)}$, and then two $H_{2(g)}$ balloons were attached. The reaction was stirred for 12 hours, and then filtered through celite. The product was dried to yield a yellow oil (3.4 g, 85% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.66 (s, 1H), δ 5.58 (t, J = 5.8 Hz, 1H), δ 3.65 (dd, J = 9.0, 5.5 Hz, 1H), δ 3.33 (m, 2H), δ 3.06-2.85 (m, 2H), δ 2.12 (m, 1H), δ 1.87 (m, 1H), δ 1.71 (m, 2H), δ 1.44 (s, 9H), δ 1.29 (s, 6H). ESI-MS $C_{14}H_{27}N_3O_3$: $[M+H^+]^+$ calculated = 286.22, observed = 286.2.

Cbz-Val-^{D}Pro-DADME-Boc. Cbz-Val-OH (3.0 g, 11.9 mmol) was activated for 2 min with PyBOP (6.2 g, 11.9 mmol) and DIEA (4.2 mL, 23.8 mmol), in 70 mL DCM. H-^{D}Pro-Dadme-Boc (3.4 g, 11.9 mmol) was added, and the reaction was stirred for 1-2 days. The product was purified by column chromatography on silica gel (3:1 EtOAc:hexane, R_f = 0.41) to yield a colorless oil (5.7 g, 92% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.33 (m, 5H), δ 6.05 (broad s, 1H), δ 5.73 (broad t, J = 7.5 Hz, 1H), δ 5.59 (broad d, 1H), δ 5.21-4.94 (m, 3H), δ 4.34-4.18 (m, 2H), δ 3.96 (m, 1H), δ 3.58 (m, 1H), δ 3.36 (m, 2H), δ 2.28-1.81 (m, 4H), δ 1.60 (s, 6H), δ 1.41 (s, 9H), δ 0.98 (d, J = 6.9 Hz, 3H), δ 0.96 (d, J = 7.1 Hz, 3H). ESI-MS $C_{27}H_{42}N_4O_6$: $[M+H^+]^+$ calculated = 519.32, observed = 519.3.

H-Val-^{D}Pro-DADME-Boc. Cbz-Val-^{D}Pro-DADME-Boc (5.7 g, 10.9 mmol) was dissolved in 180 mL MeOH. Pd/C Degussa (3 g) was added, the flask was purged with $N_{2(g)}$, and then two $H_{2(g)}$ balloons were attached. The reaction was stirred for 12 hours, and then filtered through celite. The product was dried to yield a white oil (4.3 g, quantitative yield).

^{1}H NMR (300 MHz, $CDCl_3$): δ 6.49 (s, 1H), δ 5.63 (t, 1H), δ 4.31 (m, 1H), δ 3.70 (m, 1H), δ 3.63-3.29 (m, 3H), δ 3.26 (m, 1H), δ 2.35-1.78 (m, 5H), δ 1.61 (broad s, 2H), δ 1.45 (s, 9H), δ 1.30 (s, 3H), δ 1.28 (s, 3H), δ 0.98 (d, J = 6.8 Hz, 3H), δ 0.96 (d, J = 6.8 Hz, 3H). ESI-MS $C_{19}H_{36}N_4O_4$: $[M+H^+]^+$ calculated = 385.28, observed = 385.2.

Alloc-Glu(O*t*Bu)-OH. H-Glu(O*t*Bu)-OH (5.0 g, 29.2 mmol), allyl chloroformate (3.1 mL, 29.2 mmol), and $NaHCO_3$ (146 mL of a saturated aqueous solution) were stirred in a flask with 146 mL dioxane. A white precipitate was observed in the reaction mixture. After 1 day, the solution was acidified with 12 M HCl (solution became transparent), and then extracted with DCM (3 x 400 mL). The combined organic layers were dried over anhydrous Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a runny, light yellow oil (10.5 g, quantitative yield). (Aside: R_f = 0.24, on silica gel with 2:1 EtOAc:hexane.) ^{1}H NMR (300 MHz, $CDCl_3$): δ 5.90 (m (ddt), 1H), δ 5.50 (d, J = 8.0 Hz, 1H), δ 5.31 (dd, J = 17.1, 1.6 Hz, 1H), δ 5.22 (dd, J = 10.4, 1.3 Hz, 1H), δ 4.58 (d, J = 5.7 Hz, 2H), δ 4.37 (m, 1H), δ 2.50-2.28 (m, 2H), δ 2.19 (dt, J = 12.2, 7.1 Hz, 1H), δ 2.00 (dt, J = 14.5, 7.4 Hz, 1H), δ 1.45 (s, 9H). ESI-MS $C_{13}H_{21}NO_6$: $[M+H^+]^+$ calculated = 288.15, observed = 288.1.

Alloc-Glu(O*t*Bu)-Val-^{D}Pro-DADME-Boc. Alloc-Glu(O*t*Bu)-OH (3.2 g, 11.0) was activated for 2 min with PyBOP (5.7 g, 11.0 mmol) and DIEA (3.8 mL, 22.0 mmol), in 110 mL DCM. H-Val-^{D}Pro-DADME-Boc (4.2 g, 11.0 mmol) was added, and the reaction was stirred for 1 day. A white suspension was observed in the reaction mixture. The mixture was washed with saturated $NaHCO_{3(aq)}$ (2 x 400 mL), which solubilized the white precipitate, and the resulting aqueous layers were back-extracted with DCM. The combined organic layers were then washed with brine (1 x 400 mL), dried over anhydrous

Na_2SO_4, filtered, and concentrated under reduced pressure, to yield a golden oil. The product was purified by column chromatography on silica gel (7:3 EtOAc:hexane, R_f = 0.24) to yield a white solid (5.0 g, 69% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.18 (broad d, 1H), δ 6.61 (broad s, 1H), δ 5.89 (ddt, 1H), δ 5.75 (broad d, J = 6.8 Hz, 1H), δ 5.62 (broad t, J = 6.6 Hz, 1H), δ 5.29 (dd, J = 17.2, 1.4 Hz, 1H), δ 5.20 (d, J = 10.5 Hz, 1H), δ 4.56 (d, J = 5.5 Hz, 2H), δ 4.45-4.20 (m, 3H), δ 3.89 (m, 1H), δ 3.52 (m, 2H), δ 3.30 (d, J = 6.8 Hz, 2H), δ 2.46 (m, 2H), δ 2.26 (m, 1H), δ 2.15-1.83 (m, 5H), δ 1.63 (s, 6H), δ 1.45 (s, 18H), δ 1.01 (d, J = 6.8 Hz, 3H), δ 0.99 (d, J = 6.8 Hz, 3H). ESI-MS $C_{32}H_{55}N_5O_9$: $[M+H^+]^+$ calculated = 654.41, observed = 654.4.

Alloc-Glu-Val-^{D}Pro-DADME-Fmoc. Alloc-Glu(O*t*Bu)-Val-^{D}Pro-DADME-Boc (5.0 g, 7.6 mmol) was dissolved in 75 mL of 4 M HCl in dioxane (anhydrous), and stirred for 2 hours. The solution was concentrated under reduced pressure (10 mbar, 45 °C). The resulting Boc-deprotected product was dissolved with $NaHCO_3$ (40 mL of a saturated aqueous solution) in 40 mL dioxane. Fmoc-OSu (2.6 g, 7.6 mmol) was added, and the reaction was stirred for 4 hours. The solution was then acidified with 12 N HCl. The product was extracted with DCM (5 x 150 mL), dried over anhydrous $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting yellow solid (~6 g, quantitative yield) may be used without further purification in peptide syntheses. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.80-7.32 (m (AA'XX'), 4H), δ 7.80-7.26 (m (AA'XX'), 4H), δ 6.82 (d, J = 7.3 Hz, 1H), δ 6.36 (s, 1H), δ 6.13 (m (X part of ABX), 1H), δ 5.87 (m (ddt), 1H), δ 5.72 (d, J = 7.7 Hz, 1H), δ 5.28 (d, J = 17.4 Hz, 1H), δ 5.18 (d, J = 10.1 Hz, 1H), δ 4.57 (dd, J = 7.5 Hz, 1H), δ 4.53 (d, J = 4.9 Hz, 2H), δ 4.48 (m (dd), 1H), δ 4.27 (m (ddd), 1H), δ 4.45-4.17 (m, 2H), δ 3.92 (m (X part of ABX), 1H), δ 3.58 (m (AB part of ABX), 2H), δ 3.37 (m (AB part of ABX), 2H), δ 2.40 (m

(BB' part of AA'BB'), 2H), δ 2.02 (m (AA' part of AA'BB'), 2H), δ 2.02 (m (dqq), 1H), δ 2.20-1.90 (m, 4H), δ 1.35 (s, 3H), δ 1.27 (s, 3H), δ 0.96 (d, J = 6.4 Hz, 3H), δ 0.94 (d, J = 6.5 Hz, 3H). ESI-MS $C_{38}H_{49}N_5O_9$: $[M+H^+]^+$ calculated = 720.36, observed = 720.3.

A1.3.4. Alloc-Glu-Val-^{L}Pro-DADME-Fmoc Linker.

The Alloc-Glu-Val-^{L}Pro-DADME-Fmoc linker was synthesized via the route depicted in **Figure A1.7**, except using Cbz-^{L}Pro-OH instead of Cbz-^{D}Pro-OH. **Figures A1.29-A1.34** are ^{1}H NMR spectra from this synthesis, taken in $CDCl_3$, with an internal standard of TMS, on a 400 MHz spectrometer.

A1.3.5. Succinyl-Gly Linker.

The succinyl-Gly linker was synthesized via the route depicted in **Figure A1.8**. **Figures A1.35-A1.37** are ^{1}H NMR spectra from this synthesis, taken in $CDCl_3$ (no TMS) on a 300 MHz spectrometer.

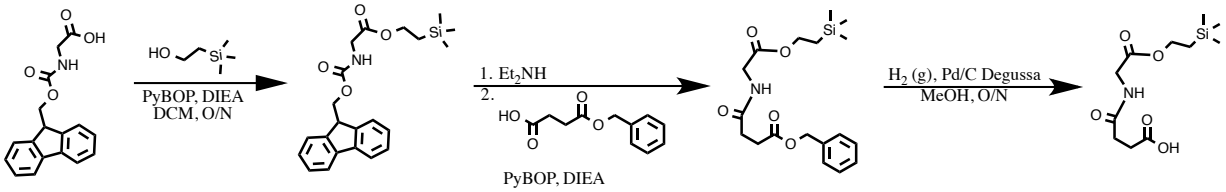

Figure A1.8. Synthetic route to succinyl-Gly linker.

Fmoc-Gly-OTMSE. Fmoc-Gly-OH (1.0 g, 3.4 mmol), DIEA (1.8 mL, 10.3 mmol), and PyBOP (1.7 g, 3.3 mmol) were dissolved in 20 mL DCM. 2-(Trimethylsilyl)ethanol (0.5 mL, 3.3 mmol) was added, and the reaction was stirred for 12 hours. A white precipitate was observed on addition of 2-(trimethylsilyl)ethanol, and this precipitate disappeared over time. The crude product was dried to a transparent, yellow oil. The product was purified by column chromatography on silica gel (9:1 hexane:EtOAc, R_f = 0.05), to yield a white, crystalline solid (0.9 g, 70% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.77 (~dt, J = 7.5, 1.0 Hz,

2H), δ 7.61 (~broad d, J = 7.4 Hz, 2H), δ 7.40 (~td, J = 7.5, 1.3 Hz, 2H), δ 7.31 (~td, J = 7.4, 1.2 Hz, 2H), δ 5.29 (t, J = 5.5 Hz, 1H), δ 4.41 (d, J = 7.1 Hz, 2H), δ 4.27 (XX' part of AA'XX', 2H), δ 4.24 (t, J = 7.0 Hz, 1H), δ 3.98 (d, J = 5.5 Hz, 2H), δ 1.02 (AA' part of AA'XX', $J_{AA'}$ = 14.3, $J_{XX'}$ = 11.3, J_{AX} = 11.0, $J_{AX'}$ = 6.2 Hz, 2H), δ 0.05 (s, 9H).

BnO-Succinyl-Gly-OTMSE. Fmoc-Gly-OTMSE (0.8 g, 1.9 mmol) was combined with an excess of diethylamine (7.5 mL, 72.7 mmol) in 15 mL DCM, and the reaction was stirred for 1 hour. The crude product was dried and resuspended in 50 mL MeCN 6x, and then dried to a light yellow, opaque oil. This oil was resuspended in 20 mL DCM. Succinic acid monobenzyl ester (0.4 g, 2.0 mmol), DIEA (0.7 mL, 4.2 mmol), and PyBOP (1.1 g, 2.1 mmol) were added, and the reaction was stirred for 12 hours. The crude product was dried to yield an orange-yellow, opaque oil. The product was purified by column chromatography on silica gel (9:1 hexane:EtOAc, then 1:1 hexane:EtOAc; R_f = 0.13, for 7:3 hexane:EtOAc), to yield a transparent, light yellow oil (0.4 g, 59% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.35 (m, 5H), δ 6.13 (broad t, J = 4.2 Hz, 1H), δ 5.13 (s, 2H), δ 4.16 (XX' part of AA'XX', 2H), δ 4.00 (d, J = 5.0 Hz, 2H), δ 2.74 (BB' part of AA'BB', 2H), δ 2.57 (AA' part of AA'BB', 2H), δ 0.92 (AA' part of AA'XX', 2H), δ 0.05 (s, 9H).

HO-Succinyl-Gly-OTMSE. BnO-Succinyl-Gly-OTMSE (0.4 g, 1.0 mmol) was dissolved in 12 mL MeOH. Pd/C Degussa (0.2 g) was added, the flask was purged with $N_{2(g)}$, and then two $H_{2(g)}$ balloons were attached. The reaction was stirred for 12 hours, and then filtered through celite. The product was dried to yield a transparent, light yellow oil (0.3 g, 96% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.83 (broad t, J = 5.4 Hz, 1H), δ 4.23 (XX' part of AA'XX', 2H), δ 3.99 (d, J = 5.3 Hz, 2H), δ 2.64 (BB' part of AA'BB', 2H), δ 2.53 (AA' part of AA'BB', 2H), δ 1.00 (AA' part of AA'XX', 2H), δ 0.04 (s, 9H).

A1.3.6. ^{1}H NMR Spectra for Linker Syntheses

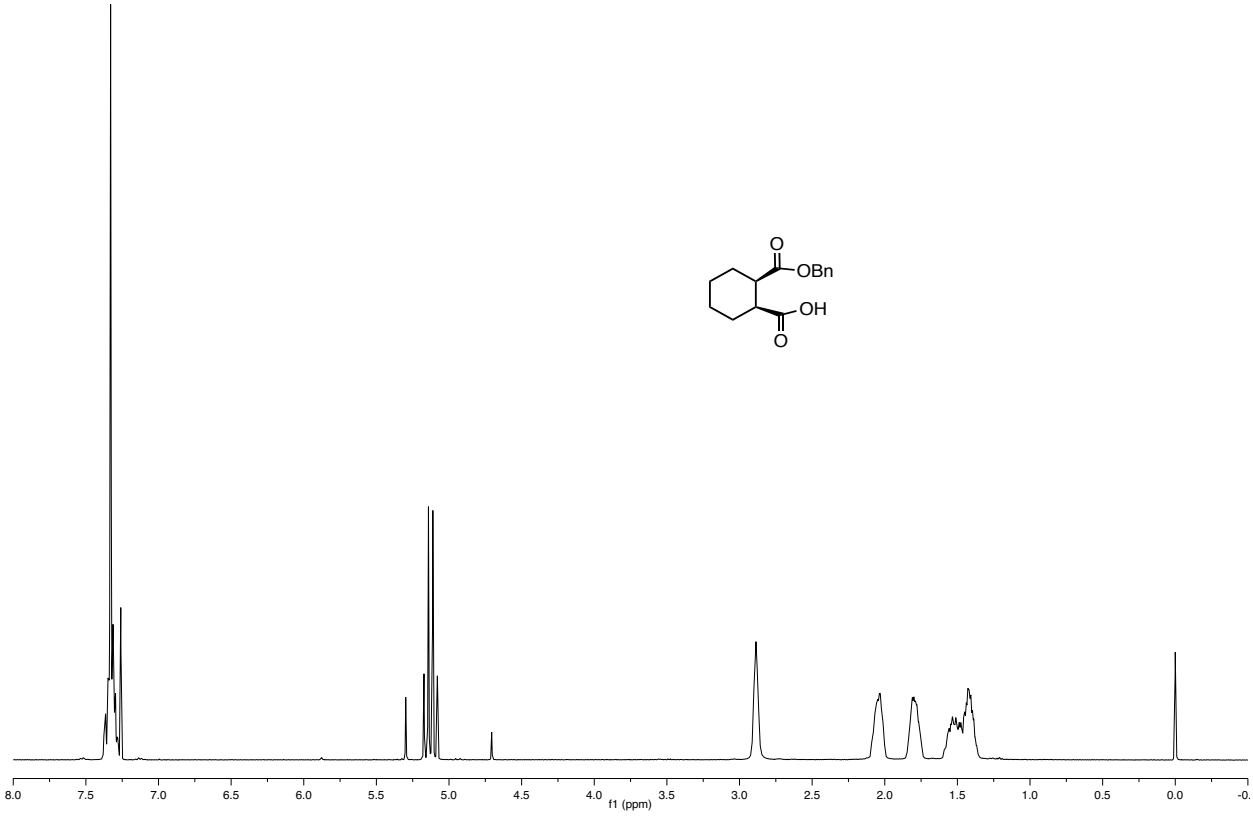

Figure A1.9. ^{1}H NMR of **BnO-(1*S*,2*R*)-CHDA-OH** in $CDCl_3$, 400 MHz.

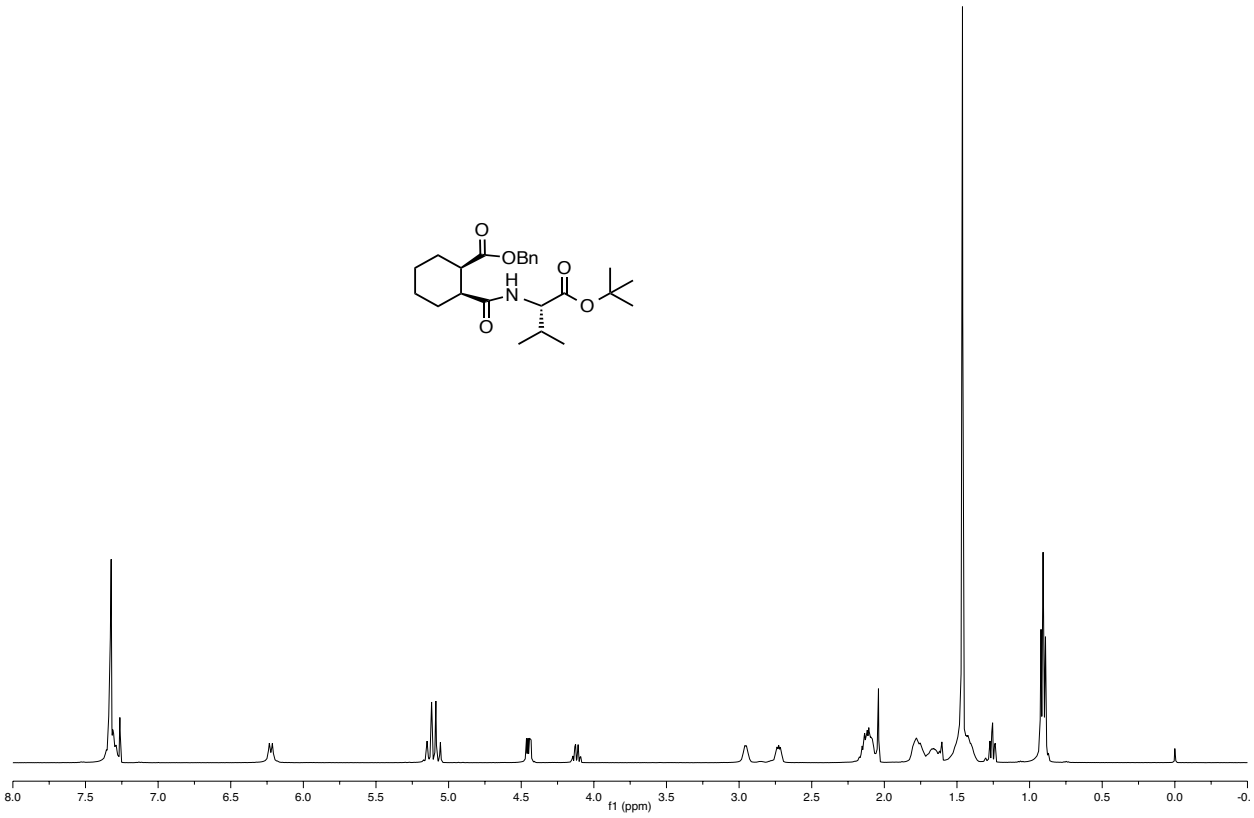

Figure A1.10. ^{1}H NMR of **BnO-(1*R*,2*S*)-CHDA-Val-O*t*Bu** in $CDCl_3$, 400 MHz.

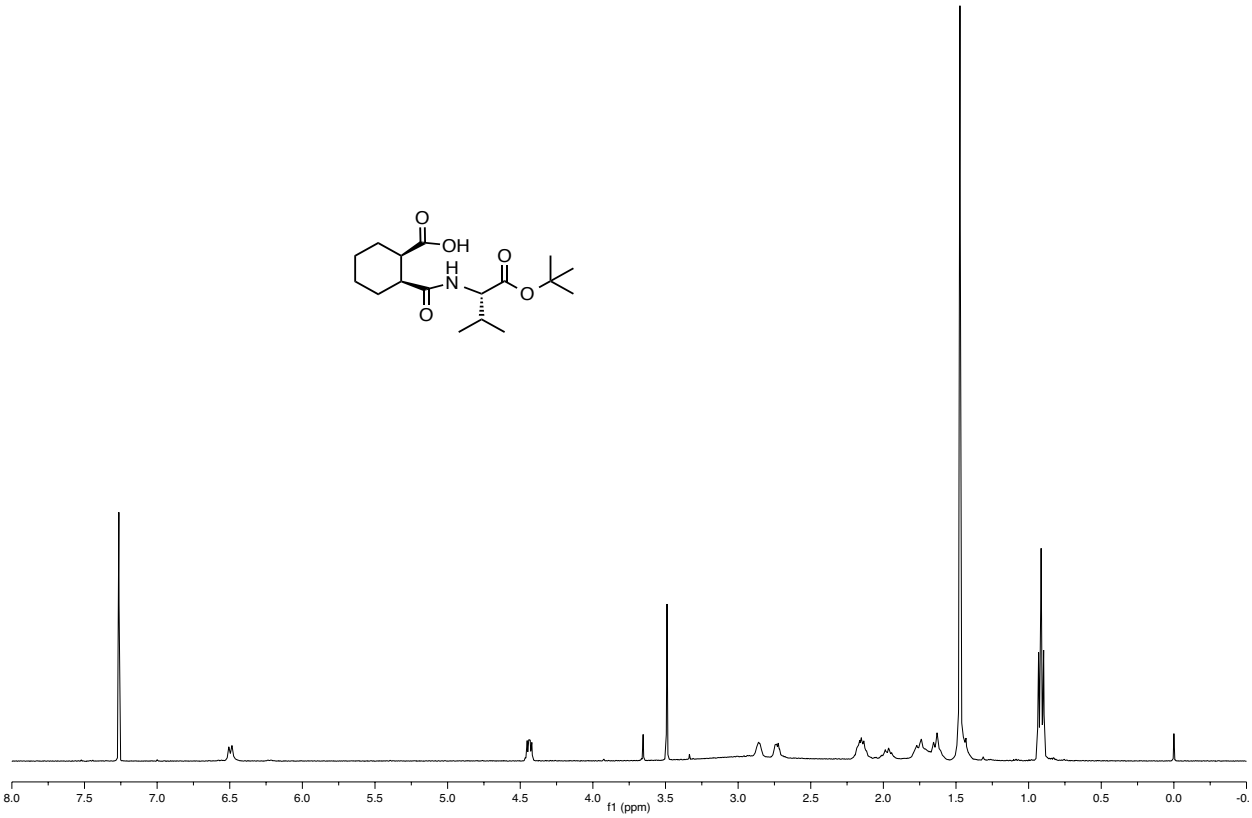

Figure A1.11. ^{1}H NMR of **HO-(1*R*,2*S*)-CHDA-Val-OtBu** in $CDCl_3$, 400 MHz.

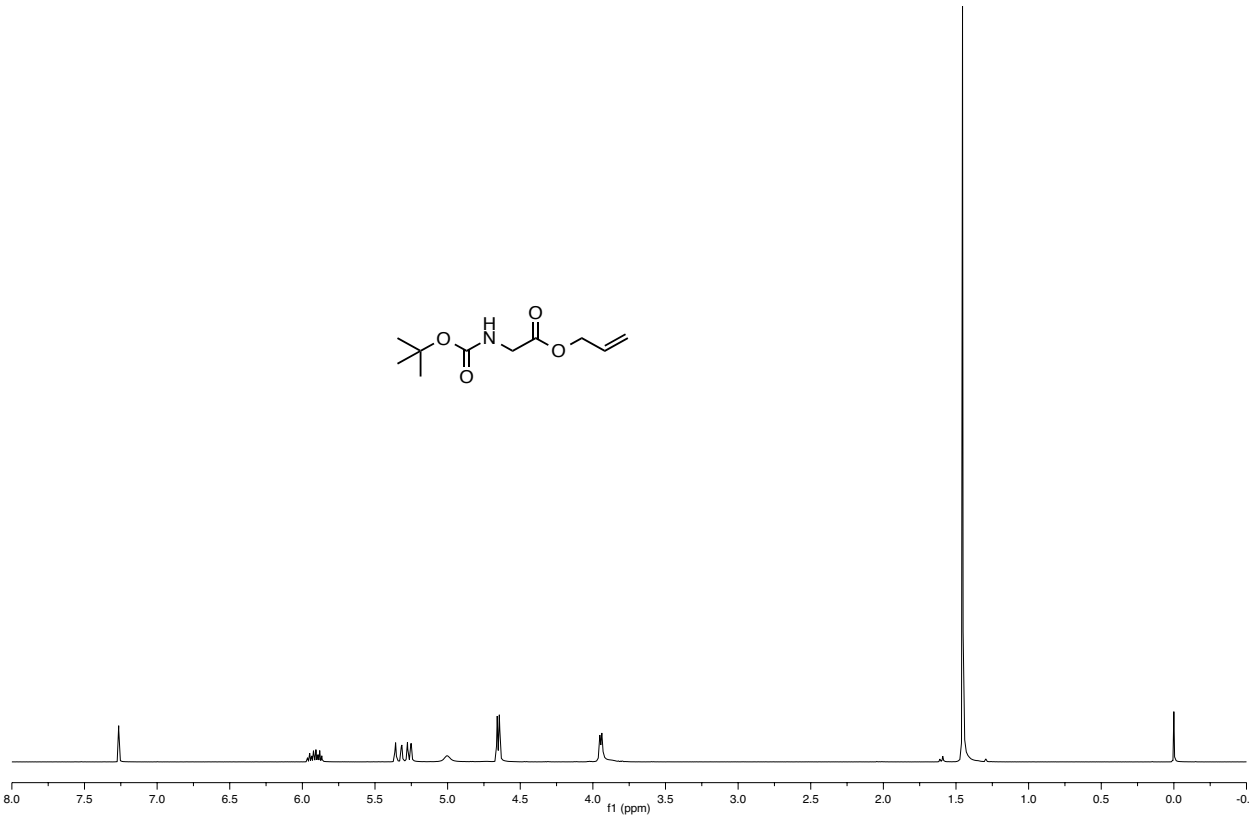

Figure A1.12. ^{1}H NMR of **Boc-Gly-OAllyl** in $CDCl_3$, 400 MHz.

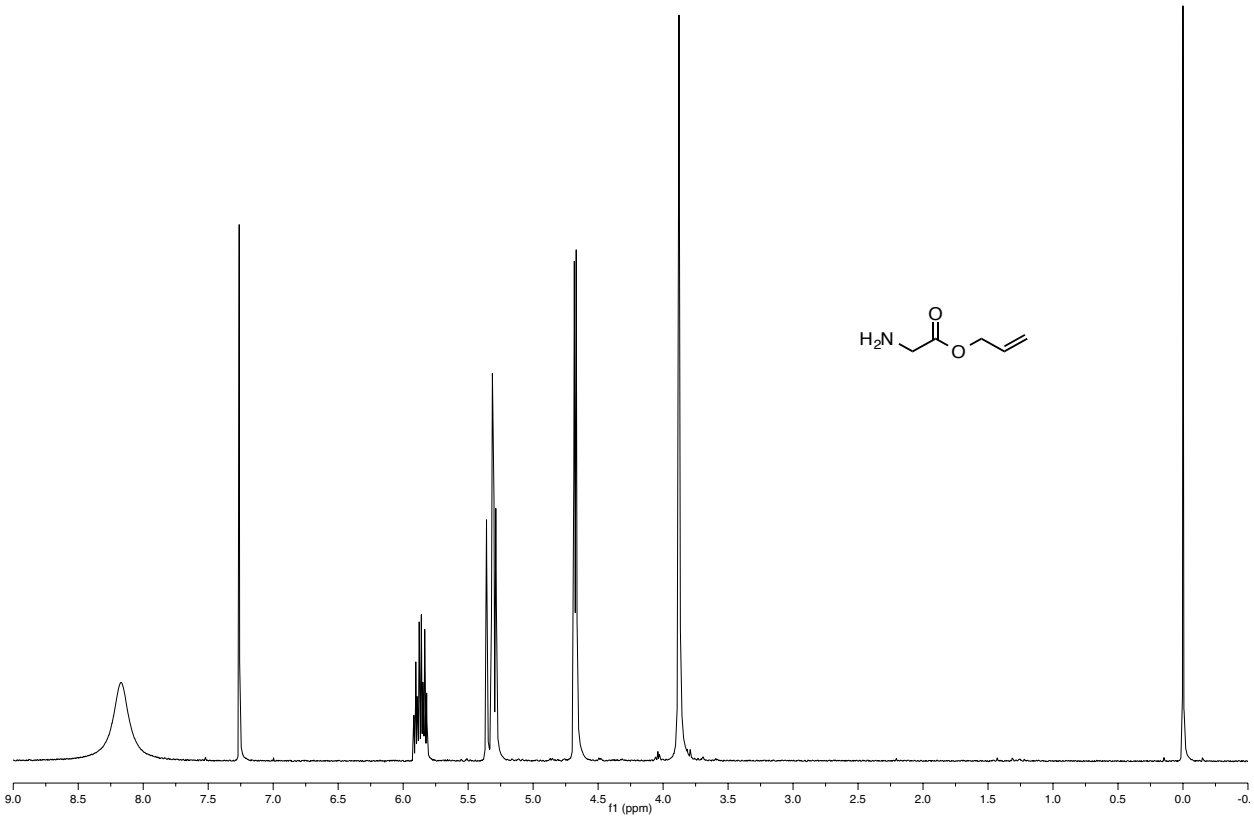

Figure A1.13. ^{1}H NMR of **TFA•H$_2$N-Gly-OAllyl** in $CDCl_3$, 400 MHz.

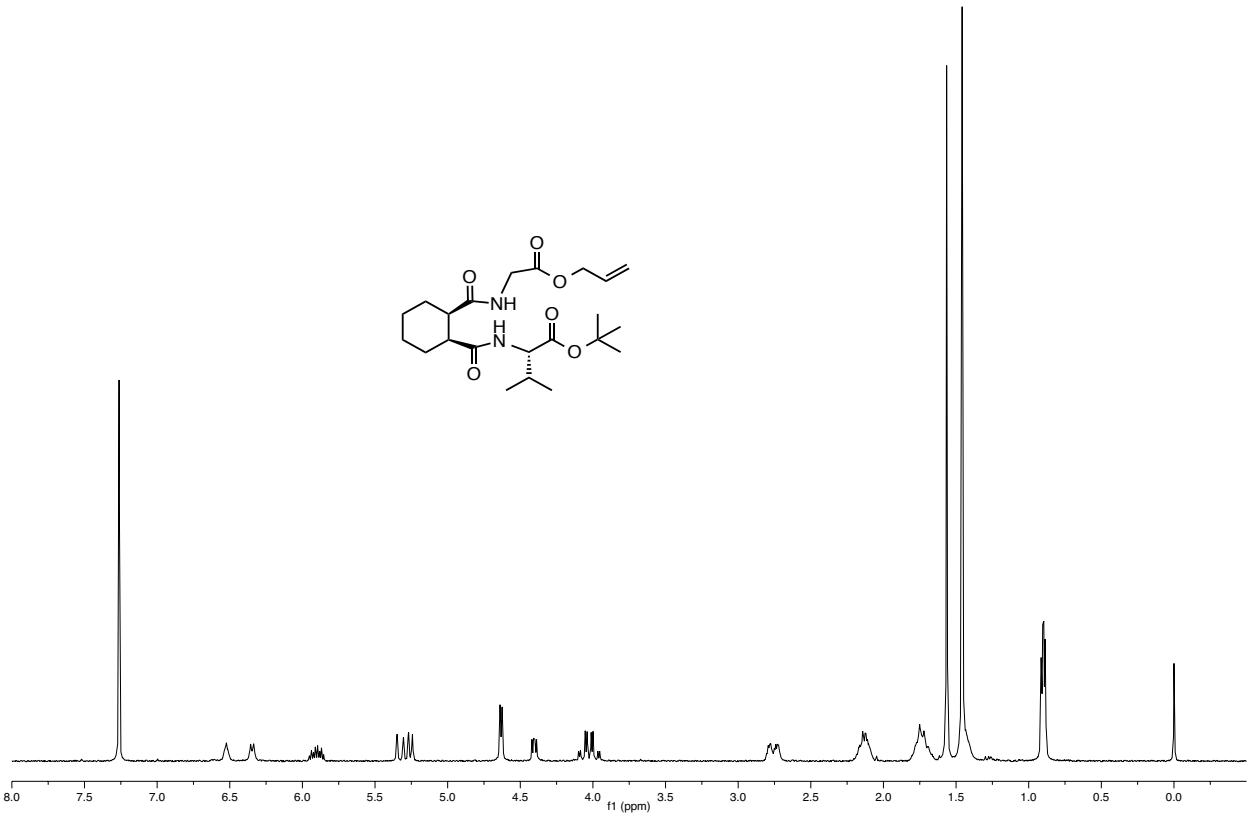

Figure A1.14. ^{1}H NMR of **OAllyl-Gly-(1S,2R)-CHDA-Val-OtBu** in $CDCl_3$, 400 MHz.

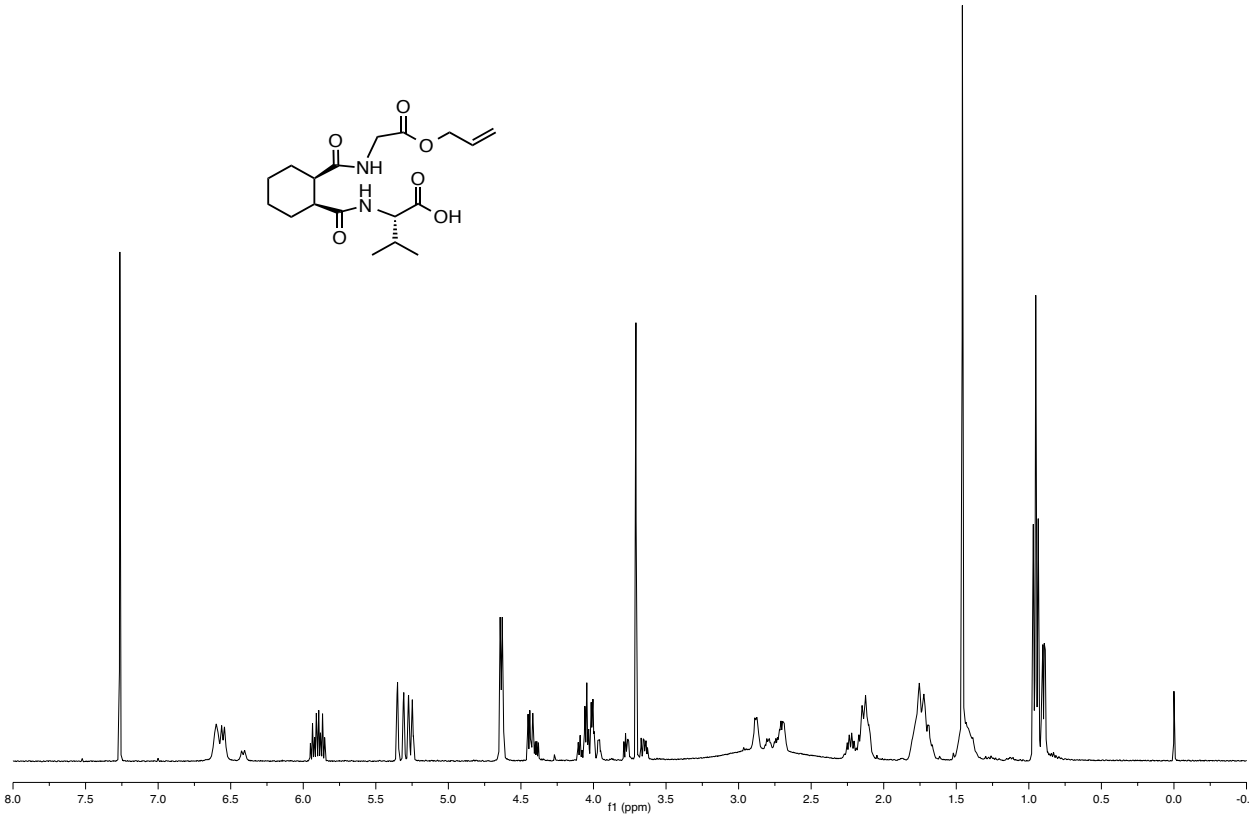

Figure A1.15. ^{1}H NMR of **OAllyl-Gly-(1S,2R)-CHDA-Val-OH** in $CDCl_3$, 400 MHz.

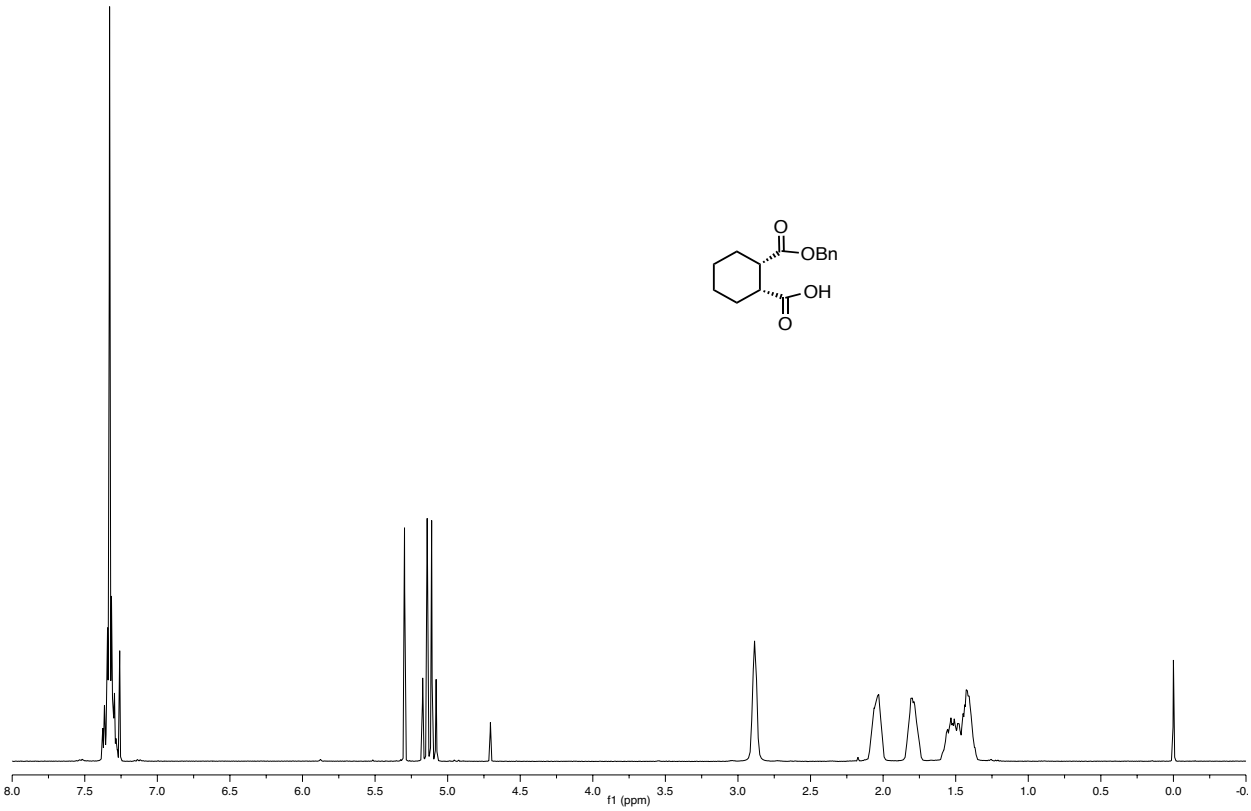

Figure A1.16. ^{1}H NMR of **BnO-(1*R*,2*S*)-CHDA-OH** in $CDCl_3$, 400 MHz.

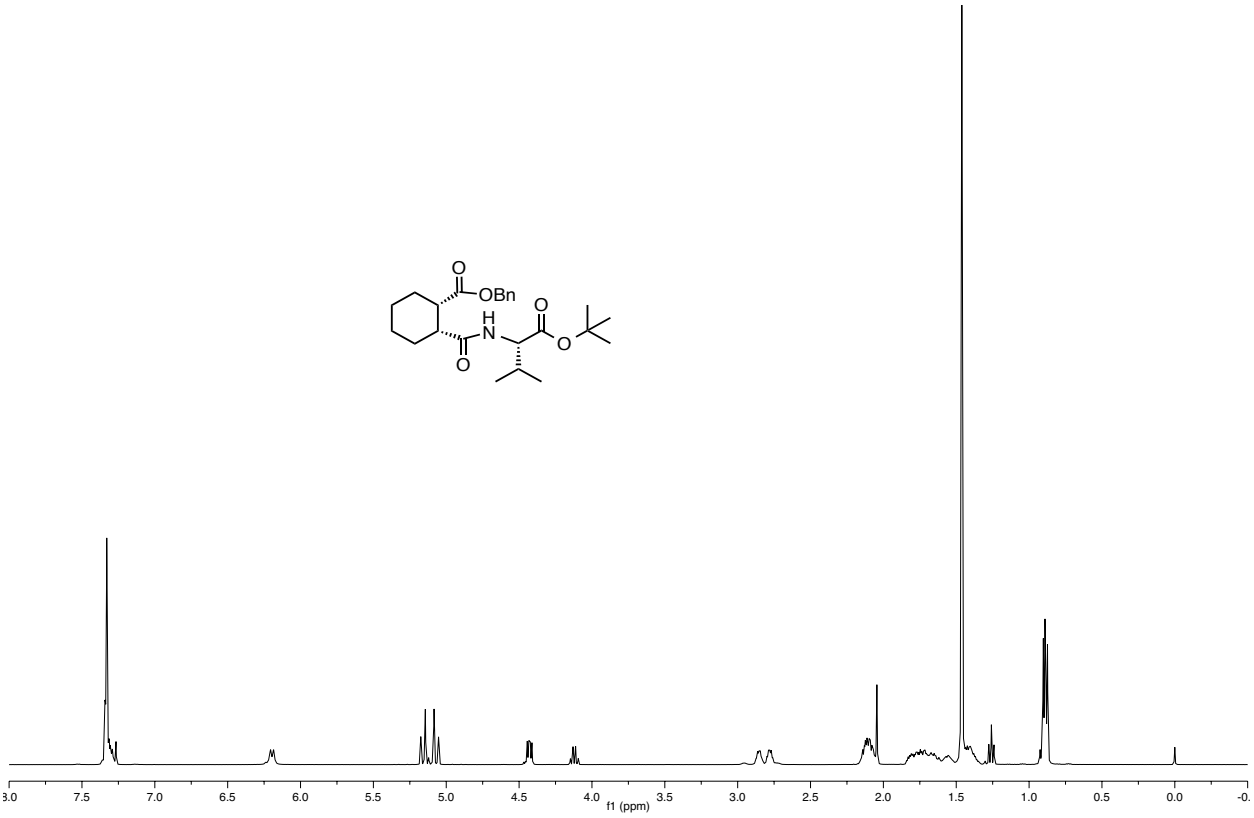

Figure A1.17. ^{1}H NMR of **BnO-(1*S*,2*R*)-CHDA-Val-O*t*Bu** in $CDCl_3$, 400 MHz.

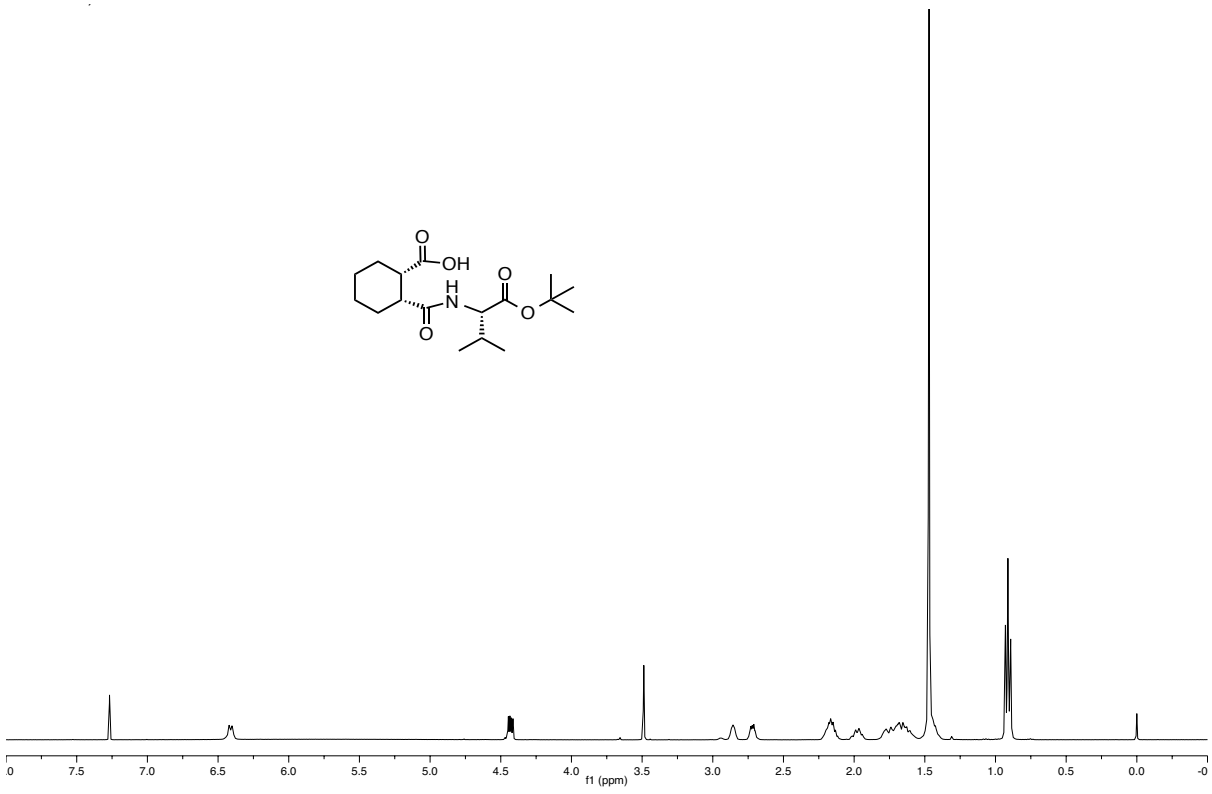

Figure A1.18. ^{1}H NMR of **HO-(1*S*,2*R*)-CHDA-Val-OtBu** in $CDCl_3$, 400 MHz.

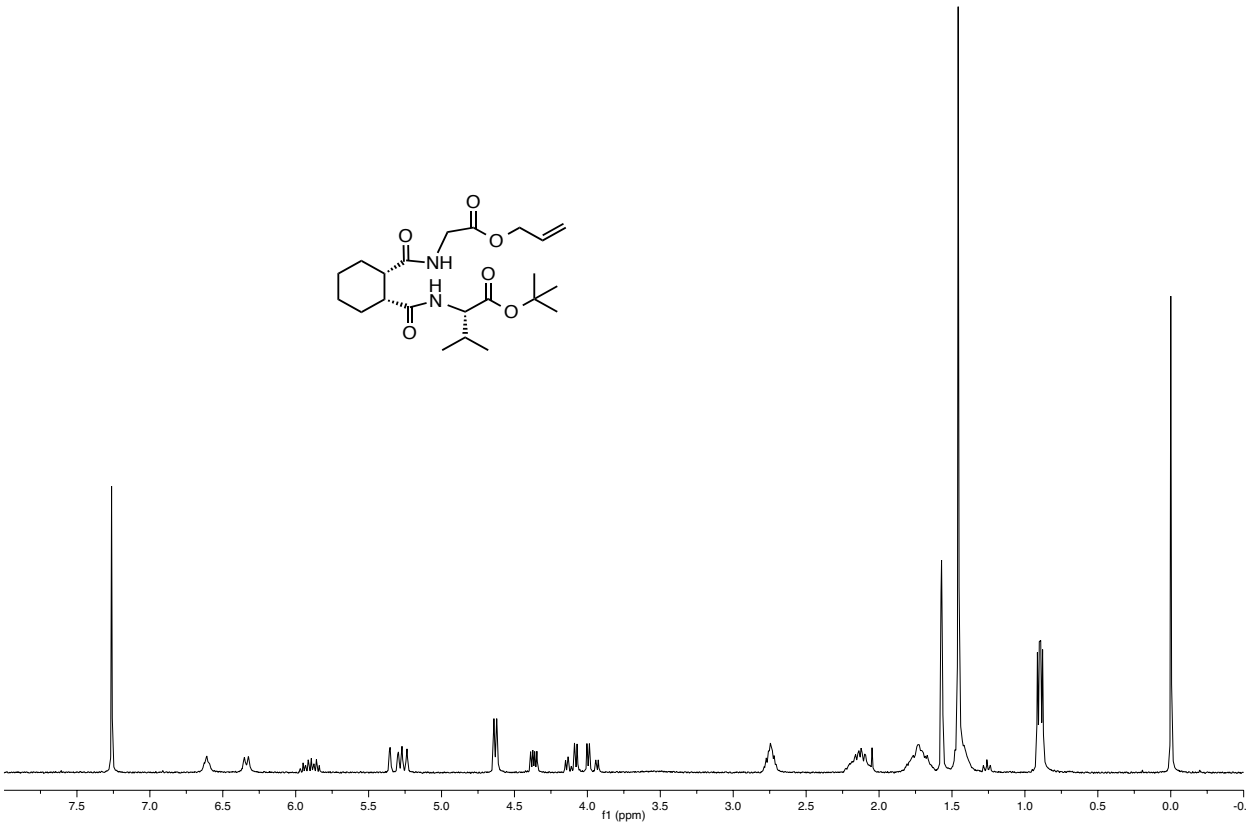

Figure A1.19. ^{1}H NMR of **OAllyl-Gly-(1R,2S)-CHDA-Val-OtBu** in $CDCl_3$, 300 MHz.

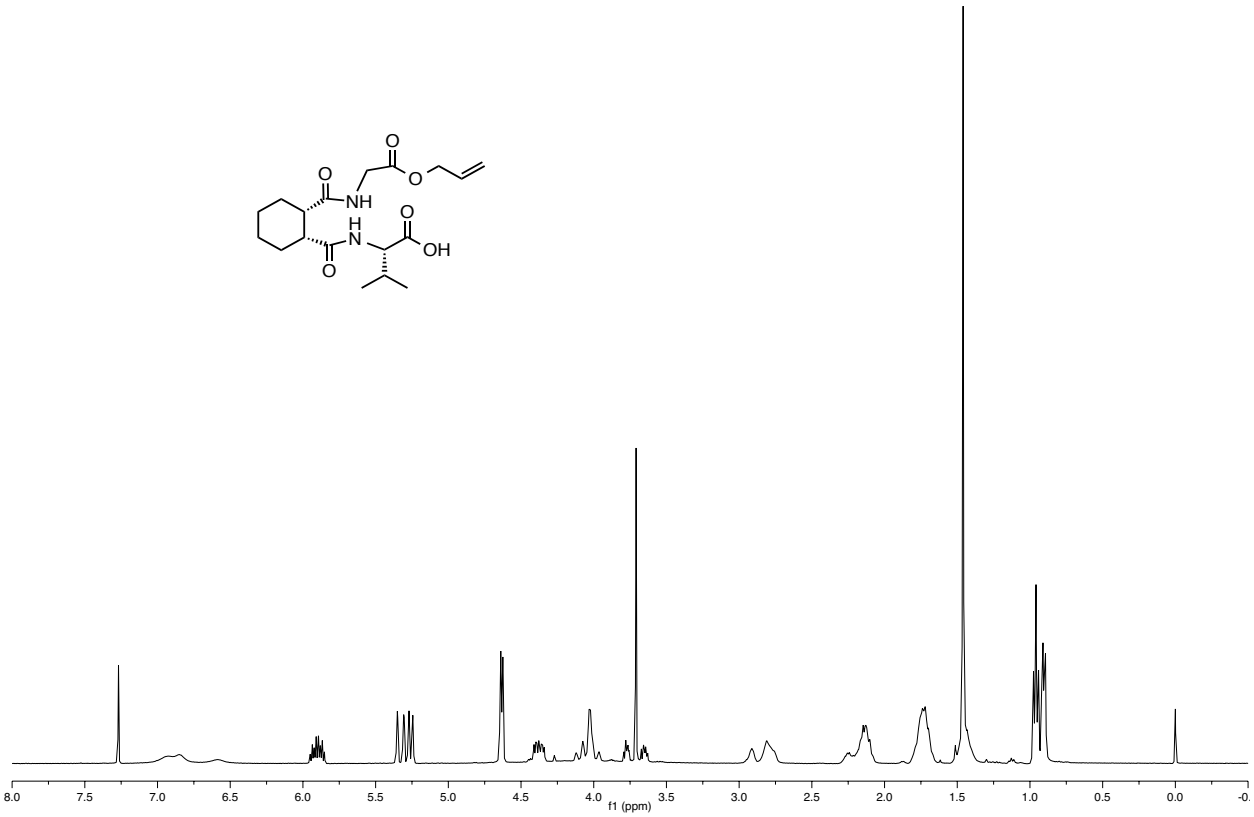

Figure A1.20. ^{1}H NMR of **OAllyl-Gly-(1R,2S)-CHDA-Val-OH** in $CDCl_3$, 400 MHz.

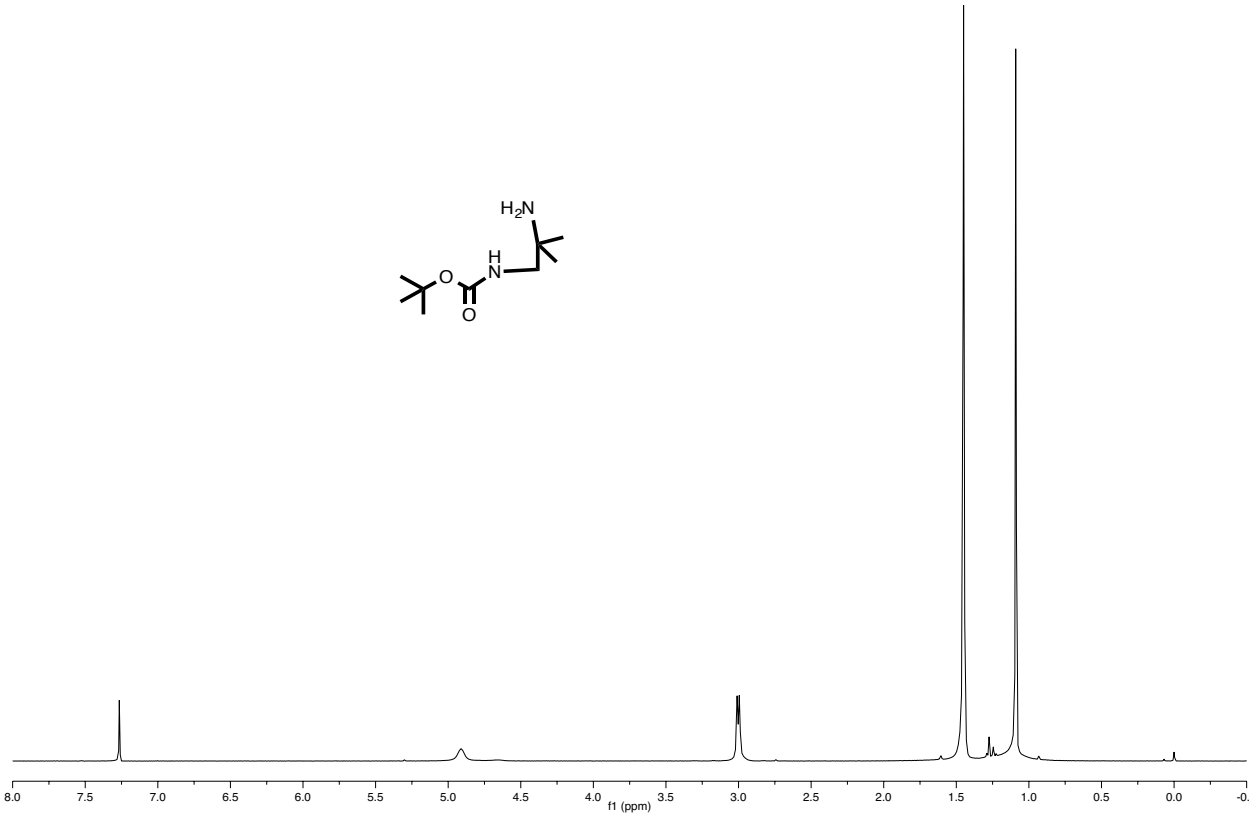

Figure A1.21. ^{1}H NMR of **H-DADME-Boc** in $CDCl_3$, 400 MHz.

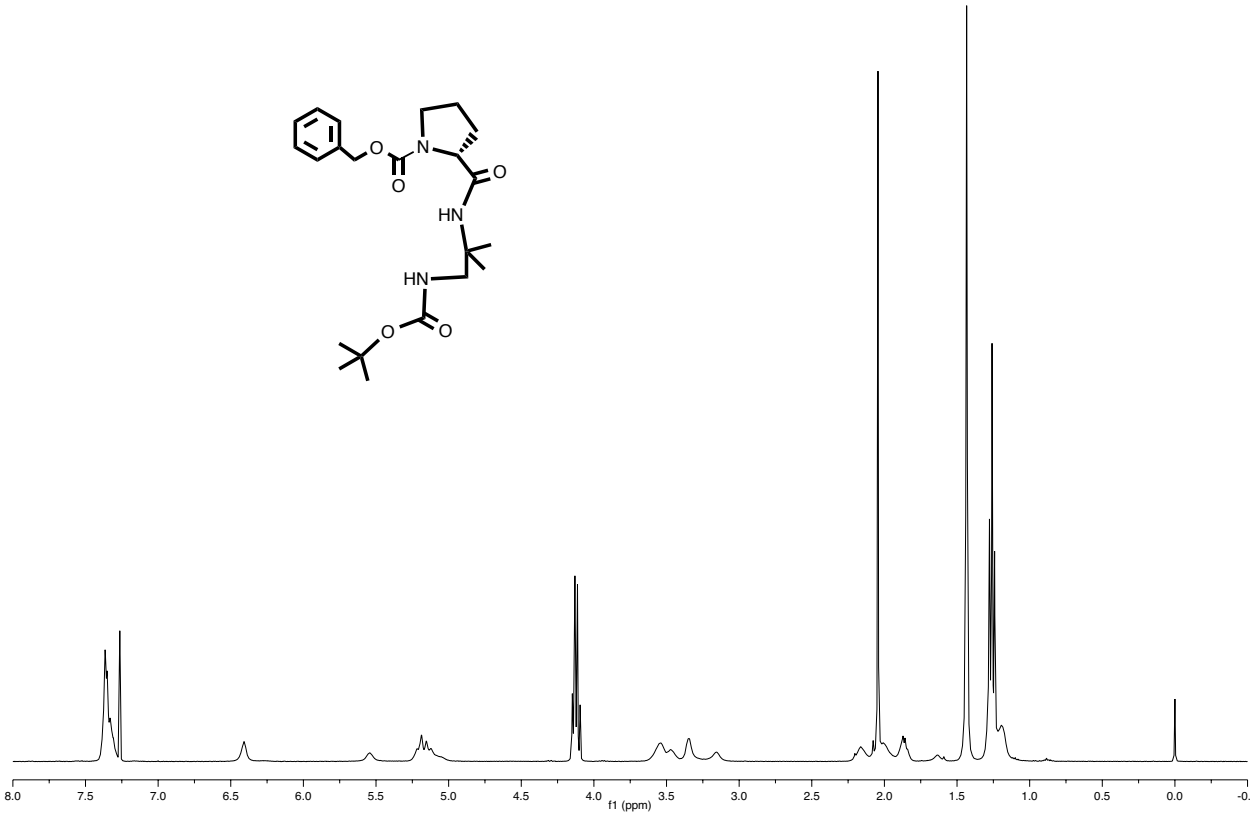

Figure A1.22. ^{1}H NMR of **Cbz-^{D}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

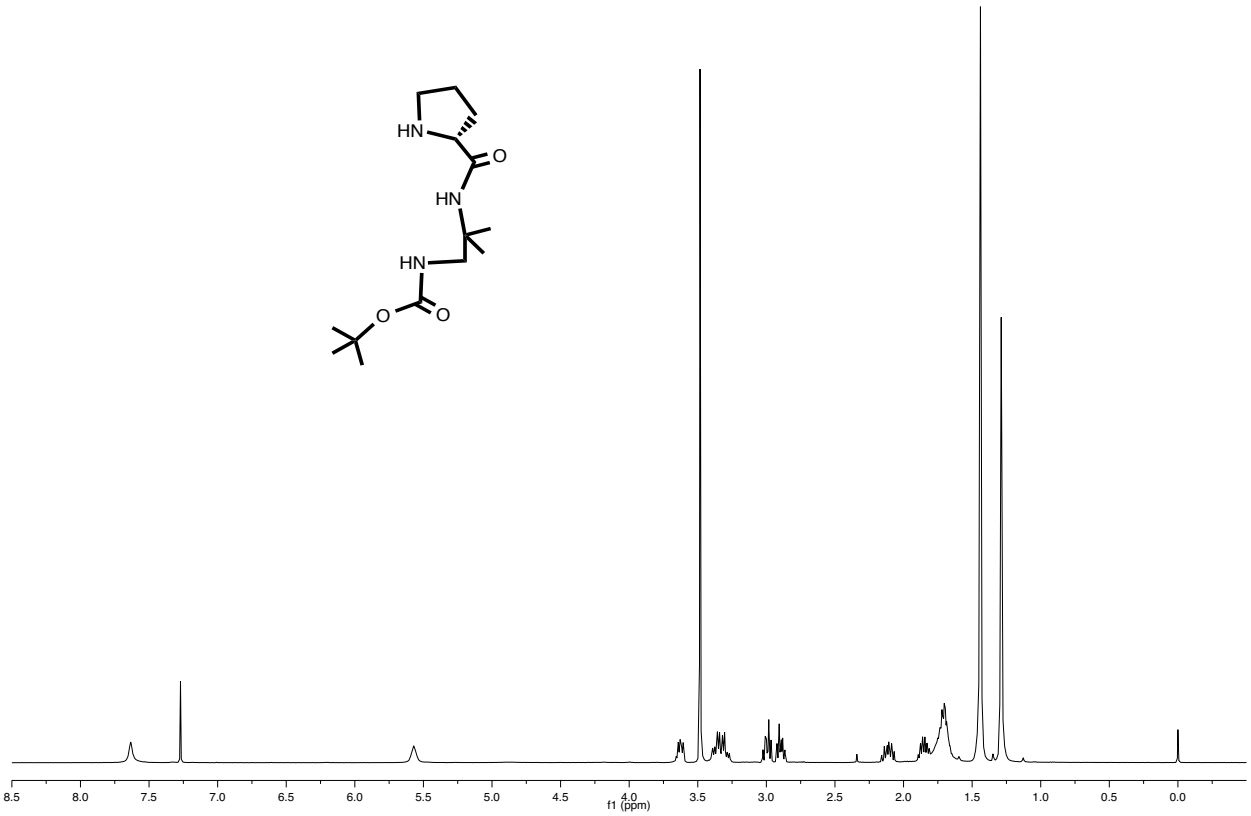

Figure A1.23. ^{1}H NMR of **H-^{D}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

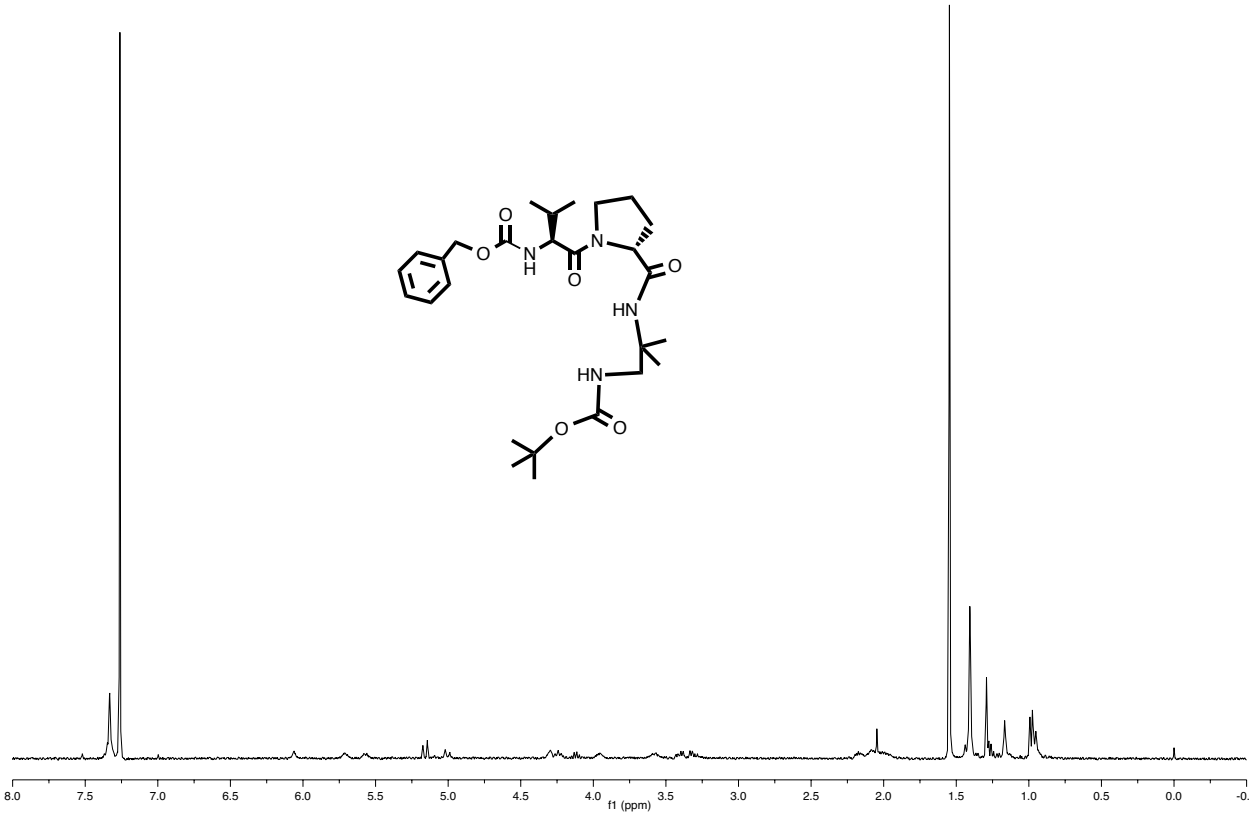

Figure A1.24. ^{1}H NMR of **Cbz-Val-^{D}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

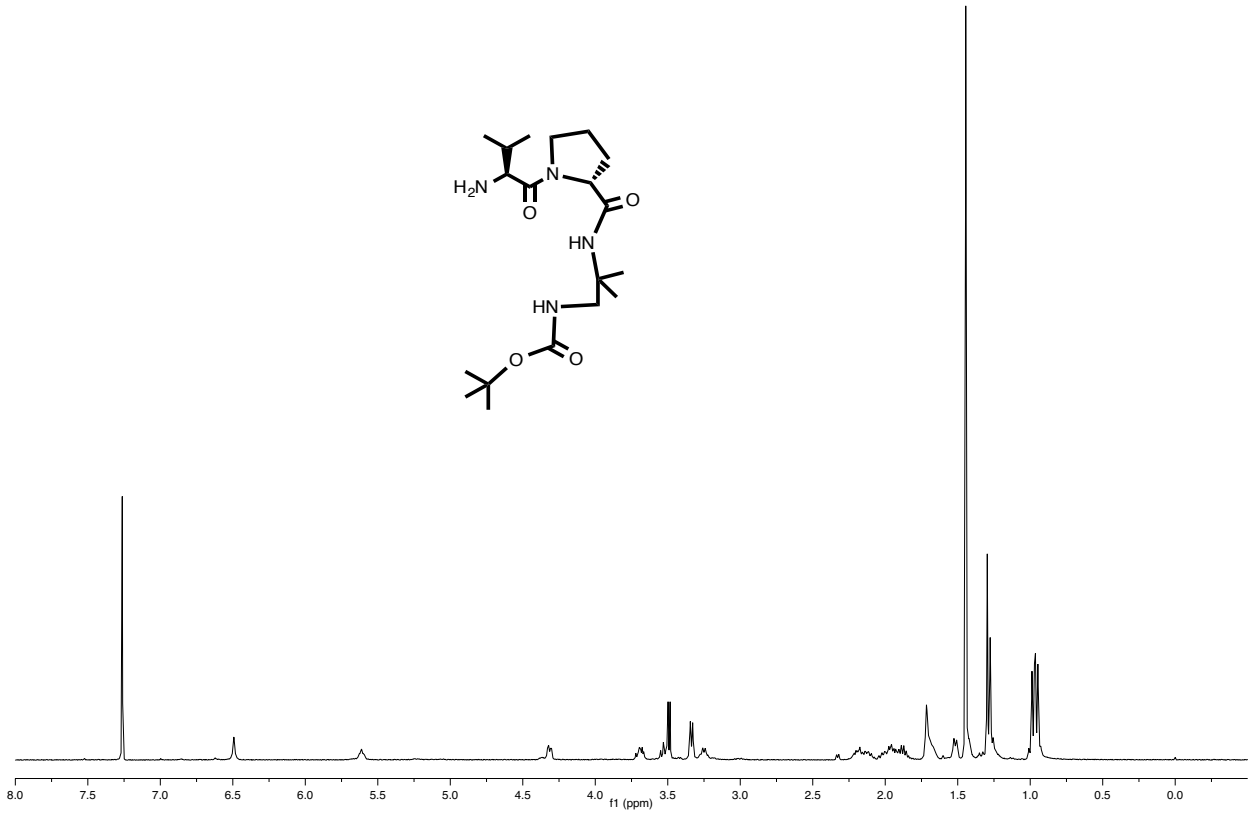

Figure A1.25. ^{1}H NMR of **H-Val-^{D}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

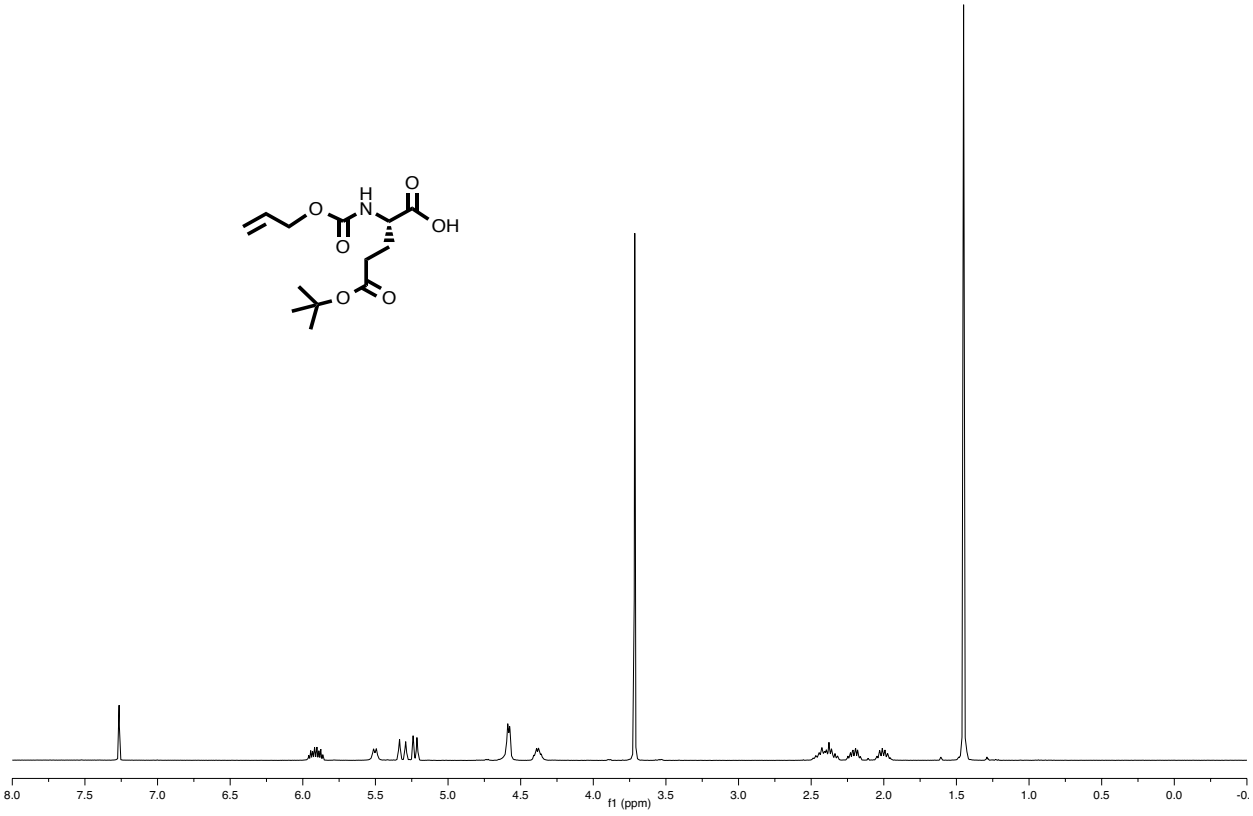

Figure A1.26. ^{1}H NMR of **Alloc-Glu(O*t*Bu)-OH** in $CDCl_3$, 400 MHz.

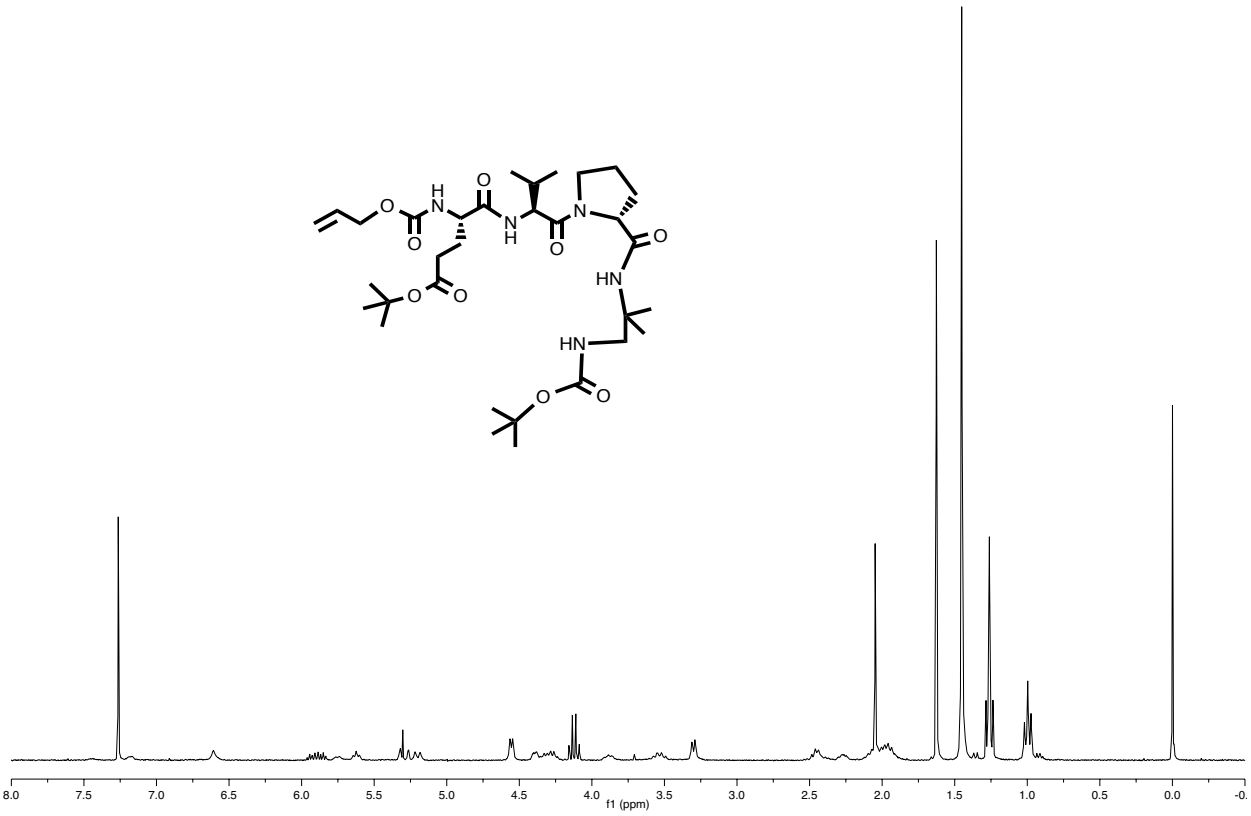

Figure A1.27. ^{1}H NMR of **Alloc-Glu(O*t*Bu)-Val-^{D}Pro-DADME-Boc** in $CDCl_3$, 300 MHz.

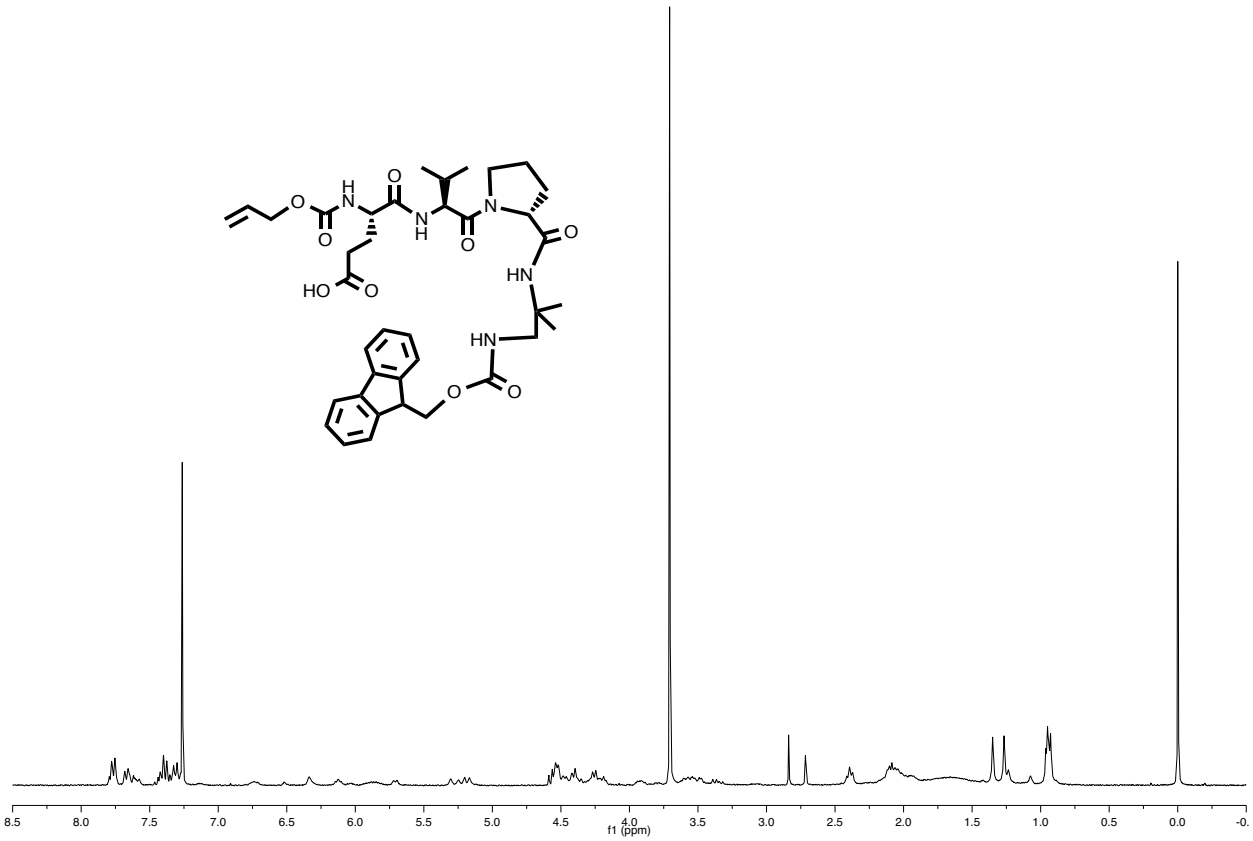

Figure A1.28. ^{1}H NMR of **Alloc-Glu-Val-^{D}Pro-DADME-Fmoc** in $CDCl_3$, 300 MHz.

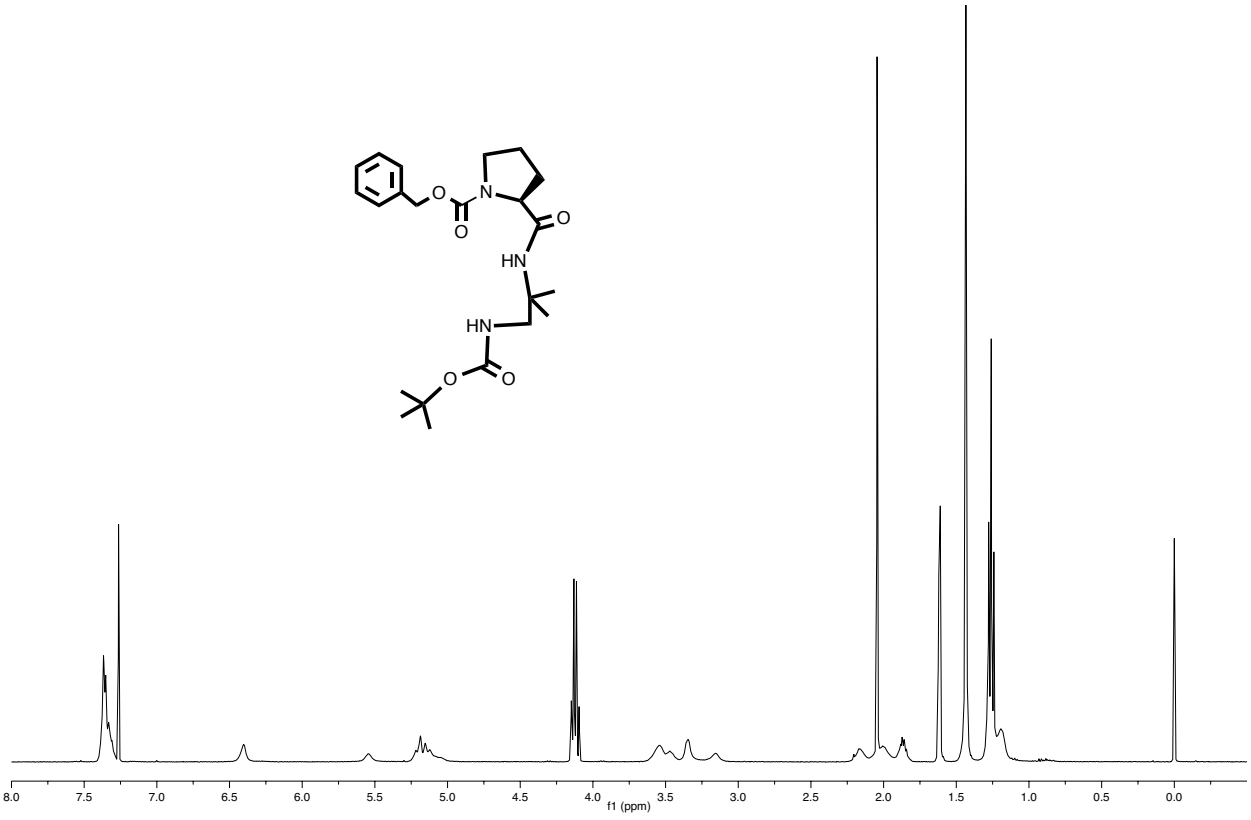

Figure A1.29. ^{1}H NMR of **Cbz-^{L}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

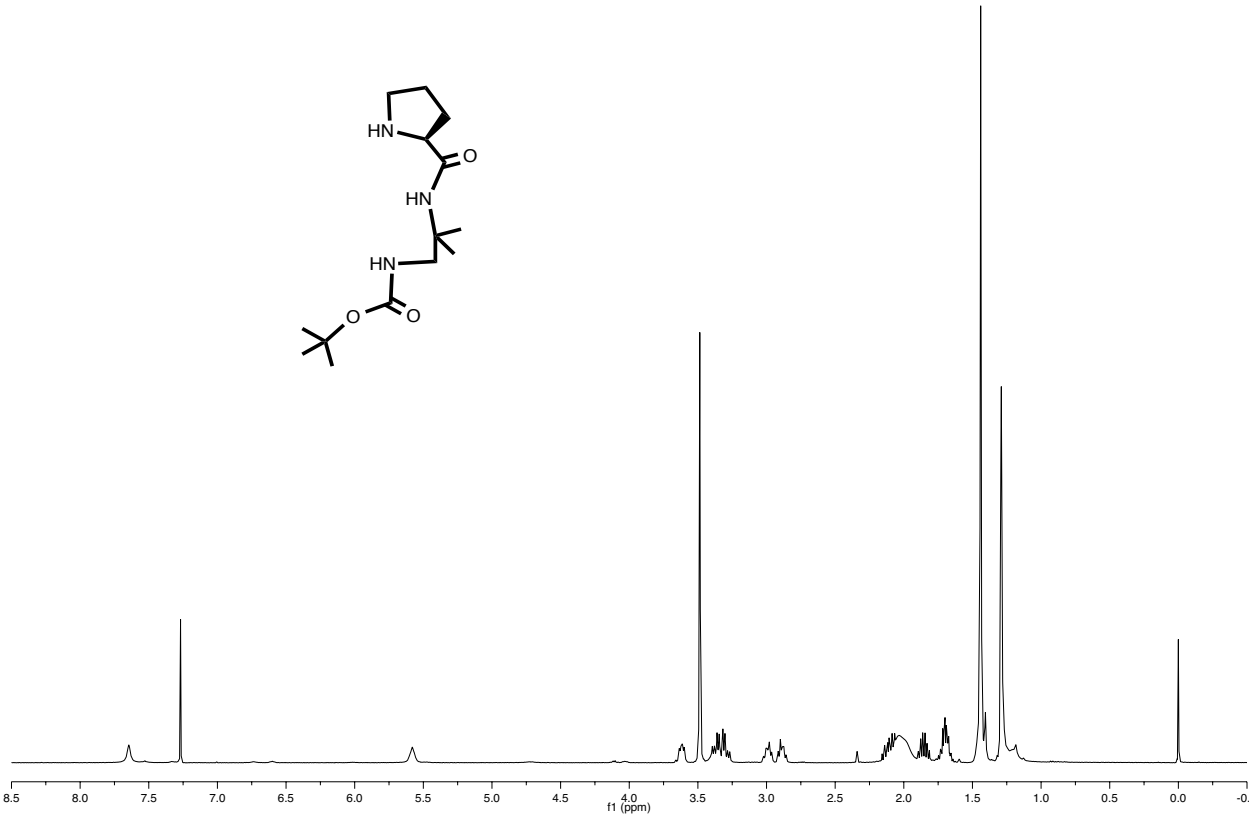

Figure A1.30. ^{1}H NMR of **H-^{L}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

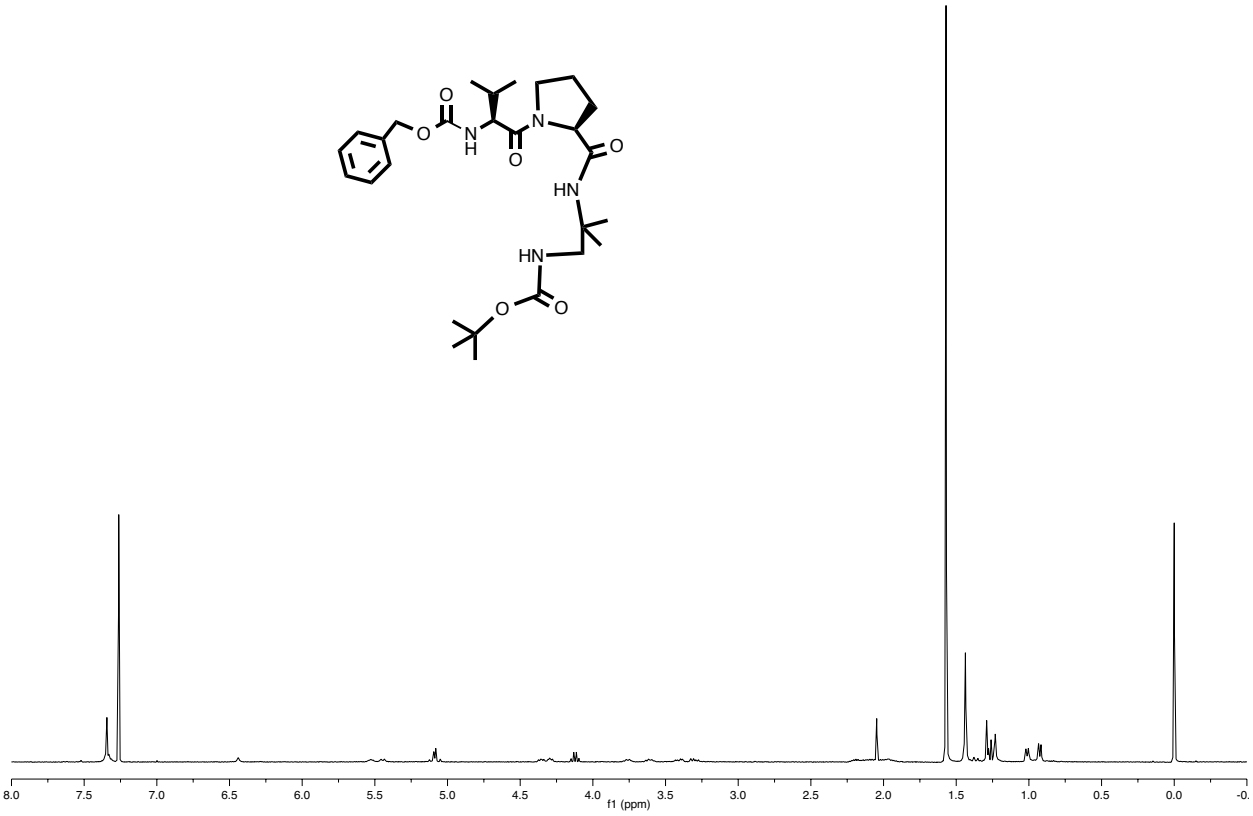

Figure A1.31. ^{1}H NMR of **Cbz-Val-^{L}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

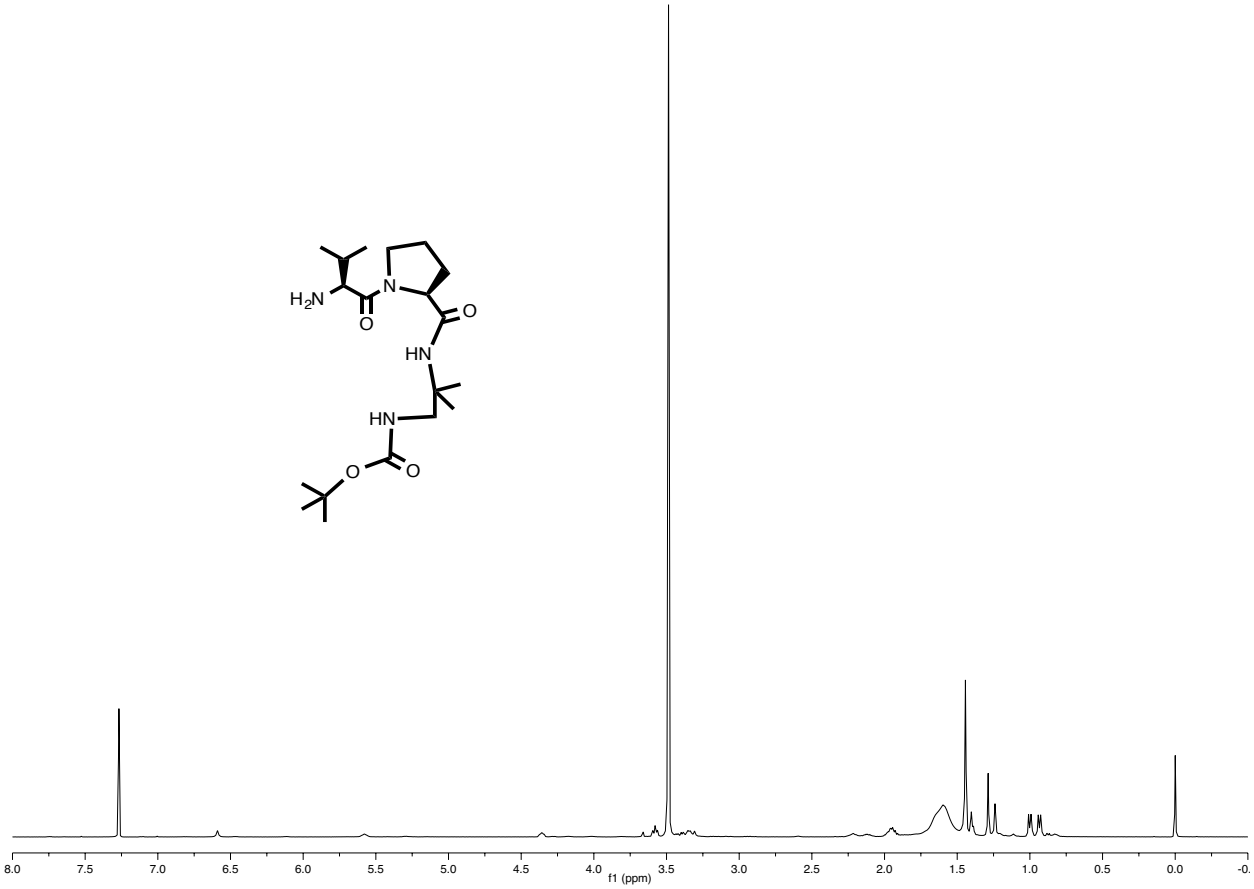

Figure A1.32. ^{1}H NMR of **H-Val-^{L}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

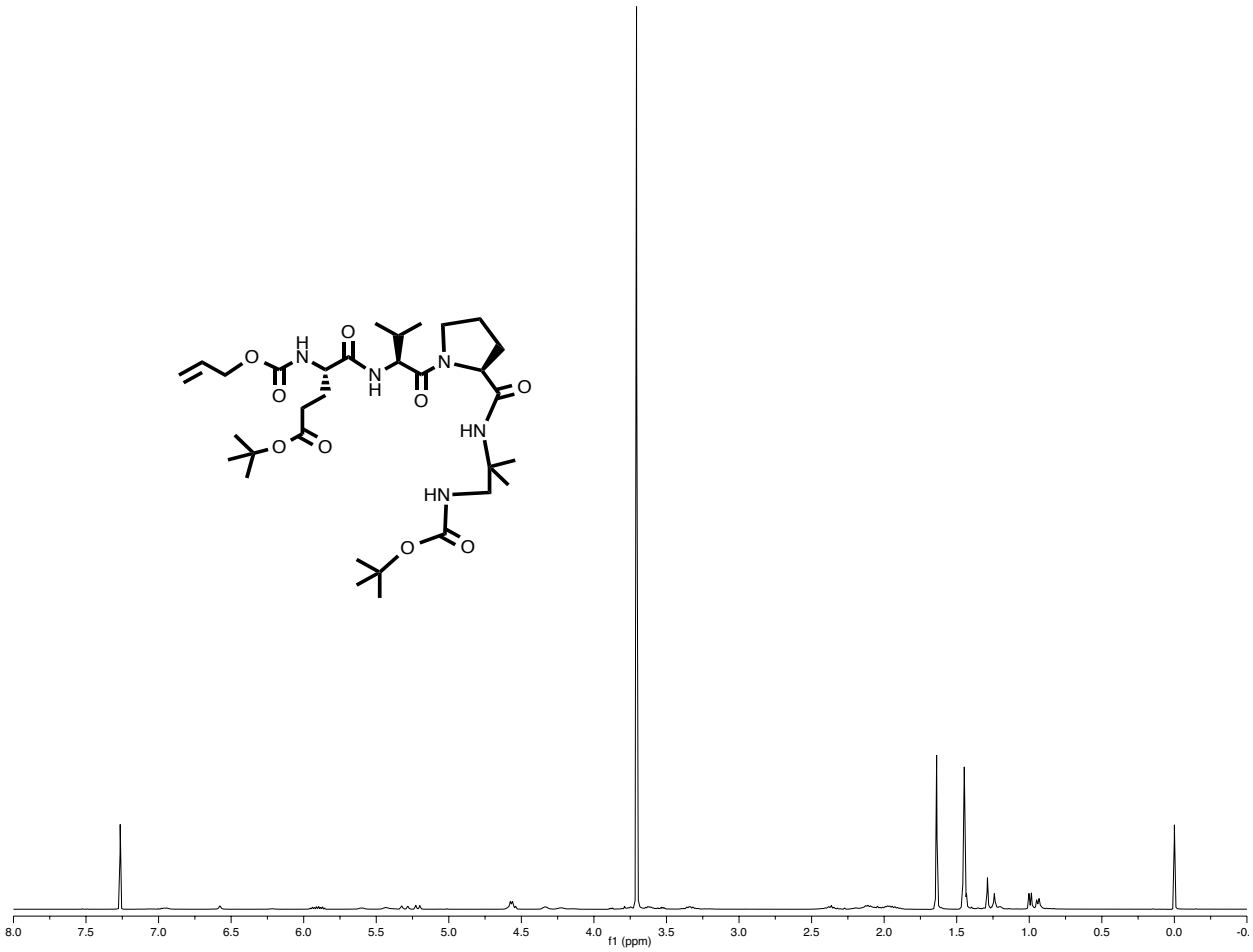

Figure A1.33. ^{1}H NMR of **Alloc-Glu(O*t*Bu)-Val-^{L}Pro-DADME-Boc** in $CDCl_3$, 400 MHz.

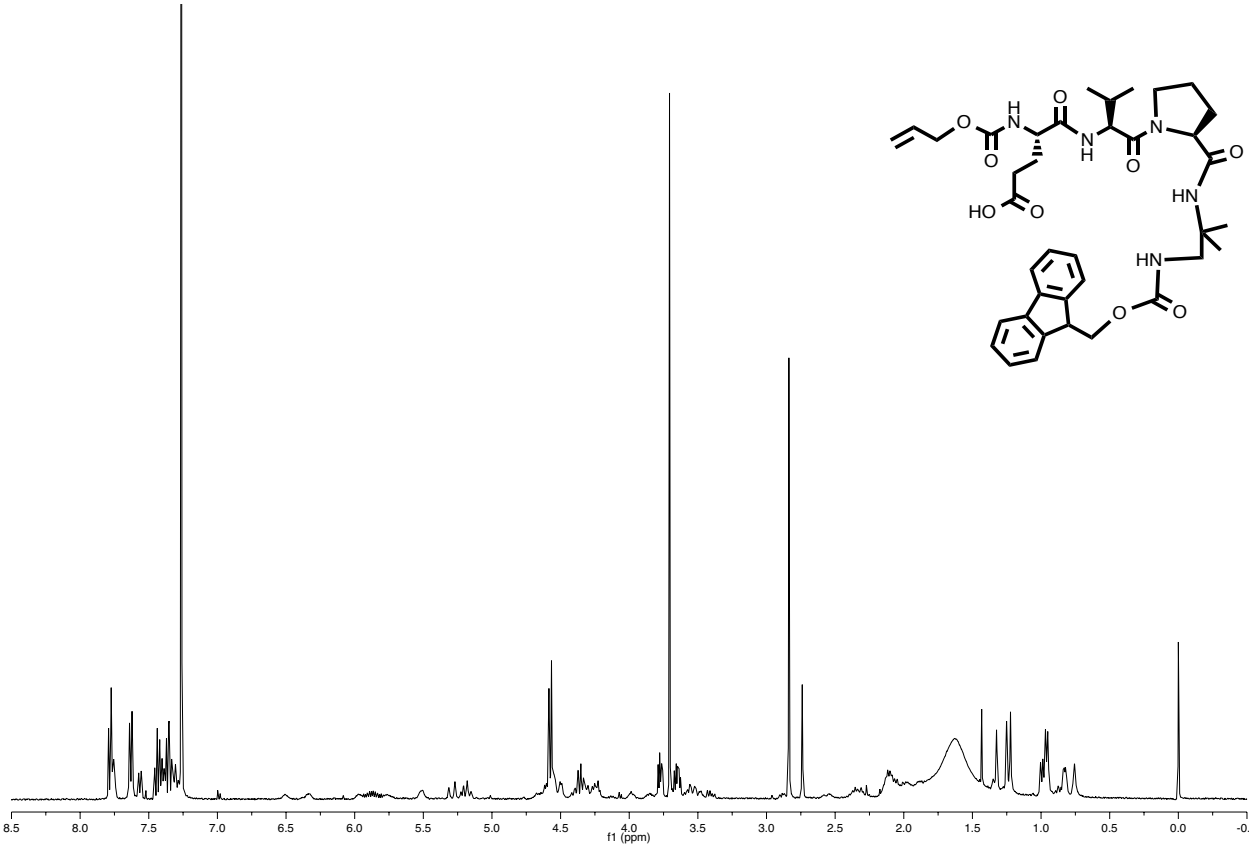

Figure A1.34. ^{1}H NMR of **Alloc-Glu-Val-^{L}Pro-DADME-Fmoc** in $CDCl_3$, 400 MHz.

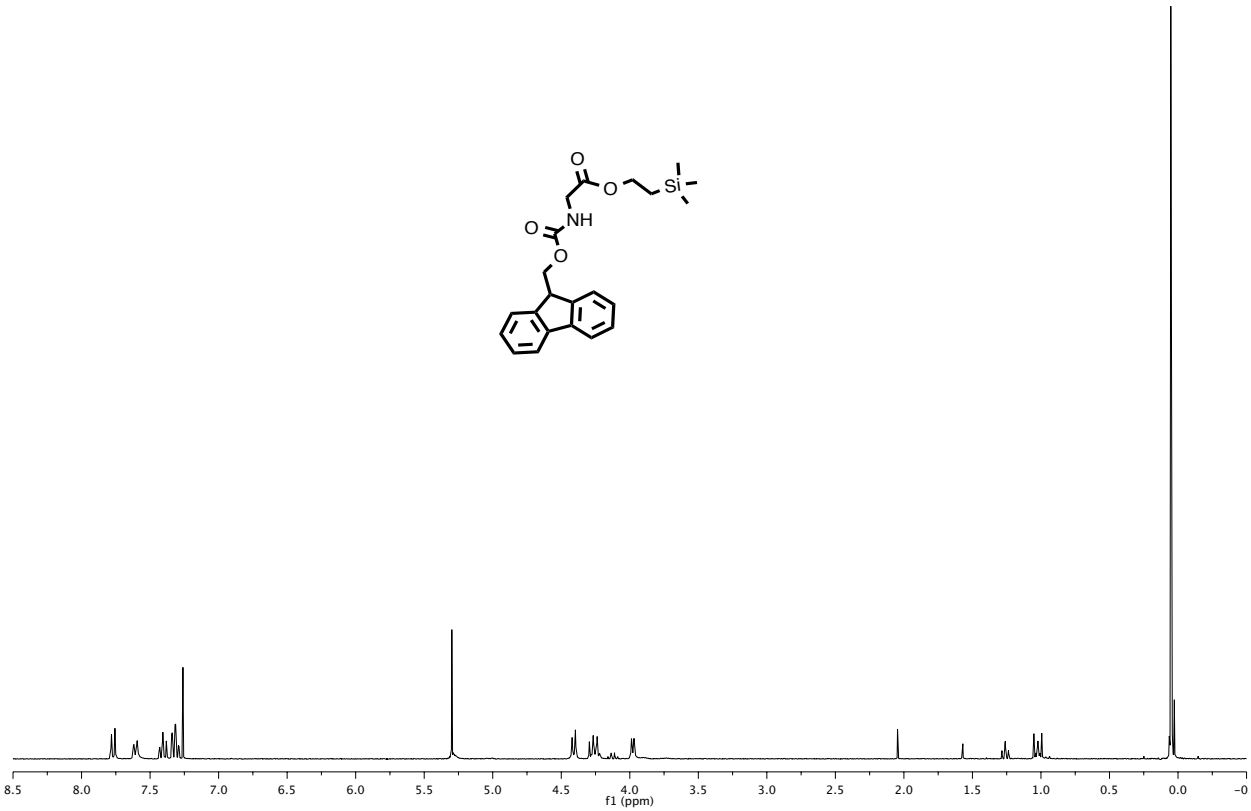

Figure A1.35. ^{1}H NMR of **Fmoc-Gly-OTMSE** in $CDCl_3$, 300 MHz.

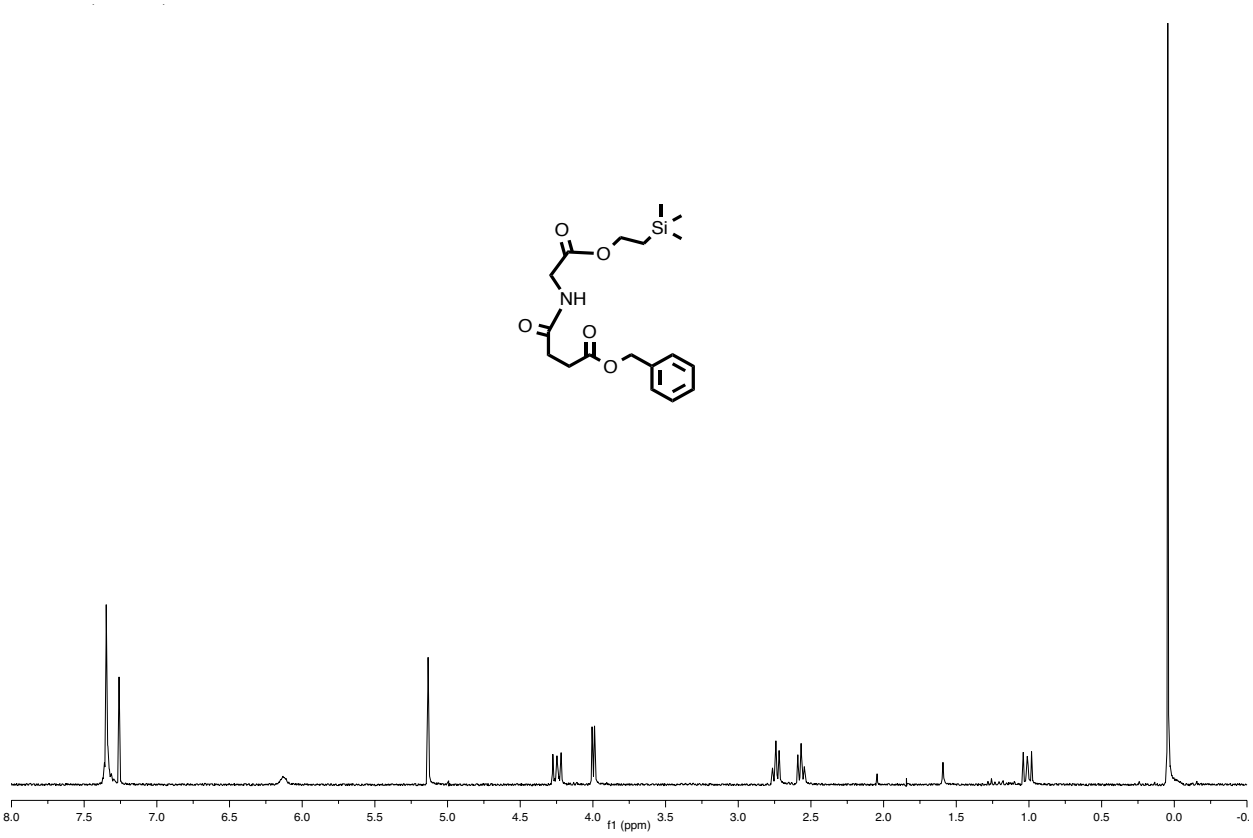

Figure A1.36. ^{1}H NMR of **BnO-Succinyl-Gly-OTMSE** in $CDCl_3$, 300 MHz.

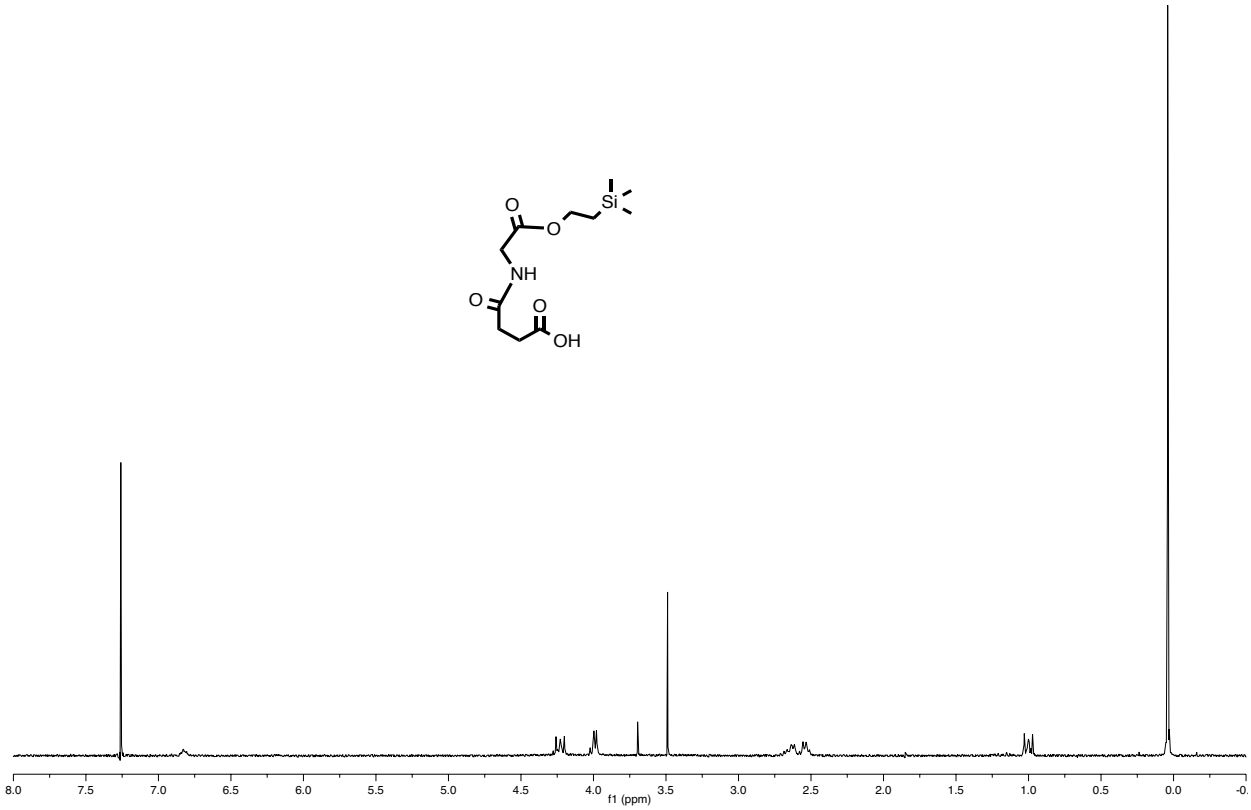

Figure A1.37. ^{1}H NMR of **HO-Succinyl-Gly-OTMSE** in $CDCl_3$, 300 MHz.

A1.4. Supplemental Peptide Characterization Data for Appendix 1.

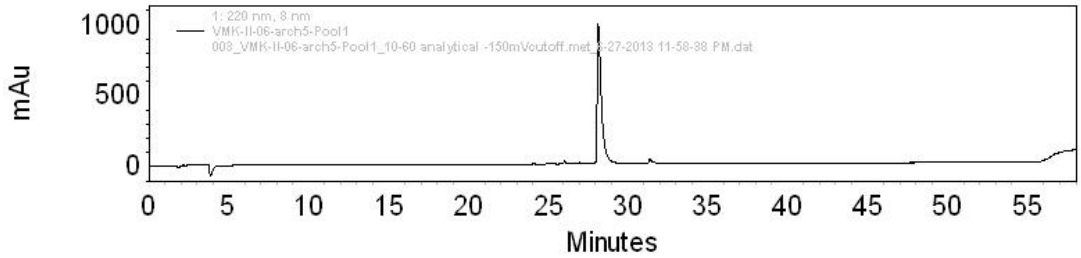

Figure A1.38. Analytical HPLC of peptide **A1.L**.

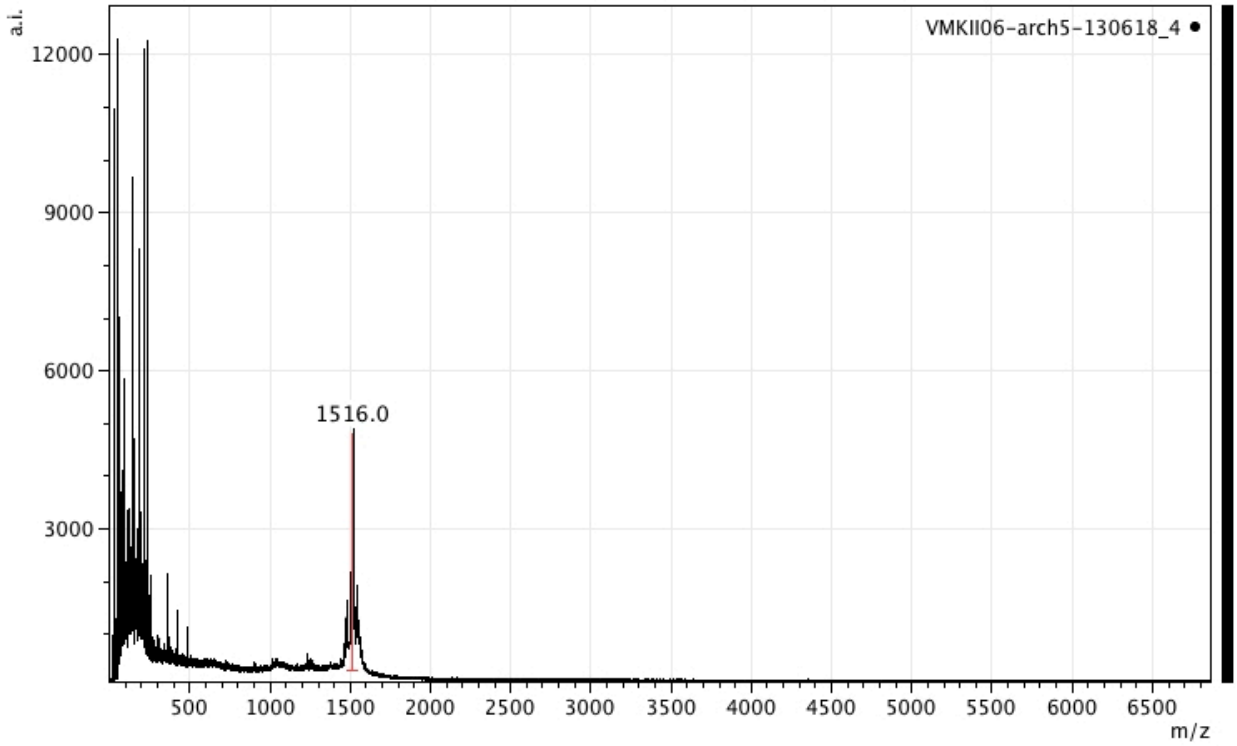

Figure A1.39. MALDI-TOF MS of peptide **A1.L**.

A1.5. References

(1) Almeida, A. M., Doctoral Dissertation Thesis, University of Wisconsin - Madison, 2012.

(2) Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: New York, 2000.

(3) Valeur, E.; Bradley, M. *Chem Soc Rev* **2009**, *38*, 606.

(4) El-Faham, A.; Albericio, F. *Chem Rev* **2011**, *111*, 6557.

(5) Angell, Y. M.; Alsina, J.; Albericio, F.; Barany, G. *J Pept Res* **2002**, *60*, 292.

(6) Beavis, R. C.; Chaudhary, T.; Chait, B. T. *Org Mass Spectrom* **1992**, *27*, 156.

(7) Zenobi, R.; Knochenmuss, R. *Mass Spectrom Rev* **1998**, *17*, 337.

(8) Beavis, R. C.; Chait, B. T.; Fales, H. M. *Rapid Comm Mass Spectrom* **1989**, *3*, 432.

(9) King, D. S.; Fields, C. G.; Fields, G. B. *Int J Pept Protein Res* **1990**, *36*, 255.

(10) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.

(11) Freire, F.; Fisk, J. D.; Peoples, A. J.; Ivancic, M.; Guzei, I. A.; Gellman, S. H. *J Am Chem Soc* **2008**, *130*, 7839.

(12) Bolm, C.; Schiffers, I.; Dinter, C. L.; Gerlach, A. *J Org Chem* **2000**, *65*, 6984.

(13) Li, H. M.; Liu, X. F.; Wu, F. H.; Tang, L. A.; Deng, L. *Proc Natl Acad Sci USA* **2010**, *107*, 20625.

(14) Freire, F.; Gellman, S. H. *J Am Chem Soc* **2009**, *131*, 7970.

Appendix 2. β-Arch Peptide Designs Preceding "Peptide 3.A"

A2.1. β-Arch Peptide Designs Not Based on Consensus Sequences

Aaron M. Almeida and Samuel H. Gellman designed a series of peptides intended to form three-stranded β-arches.[1] The sequences of the intended β-arcs were based on individual (rather than consensus) sequences of β-arcs found in nature (e.g., residues 156-160 of polygalacturonase I of *Aspergillus niger*),[2] and the sequences of the intended β-strands were selected via trial-and-error variation of individual residues at particular positions. Characterization of these peptides did not provide strong evidence for the desired β-arch-like folds or the solubility of these peptides in aqueous solution.[1] We nevertheless began our work by synthesizing peptide **A2.X** (**Figure A2.1X**), a peptide intended to form a four-stranded β-arch, with a sequence based on what Almeida and Gellman described as their most promising results. As with previous peptide designs, peptide **A2.X** had limited solubility in aqueous solution.

A2.2. β-Arch Peptide Designs Based on a 2006 Consensus Sequence

We subsequently turned to β-arch peptide designs based on a published consensus sequence by Kajava and coworkers.[3] When isolated, a short peptide sequence extracted from a larger protein often loses the conformation that it had in the structural context of the folded, native protein. It follows that any β-arch peptide design based on a single β-arc found in nature ran the risk of not folding into a β-arc once it was removed from the context of the whole protein. We thought that introducing consensus sequences to the peptide designs might remove an element of trial-and-error from the design process, as well as provide greater significance to any successfully generated peptide model. We hypothesized that incorporating a β-arch consensus sequence into the peptide design would increase the propensity of the peptide to adopt a β-arch fold – i.e., that designing a

peptide sequence that had at each position a residue that was common to multiple β-arches, would decrease the possibility that the sequence formed a β-arch only in the presence of features (i.e., the chemical environment produced by nearby residues) specific to a single protein. Based on the sequence logo by Kajava and coworkers (reproduced in **Figure A2.3I**), we designed and synthesized peptide **A2.Y** (**Figure A2.1Y**). However, peptide **A2.Y**, which like peptide **A2.X** had a net charge of +6 in aqueous solution at pH 3.8, also had limited solubility in aqueous solution (**Figure A2.2Y**).

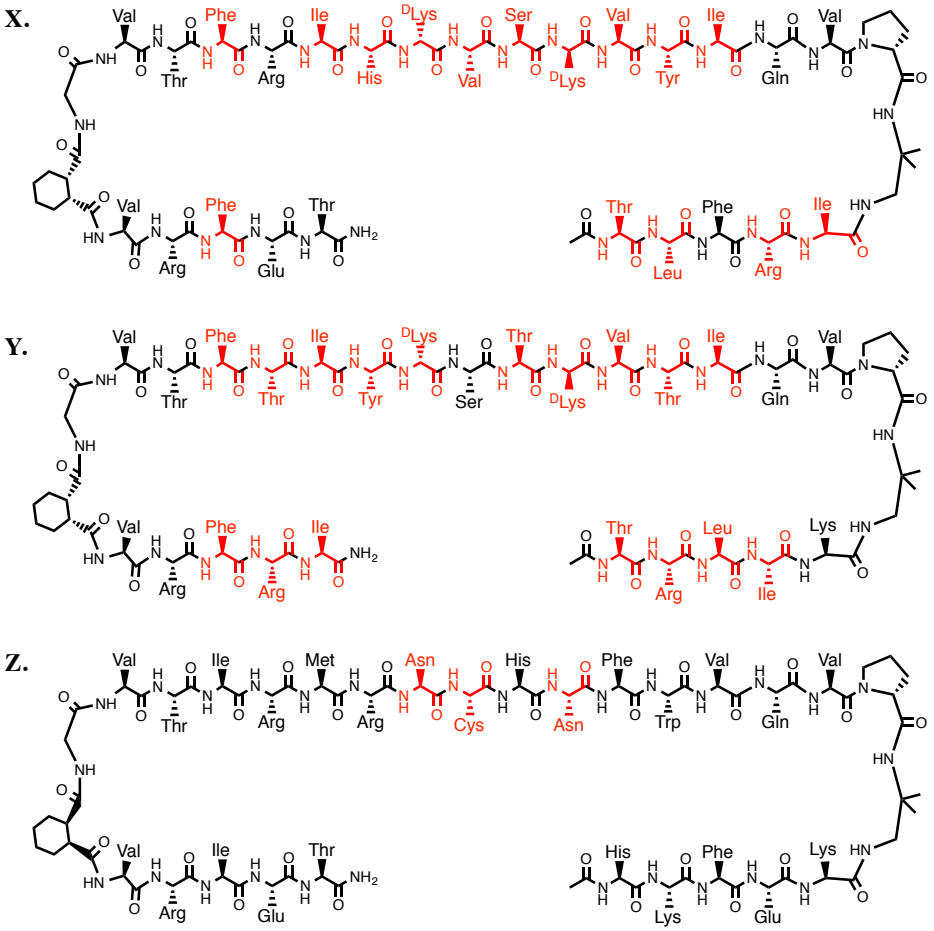

Figure A2.1. Peptides **A2.X**, **A2.Y**, and **A2.Z**, with differences from peptide **3.A** (**Figure 3.9**) highlighted in red. The intended β-arc of peptide **A2.X** was based on residues 156-160 of polygalacturonase I of *Aspergillus niger*, and the sequences of the intended β-strands were not based on consensus sequences or sequences of natural proteins. The sequence of peptide **Y** was designed in a similar manner to the design of peptide **3.A**, but instead employed the *blbbl* β-arch consensus sequence reported by Kajava and coworkers. Peptide **A2.Z** has a design similar to that of peptide **3.A**.

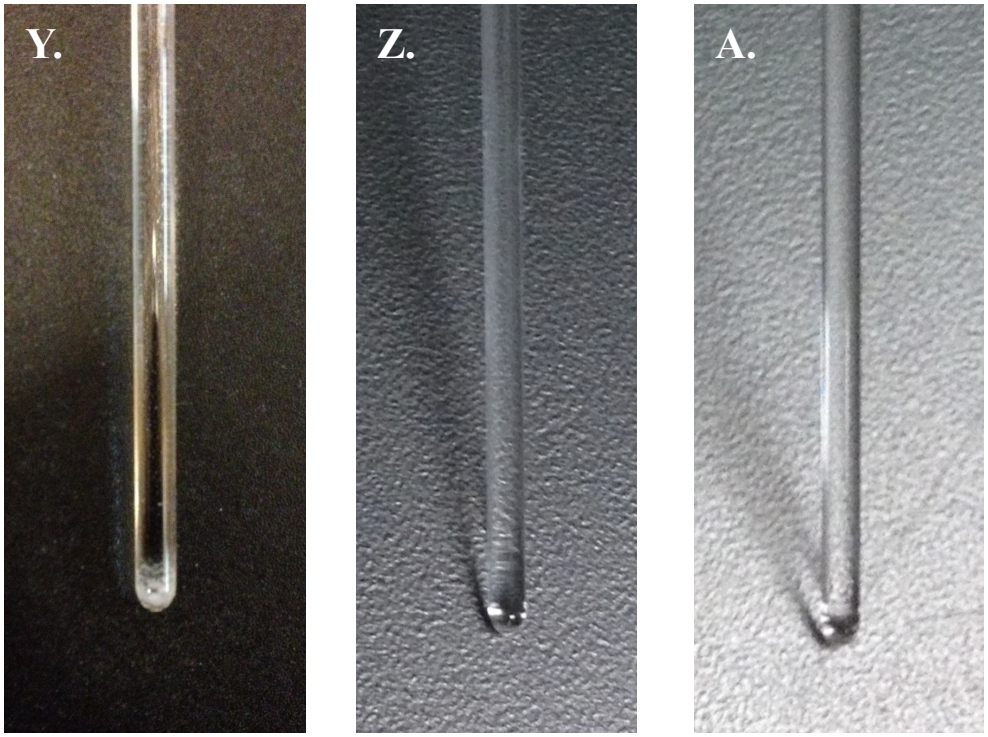

Figure A2.2. NMR tubes containing 2 mM peptide in 3 mM NaOAc-d_3, pH 3.8. (Y) Peptide **A2.Y** began to visibly precipitate from the tube within minutes of sample preparation (see white precipitate at bottom of tube). In contrast, there was no visible heterogeneity, formation of viscous gels, or precipitation from NMR tubes, for peptide **A2.Z** (Z) or **3.A** (A), for >48 hours at 4-20 °C.

A2.3. Motivation for New Bioinformatics Studies, and Consequent Peptide Designs

To study the folded conformation of the model β-arch peptide in aqueous solution, we needed to increase the water solubility of the peptide, by increasing the number of charged residues. With the limited amount of information provided in the Kajava sequence logo (**Figure A2.3I**), however, it was impossible to determine where charged residues were generally found within the β-arch, and where they might accordingly be incorporated into a β-arch model peptide. Moreover, we noted a few areas in which we might improve upon the inspiring work of Kajava and coworkers: (1) the PDB had grown since their 2006 paper; (2) we were not able to reproduce the published *blbbl* sequence logo from the data

set in their supplementary information or from the data set that Kajava sent to us in a personal communication on January 22, 2013; and (3) the 2006 sequence logo had not been normalized. We therefore conducted the bioinformatics studies described in **Chapter 3**. The differences between our sequence logo for the *blbbl* β-arch, and that of Kajava and coworkers, are highlighted in **Figure A2.3**. A key difference between the two consensus sequences is that ours shows L-asparagine (Asn, N) at the *l* positions of the β-arc. This finding agrees with established observations on the unusual conformational characteristics of asparaginyl residues, possibly resulting from the stabilization of *l* conformations via carbonyl-carbonyl interactions.[4,5] From our consensus sequence for a *blbbl* β-arch (**Figure 3.6**), we designed peptide **A2.Z** (**Figure A2.1Z**), using the principles described in **Chapter 3** for the design of peptide **3.A**. The sequence of peptide **A2.Z** differed from that of peptide **3.A** only at three residues, with Arg13-Asn14-Cys15-His16-Asn17 (rather than Arg13-^{D}Ala14-Ser15-His16-^{D}Ala17) as the intended β-arc. The residues of the intended β-arc of peptide **A2.Z** were chosen to exactly correspond to the residues in our β-arc consensus sequence (**Figure 3.6**). Peptide **A2.Z** was the first peptide design to not visibly precipitate from aqueous solution, under traditional conditions for obtaining NMR data (**Figure A2.2Z**). The net charge of peptide **A2.Z** at pH 3.8 is +7, with the simplifying assumption of 100% protonation of Glu side-chains. In other words, peptide **A2.Z** had a similar net charge to previous peptide designs (peptides **A2.X**, **A2.Y**), but did not precipitate under the conditions that had caused previous peptide designs to precipitate. We expect that this dramatic increase in solubility reflected a decrease in unstructured peptide aggregation, which in turn reflected peptide folding into the intended β-arch. Our bioinformatics work had yielded a consensus sequence that had enabled

strategic placement of charged and other residues, to promote water-solubility and folding of the β-arch.

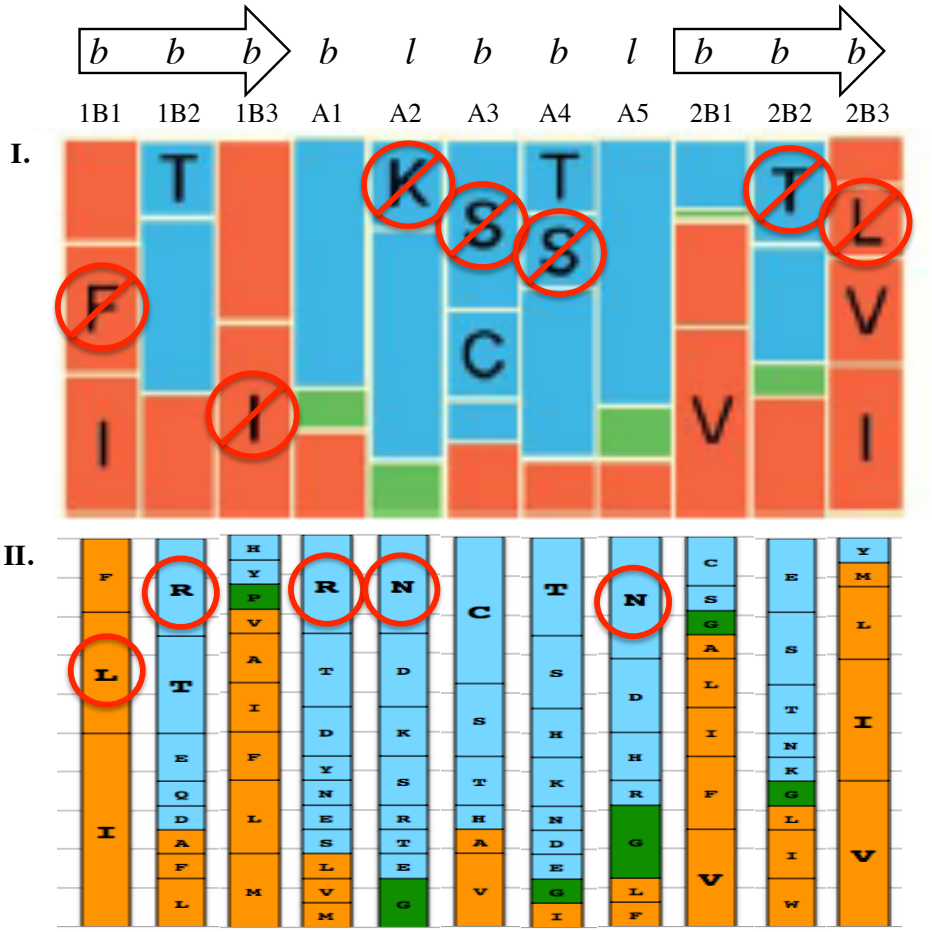

Figure A2.3. Comparison of *blbbl* β-arch sequence logos by (I) Kajava and coworkers in 2006,[3] and (II) Kung, *et al.* in 2013, with differences highlighted in red circles. Unlike in **Figure 3.6**, the logo in (II) is not normalized for the natural abundance of amino acids, because Kajava and coworkers did not normalize their data. **Figure A2.3I** reprinted with permission from Elsevier.

From peptide **A2.Z**, we moved onto the design of peptide **3.A** and its variants (e.g., peptide **3.B**, which like **A2.Z** has Asn at positions 14 and 17), choosing to place Ser instead of Cys at position 15 of the β-arc, to avoid oxidation reactions that can complicate syntheses and characterizations. We did not observe Cys oxidation or peptide dimerization at the conditions used to obtain NMR spectra (0.02-2 mM peptide, 2-100 mM NaOAc-d_3,

pH 3.8, 4-5 °C, 0-48 hours), by analytical HPLC and MALDI-TOF MS. However, we did observe Cys oxidation when samples were left at room temperature for >12 hours, which occurred when various pieces of laboratory equipment (e.g., refrigerator, lyophilizer) broke. Ser was chosen to replace the consensus Cys, because Ser had the closest structure to Cys.

Another of our designs that was not discussed in **Chapter 3** was peptide **A2.Q**, the variant of peptide **3.A** with ^{D}Ala14^{D}Glu and ^{D}Ala17^{D}Glu substitutions in the β-arc. This variant was designed because L-glutamate (Glu), like Ala, has a high propensity for adopting backbone dihedral angles in the right-handed α-helical basin of the Ramachandran plot.[6,7] We therefore expected ^{D}Glu, like ^{D}Ala, to have a high propensity for adopting left-handed α-helical (*l*) Ramachandran angles. However, some precipitation of peptide **A2.Q** from aqueous solution was observed (though not as much precipitation as was observed for peptides **A2.X** or **A2.Y**). The limited solubility of peptide **A2.Q** in aqueous solution is likely attributable to the two additional glutamate residues that it contains. Assuming a pKa of 4.2 for the ε-proton of glutamic acid, we would expect 28% of Glu side-chains to be negatively-charged at pH 3.8.

A2.4. Supplementary Peptide Characterization Data for Appendix 2.

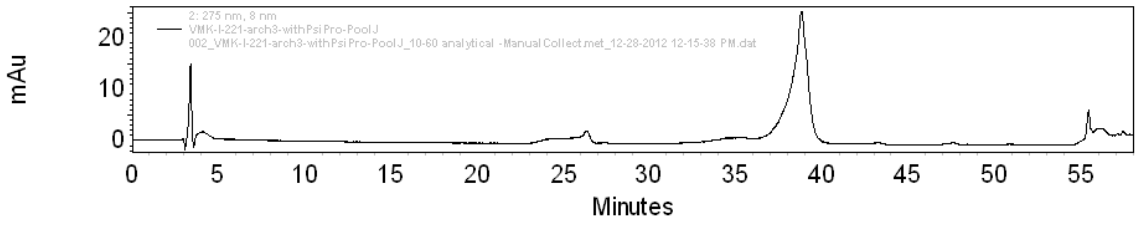

Figure A2.4Y. Analytical HPLC of peptide **A2.Y**.

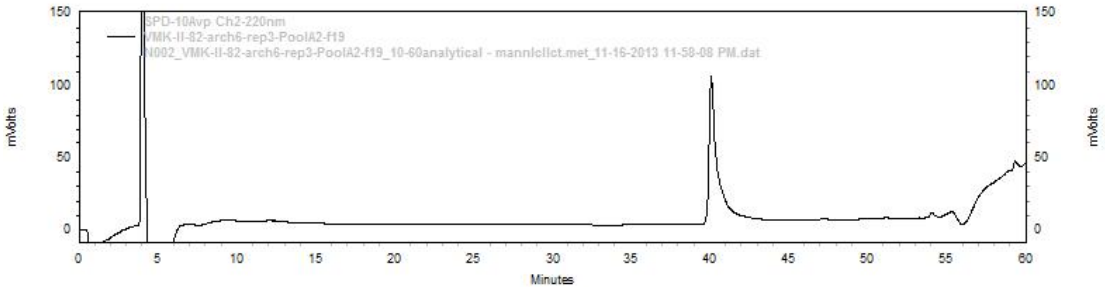

Figure A2.4Z. Analytical HPLC of peptide **A2.Z**.

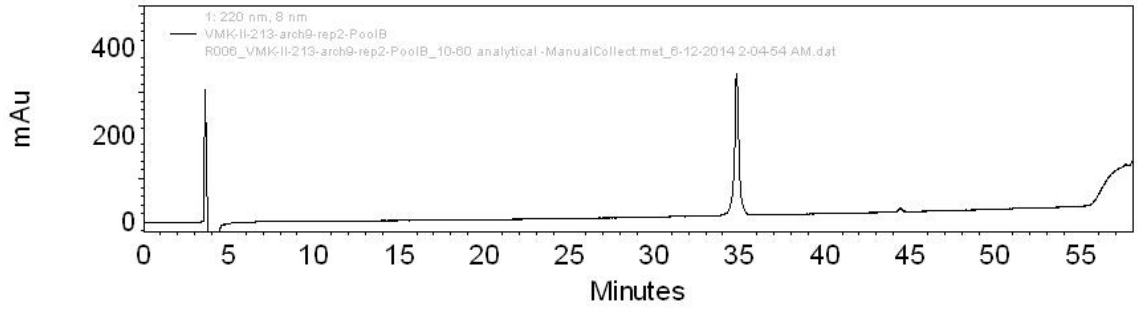

Figure A2.4Q. Analytical HPLC of peptide **A2.Q**.

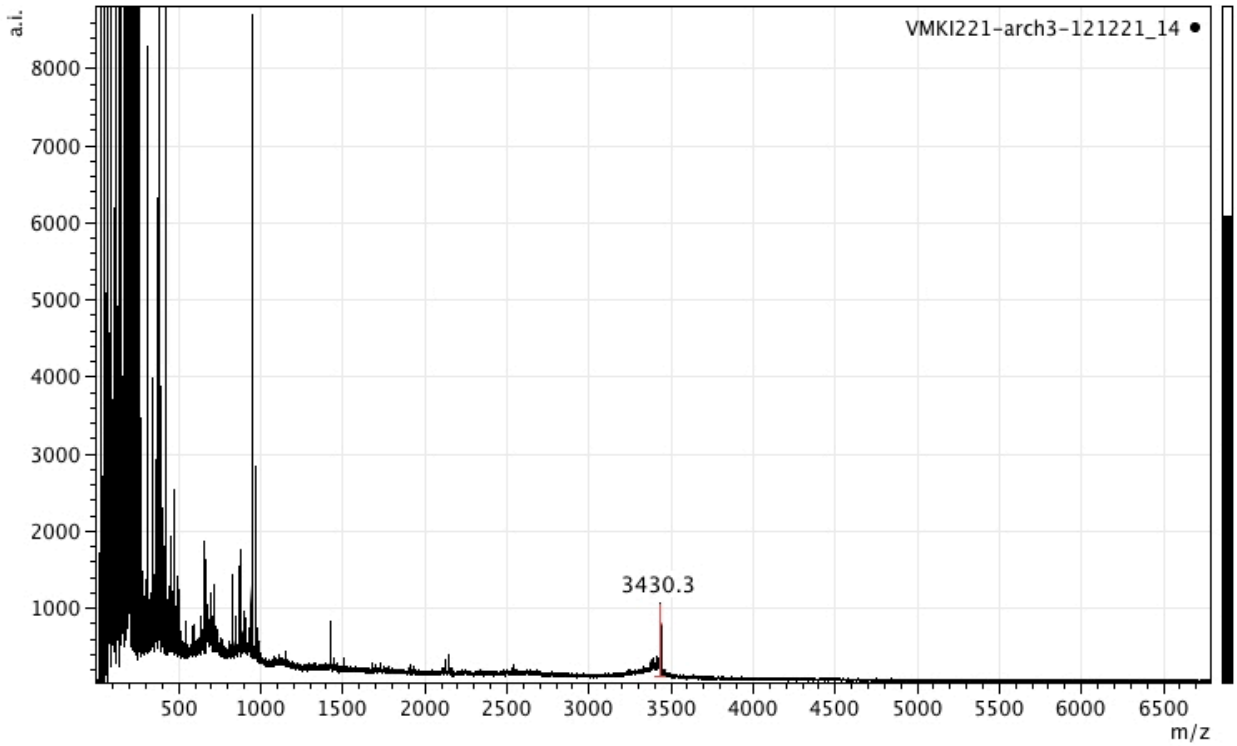

Figure A2.5Y. MALDI-TOF MS of peptide **A2.Y**. Calculated m/z of $[M+H^+]^+$ = 3431.09; observed = 3430.3.

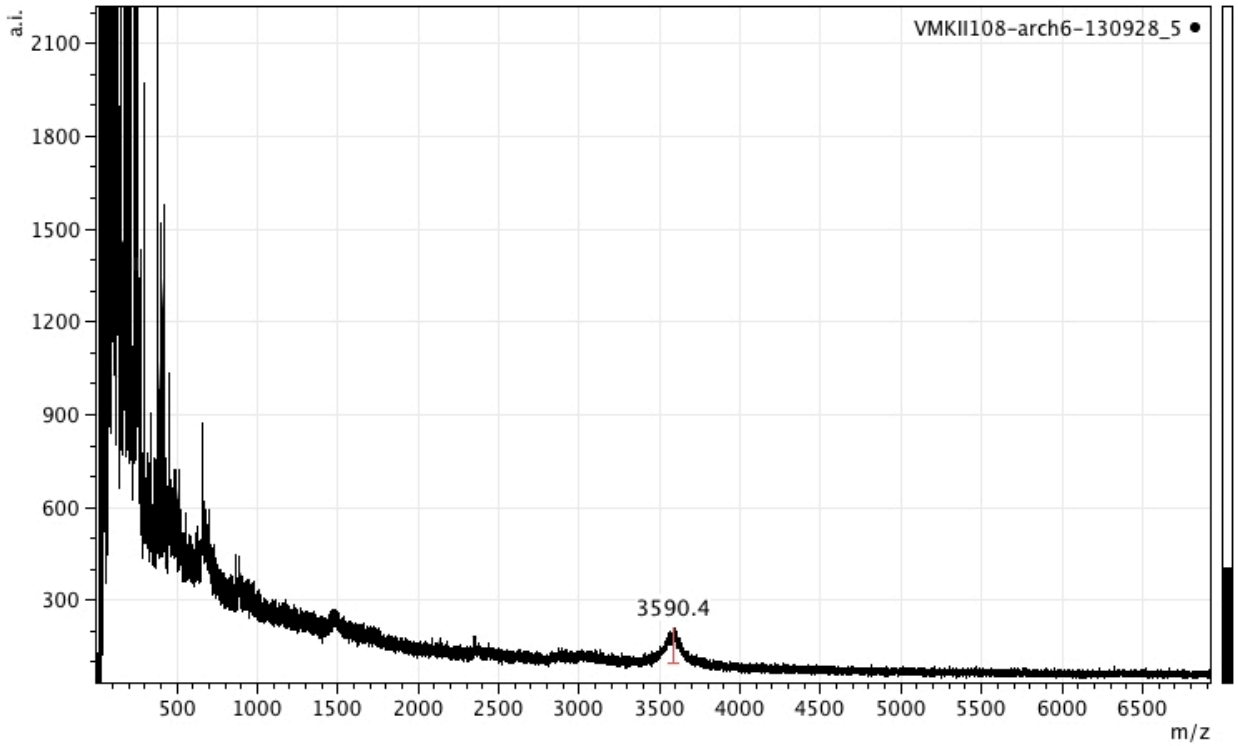

Figure A2.5Z. MALDI-TOF MS of peptide **A2.Z**. Calculated m/z of $[M+H^+]^+$ = 3590.92; observed = 3590.4.

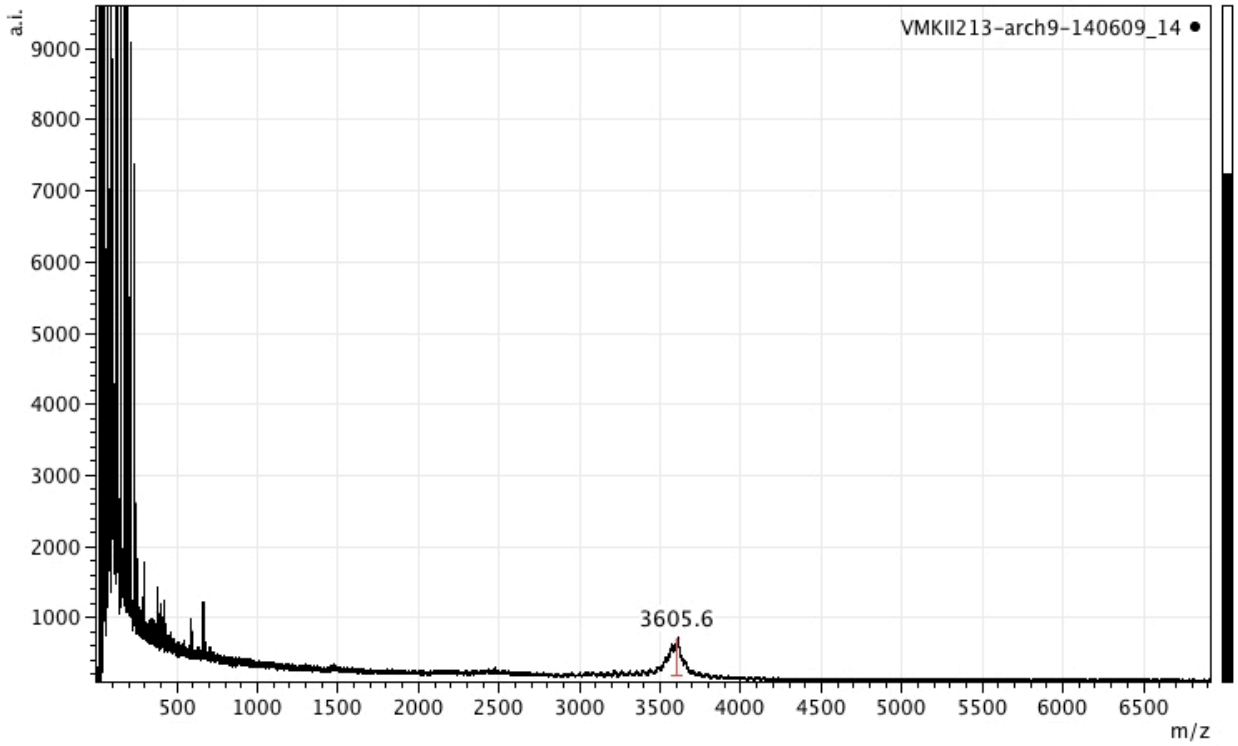

Figure A2.5Q. MALDI-TOF MS of peptide **A2.Q**. Calculated m/z of $[M+H^+]^+$ = 3604.94; observed = 3605.6.

A2.5. References

(1) Almeida, A. M., Doctoral Disseration Thesis, University of Wisconsin - Madison, 2012.

(2) van Pouderoyen, G.; Snijder, H. J.; Benen, J. A. E.; Dijkstra, B. W. *FEBS Lett* **2003**, *554*, 462.

(3) Hennetin, J.; Jullian, B.; Steven, A. C.; Kajava, A. V. *J Mol Biol* **2006**, *358*, 1094.

(4) Srinivasan, N.; Anuradha, V. S.; Ramakrishnan, C.; Sowdhamini, R.; Balaram, P. *Int J Pept Prot Res* **1994**, *44*, 112.

(5) Deane, C. M.; Allen, F. H.; Taylor, R.; Blundell, T. L. *Protein Eng* **1999**, *12*, 1025.

(6) Williams, R. W.; Chang, A.; Juretic, D.; Loughran, S. *Biochim Biophys Acta* **1987**, *916*, 200.

(7) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr D* **2002**, *58*, 768.

Appendix 3. Supplemental Information for Computer Scripts

A3.1. Perl Script for Searching for Sequences Containing Particular Sets of Backbone Dihedral Angles (e.g., *blbbl*)

```
#!/usr/bin/perl -w
#
#
#dssp reader for finding blbbl sequences
#Vanessa M. Kung, Gail J. Bartlett, version 20130209
#
#
#
================================================================================
=========
use strict;
use Data::Dumper;

my $flag = 0;


my $aData = [];

open (FILE, $ARGV[0]);
while (defined (my $line = <FILE>)) {

    chomp $line;

    if ($flag == 0) {

     if ($line =~ /^\s+#/) {

          $flag = 1;
          next;
     }
    }
    else {

     my $residue_num = substr($line,5,5);
     my $residue_id = substr($line,13,1);
     my $phi = substr($line,103,6);
     my $psi = substr($line,109,6);

     $residue_num =~ s/^\s+//;
     $phi =~ s/\s+//;
     $psi =~ s/\s+//;

     my $ss;

     if (($phi > -180) && ($phi < -90) && ($psi > 90) && ($psi < 180))
{
```

```
            $ss = "b";
      }

      elsif (($phi > 30) && ($phi < 100) && ($psi > -20) && ($psi <
90)) {

            $ss = "l";
      }
      else {

            $ss = "-";
      }


      push @{$aData}, [$residue_num, $residue_id, $phi, $psi, $ss];


    }

}

for (my $i = 0; $i < @$aData - 5; $i++) {

    my $ss_string;

    my $letter_1 = $aData->[$i]->[4];
    my $letter_2 = $aData->[$i+1]->[4];
    my $letter_3 = $aData->[$i+2]->[4];
    my $letter_4 = $aData->[$i+3]->[4];
    my $letter_5 = $aData->[$i+4]->[4];

    $ss_string = join "", $letter_1, $letter_2, $letter_3, $letter_4,
$letter_5;


    if ($ss_string eq "$ARGV[1]") {

        print $ARGV[0];
        print "\t";
      print $aData->[$i]->[0];
      print "\t";
      for (my $j = $i; $j <= $i+4;$j++) {

            print $aData->[$j]->[1];
      }
      print "\t";
      print $aData->[$i+4]->[0];
      print "\n";
    }
}
```

```
exit;
```

A3.2. Coordinates, Topology, and Parameters Files for Building Non-Natural Residue DADME

DADME Coordinates File:

```
REMARK  Vanessa M. Kung, 20140204
REMARK
REMARK
REMARK  This file was generated by PRODRG version AA100323.0717
REMARK  PRODRG written/copyrighted by Daan van Aalten
REMARK  and Alexander Schuettelkopf
REMARK
REMARK  Questions/comments to dava@davapc1.bioch.dundee.ac.uk
REMARK
REMARK  When using this software in a publication, cite:
REMARK  A. W. Schuettelkopf and D. M. F. van Aalten (2004).
REMARK  PRODRG - a tool for high-throughput crystallography
REMARK  of protein-ligand complexes.
REMARK  Acta Crystallogr. D60, 1355--1363.
REMARK
REMARK
HETATM    1  CAA DRG    1     21.500 -20.620  29.010  1.00 20.00         C
HETATM    2  HAA DRG    1     20.650 -20.954  28.415  1.00 20.00         H
HETATM    3  HAB DRG    1     21.959 -21.478  29.501  1.00 20.00         H
HETATM    4  HAC DRG    1     22.222 -20.070  28.406  1.00 20.00         H
HETATM    5  CAF DRG    1     21.020 -19.620  30.070  1.00 20.00         C
HETATM    6  CAB DRG    1     22.260 -19.180  30.860  1.00 20.00         C
HETATM    7  HAE DRG    1     22.677 -20.037  31.390  1.00 20.00         H
HETATM    8  HAF DRG    1     21.978 -18.411  31.579  1.00 20.00         H
HETATM    9  HAD DRG    1     22.953 -18.795  30.112  1.00 20.00         H
HETATM   10  NAD DRG    1     20.090 -20.350  30.950  1.00 20.00         N
HETATM   11  HAM DRG    1     19.730 -19.740  31.650  1.00 20.00         H
HETATM   12  HAN DRG    1     20.580 -21.110  31.390  1.00 20.00         H
HETATM   13  HAL DRG    1     19.330 -20.720  30.410  1.00 20.00         H
HETATM   14  CAE DRG    1     20.380 -18.430  29.290  1.00 20.00         C
HETATM   15  HAG DRG    1     21.046 -18.260  28.444  1.00 20.00         H
HETATM   16  HAH DRG    1     19.400 -18.777  28.964  1.00 20.00         H
HETATM   17  NAC DRG    1     20.160 -17.090  29.900  1.00 20.00         N
HETATM   18  HAJ DRG    1     19.550 -17.180  30.680  1.00 20.00         H
HETATM   19  HAK DRG    1     19.740 -16.490  29.220  1.00 20.00         H
HETATM   20  HAI DRG    1     21.040 -16.710  30.200  1.00 20.00         H
CONECT    1    2    3    4    5
CONECT    2    1
```

```
CONECT   3   1
CONECT   4   1
CONECT   5   1   6  10  14
CONECT   6   5   7   8   9
CONECT   7   6
CONECT   8   6
CONECT   9   6
CONECT  10   5  11  12  13
CONECT  11  10
CONECT  12  10
CONECT  13  10
CONECT  14   5  15  16  17
CONECT  15  14
CONECT  16  14
CONECT  17  14  18  19  20
CONECT  18  17
CONECT  19  17
CONECT  20  17
END
```

DADME Topology File:

```
!Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version 20150224
!
!
!     This file was generated by PRODRG version AA100323.0717
!     PRODRG written/copyrighted by Daan van Aalten
!     and Alexander Schuettelkopf
!
!     Questions/comments to dava@davapc1.bioch.dundee.ac.uk
!
!     When using this software in a publication, cite:
!     A. W. Schuettelkopf and D. M. F. van Aalten (2004).
!     PRODRG - a tool for high-throughput crystallography
!     of protein-ligand complexes.
!     Acta Crystallogr. D60, 1355--1363.
!
!
!  *** NOTE *** IF YOU USE MORE THAN ONE PRODRG-GENERATED TOPOLOGY IN CNS,
!            PLEASE CAREFULLY READ THE FAQ AS THIS MAY CAUSE PROBLEMS
!
!
set echo=false end
AUTOGENERATE  ANGLES=FALSE  END

mass H    1.008
```

```
mass C   12.011
mass O   15.999
mass N   14.007
MASS CGAA  12.0110
MASS HGAA   1.0080
MASS HGAB   1.0080
MASS HGAC   1.0080
MASS CGAF  12.0110
MASS CGAB  12.0110
MASS HGAE   1.0080
MASS HGAF   1.0080
MASS HGAD   1.0080
MASS NGAD  14.0067
MASS HGAM   1.0080
MASS HGAN   1.0080
MASS HGAL   1.0080
MASS CGAE  12.0110
MASS HGAG   1.0080
MASS HGAH   1.0080
MASS NGAC  14.0067
MASS HGAJ   1.0080
MASS HGAK   1.0080
MASS HGAI   1.0080

Residue DADM
GROUP
 ATOM CAA  TYPE=CGAA CHARGE= 0.139 END
 ATOM HAA  TYPE=HGAA CHARGE= 0.004 END
 ATOM HAB  TYPE=HGAB CHARGE= 0.004 END
 ATOM HAC  TYPE=HGAC CHARGE= 0.004 END
 ATOM CAF  TYPE=CGAF CHARGE= 0.169 END
 ATOM CAB  TYPE=CGAB CHARGE= 0.139 END
 ATOM HAE  TYPE=HGAE CHARGE= 0.004 END
 ATOM HAF  TYPE=HGAF CHARGE= 0.004 END
 ATOM HAD  TYPE=HGAD CHARGE= 0.004 END
 ATOM NAD  TYPE=NGAD CHARGE= 0.657 END
 ATOM HAM  TYPE=HGAM CHARGE= 0.011 END
 ATOM HAN  TYPE=HGAN CHARGE= 0.011 END
 ATOM HAL  TYPE=HGAL CHARGE= 0.011 END
 ATOM CAE  TYPE=CGAE CHARGE= 0.141 END
 ATOM HAG  TYPE=HGAG CHARGE= 0.004 END
 ATOM HAH  TYPE=HGAH CHARGE= 0.004 END
 ATOM NAC  TYPE=NGAC CHARGE= 0.657 END
 ATOM HAJ  TYPE=HGAJ CHARGE= 0.011 END
 ATOM HAK  TYPE=HGAK CHARGE= 0.011 END
```

```
ATOM HAI  TYPE=HGAI CHARGE= 0.011 END

BOND CAA  HAA
BOND CAA  HAB
BOND CAA  HAC
BOND CAF  CAA
BOND CAF  CAB
BOND CAF  NAD
BOND CAF  CAE
BOND CAB  HAE
BOND CAB  HAF
BOND CAB  HAD
BOND NAD  HAM
BOND NAD  HAN
BOND NAD  HAL
BOND CAE  HAG
BOND CAE  HAH
BOND CAE  NAC
BOND NAC  HAJ
BOND NAC  HAK
BOND NAC  HAI

ANGLE HAA  CAA  HAB
ANGLE HAA  CAA  HAC
ANGLE HAA  CAA  CAF
ANGLE HAB  CAA  HAC
ANGLE HAB  CAA  CAF
ANGLE HAC  CAA  CAF
ANGLE CAA  CAF  CAB
ANGLE CAA  CAF  NAD
ANGLE CAA  CAF  CAE
ANGLE CAB  CAF  NAD
ANGLE CAB  CAF  CAE
ANGLE NAD  CAF  CAE
ANGLE CAF  CAB  HAE
ANGLE CAF  CAB  HAF
ANGLE CAF  CAB  HAD
ANGLE HAE  CAB  HAF
ANGLE HAE  CAB  HAD
ANGLE HAF  CAB  HAD
ANGLE CAF  NAD  HAM
ANGLE CAF  NAD  HAN
ANGLE CAF  NAD  HAL
ANGLE HAM  NAD  HAN
ANGLE HAM  NAD  HAL
```

```
 ANGLE HAN  NAD  HAL
 ANGLE CAF  CAE  HAG
 ANGLE CAF  CAE  HAH
 ANGLE CAF  CAE  NAC
 ANGLE HAG  CAE  HAH
 ANGLE HAG  CAE  NAC
 ANGLE HAH  CAE  NAC
 ANGLE CAE  NAC  HAJ
 ANGLE CAE  NAC  HAK
 ANGLE CAE  NAC  HAI
 ANGLE HAJ  NAC  HAK
 ANGLE HAJ  NAC  HAI
 ANGLE HAK  NAC  HAI

 IMPROPER CAA  HAA  HAC  HAB
 IMPROPER CAF  CAA  CAB  NAD
 IMPROPER CAB  CAF  HAE  HAF
 IMPROPER NAD  CAF  HAM  HAN
 IMPROPER CAE  CAF  HAH  HAG
 IMPROPER NAC  CAE  HAJ  HAK

 improper HAM HAN CAF HAL
 improper HAJ HAK CAE HAI
 improper HAE HAF CAF HAD
 improper HAA HAB CAF HAC

 improper CAA CAB NAD CAE
 improper CAF NAC HAH HAG


END { RESIdue DADM }
set echo=true end

presidue dadm

 group
  modify   atom 2HAI   type=H charge=0.0 end
  modify   atom 2HAL   type=H charge=0.0 end

   delete atom 2HAJ  charge=0.0 end
   delete atom 2HAK  charge=0.0 end

   delete atom 2HAM  charge=0.0 end
   delete atom 2HAN  charge=0.0 end
```

```
 add bond 1C 2NAC

 add angle 1CA 1C   2NAC
 add angle 1O  1C   2NAC
 add angle 1C  2NAC 2CAE
 add angle 1C  2NAC 2HAI

 add improper 1O  1C  2NAC  2CAE
 add improper 2HAI 2NAC  1C  1CA
 add improper 1CA 1C  2NAC  2CAE

 add bond 2NAD 3C

 add angle 3N  3C   2NAD
 add angle 3O  3C   2NAD
 add angle 3C  2NAD 2CAF
 add angle 3C  2NAD 2HAL

 add improper 3O   3C  2NAD  2CAF
 add improper 2HAL 2NAD  3C  3CA
 add improper 3CA  3C  2NAD  2CAF

end
```

DADME Parameters File:

```
!Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version 20150224
!
!
!     This file was generated by PRODRG version AA100323.0717
!     PRODRG written/copyrighted by Daan van Aalten
!     and Alexander Schuettelkopf
!
!     Questions/comments to dava@davapc1.bioch.dundee.ac.uk
!
!     When using this software in a publication, cite:
!     A. W. Schuettelkopf and D. M. F. van Aalten (2004).
!     PRODRG - a tool for high-throughput crystallography
!     of protein-ligand complexes.
!     Acta Crystallogr. D60, 1355--1363.
!
!
!  *** NOTE *** IF YOU USE MORE THAN ONE PRODRG-GENERATED TOPOLOGY IN CNS,
!            PLEASE CAREFULLY READ THE FAQ AS THIS MAY CAUSE PROBLEMS
!
!
```

```
set echo=off message=on end

evaluate ($pd_x = 1.0)
evaluate ($kbon = 1000)
evaluate ($kang =  500)
evaluate ($kchi =  500)
evaluate ($kback = 500)
evaluate ($kpx = $kback)
evaluate ($kpla =  500)

bond    C     NGAC          $kbon    1.305
bond    NGAD   C           $kbon    1.305

eval ($pd_v=$pd_x* 13971.0) BOND CGAA HGAA $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CGAA HGAB $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CGAA HGAC $pd_v   1.090
eval ($pd_v=$pd_x* 16001.4) BOND CGAF CGAA $pd_v   1.530
eval ($pd_v=$pd_x* 16001.4) BOND CGAF CGAB $pd_v   1.530
eval ($pd_v=$pd_x* 17993.7) BOND CGAF NGAD $pd_v   1.470
eval ($pd_v=$pd_x* 16001.4) BOND CGAF CGAE $pd_v   1.530
eval ($pd_v=$pd_x* 13971.0) BOND CGAB HGAE $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CGAB HGAF $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CGAB HGAD $pd_v   1.090
eval ($pd_v=$pd_x* 17877.6) BOND NGAD HGAM $pd_v   1.000
eval ($pd_v=$pd_x* 17877.6) BOND NGAD HGAN $pd_v   1.000
eval ($pd_v=$pd_x* 17877.6) BOND NGAD HGAL $pd_v   1.000
eval ($pd_v=$pd_x* 13971.0) BOND CGAE HGAG $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CGAE HGAH $pd_v   1.090
eval ($pd_v=$pd_x* 17993.7) BOND CGAE NGAC $pd_v   1.470
eval ($pd_v=$pd_x* 17877.6) BOND NGAC HGAJ $pd_v   1.000
eval ($pd_v=$pd_x* 17877.6) BOND NGAC HGAK $pd_v   1.000
eval ($pd_v=$pd_x* 17877.6) BOND NGAC HGAI $pd_v   1.000

eval ($pd_v=$pd_x* 17877.6) BOND NGAD H $pd_v   1.000
eval ($pd_v=$pd_x* 17877.6) BOND NGAC H $pd_v   1.000

angle CT C    NGAC          $kang    116.9
angle O  C    NGAC          $kang    122.0
angle C  NGAC CGAE          $kang    120
angle C  NGAC H             $kang    120

angle N    C    NGAD         $kang    120.0
angle O    C    NGAD         $kang    123.00
angle C    NGAD CGAF         $kang    122.0
angle C    NGAD   H          $kang    120.0
```

```
angle CGAE NGAC   H          $kang    120
angle CGAF NGAD   H          $kang    120

eval ($pd_v=$pd_x*  720.0) ANGLE HGAA CGAA HGAB $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAA CGAA HGAC $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAA CGAA CGAF $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAB CGAA HGAC $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAB CGAA CGAF $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAC CGAA CGAF $pd_v 109.500
eval ($pd_v=$pd_x*  883.5) ANGLE CGAA CGAF CGAB $pd_v 109.500
eval ($pd_v=$pd_x*  883.2) ANGLE CGAA CGAF NGAD $pd_v 111.000
eval ($pd_v=$pd_x*  883.5) ANGLE CGAA CGAF CGAE $pd_v 109.500
eval ($pd_v=$pd_x*  883.2) ANGLE CGAB CGAF NGAD $pd_v 111.000
eval ($pd_v=$pd_x*  883.5) ANGLE CGAB CGAF CGAE $pd_v 109.500
eval ($pd_v=$pd_x*  883.2) ANGLE NGAD CGAF CGAE $pd_v 111.000
eval ($pd_v=$pd_x*  720.0) ANGLE CGAF CGAB HGAE $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE CGAF CGAB HGAF $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE CGAF CGAB HGAD $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAE CGAB HGAF $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAE CGAB HGAD $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAF CGAB HGAD $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAF NGAD HGAM $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAF NGAD HGAN $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAF NGAD HGAL $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAM NGAD HGAN $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAM NGAD HGAL $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAN NGAD HGAL $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE CGAF CGAE HGAG $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE CGAF CGAE HGAH $pd_v 109.500
eval ($pd_v=$pd_x*  883.2) ANGLE CGAF CGAE NGAC $pd_v 111.000
eval ($pd_v=$pd_x*  720.0) ANGLE HGAG CGAE HGAH $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAG CGAE NGAC $pd_v 109.500
eval ($pd_v=$pd_x*  720.0) ANGLE HGAH CGAE NGAC $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAE NGAC HGAJ $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAE NGAC HGAK $pd_v 109.500
eval ($pd_v=$pd_x*  722.1) ANGLE CGAE NGAC HGAI $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAJ NGAC HGAK $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAJ NGAC HGAI $pd_v 109.500
eval ($pd_v=$pd_x*  645.6) ANGLE HGAK NGAC HGAI $pd_v 109.500


improper  O    C    NGAC  CGAE    $kpx    0    0.0
improper  H    NGAC C     CT      $kchi   0    0
improper  CT   C    NGAC  CGAE    $kpx    0  180.0
```

```
improper  O    C    NGAD  CGAF    $kpx    0    0.0
improper  H    NGAD C     CT      $kchi    0    0
improper  CT   C    NGAD  CGAF    $kpx    0  180.0

eval ($pd_v=$pd_x*   160.1) IMPR CGAA HGAA HGAC HGAB $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR CGAF CGAA CGAB NGAD $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR CGAB CGAF HGAE HGAF $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR NGAD CGAF HGAM HGAN $pd_v 0  33
eval ($pd_v=$pd_x*   160.1) IMPR CGAE CGAF HGAH HGAG $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR NGAC CGAE HGAJ HGAK $pd_v 0  33

improper  HGAM HGAN CGAF HGAL      $kchi   0   -66.514
improper  HGAJ HGAK CGAE HGAI      $kchi   0   -66.514
improper  HGAE HGAF CGAF HGAD      $kchi   0   -66.514
improper  HGAA HGAB CGAF HGAC      $kchi   0   -66.514

improper  CGAA CGAB NGAD CGAE      $kchi   0   -70
improper  CGAF NGAC HGAH HGAG      $kchi   0   -70


NBONds
 TOLERANCE=0.5 NBXMOD=5 WMIN=1.5
 REPEL=1.0 REXPONENT=4 IREXPONENT=1 RCONST=16.0
 CTONNB=5.5 CTOFNB=6.0 CUTNB=7.0
END

NONBONDED CGAA  0.10000 3.29633 0.10000 3.02906
NONBONDED HGAA  0.10000 2.13816 0.10000 1.87089
NONBONDED HGAB  0.10000 2.13816 0.10000 1.87089
NONBONDED HGAC  0.10000 2.13816 0.10000 1.87089
NONBONDED CGAF  0.10000 3.29633 0.10000 3.02906
NONBONDED CGAB  0.10000 3.29633 0.10000 3.02906
NONBONDED HGAE  0.10000 2.13816 0.10000 1.87089
NONBONDED HGAF  0.10000 2.13816 0.10000 1.87089
NONBONDED HGAD  0.10000 2.13816 0.10000 1.87089
NONBONDED NGAD  0.10000 2.67270 0.10000 2.40543
NONBONDED HGAM  0.10000 1.42544 0.10000 1.15817
NONBONDED HGAN  0.10000 1.42544 0.10000 1.15817
NONBONDED HGAL  0.10000 1.42544 0.10000 1.15817
NONBONDED CGAE  0.10000 3.29633 0.10000 3.02906
NONBONDED HGAG  0.10000 2.13816 0.10000 1.87089
NONBONDED HGAH  0.10000 2.13816 0.10000 1.87089
NONBONDED NGAC  0.10000 2.67270 0.10000 2.40543
NONBONDED HGAJ  0.10000 1.42544 0.10000 1.15817
NONBONDED HGAK  0.10000 1.42544 0.10000 1.15817
```

```
NONBONDED HGAI  0.10000 1.42544 0.10000 1.15817

set echo=on message=on end
```

A3.3. Coordinates, Topology, and Parameters Files for Building Non-Natural Residue CHDA

CHDA Coordinates File:

```
REMARK  Vanessa M. Kung, 20140203
REMARK
REMARK
REMARK  This file was generated by PRODRG version AA100323.0717
REMARK  PRODRG written/copyrighted by Daan van Aalten
REMARK  and Alexander Schuettelkopf
REMARK
REMARK  Questions/comments to dava@davapc1.bioch.dundee.ac.uk
REMARK
REMARK  When using this software in a publication, cite:
REMARK  A. W. Schuettelkopf and D. M. F. van Aalten (2004).
REMARK  PRODRG - a tool for high-throughput crystallography
REMARK  of protein-ligand complexes.
REMARK  Acta Crystallogr. D60, 1355--1363.
REMARK
REMARK
HETATM    1  OAA DRG    1      3.820   5.820   1.040  1.00 20.00         O
HETATM    2  CAI DRG    1      4.440   5.310   0.100  1.00 20.00         C
HETATM    3  FAC DRG    1      5.750   5.470  -0.120  1.00 20.00         F
HETATM    4  CAK DRG    1      3.670   4.550  -0.990  1.00 20.00         C
HETATM    5  HAM DRG    1      4.256   4.553  -1.909  1.00 20.00         H
HETATM    6  CAG DRG    1      2.340   5.290  -1.240  1.00 20.00         C
HETATM    7  HAI DRG    1      2.455   6.280  -0.799  1.00 20.00         H
HETATM    8  HAJ DRG    1      2.165   5.315  -2.316  1.00 20.00         H
HETATM    9  CAE DRG    1      1.100   4.660  -0.570  1.00 20.00         C
HETATM   10  HAE DRG    1      0.396   5.473  -0.390  1.00 20.00         H
HETATM   11  HAF DRG    1      0.711   3.873  -1.216  1.00 20.00         H
HETATM   12  CAF DRG    1      1.440   4.060   0.790  1.00 20.00         C
HETATM   13  HAG DRG    1      0.537   3.626   1.219  1.00 20.00         H
HETATM   14  HAH DRG    1      1.871   4.862   1.389  1.00 20.00         H
HETATM   15  CAH DRG    1      2.510   2.970   0.680  1.00 20.00         C
HETATM   16  HAK DRG    1      2.012   2.001   0.648  1.00 20.00         H
HETATM   17  HAL DRG    1      3.168   3.093   1.541  1.00 20.00         H
HETATM   18  CAL DRG    1      3.400   3.100  -0.570  1.00 20.00         C
HETATM   19  HAN DRG    1      2.831   2.655  -1.386  1.00 20.00         H
HETATM   20  CAJ DRG    1      4.650   2.230  -0.410  1.00 20.00         C
HETATM   21  OAB DRG    1      4.620   1.260   0.330  1.00 20.00         O
```

```
HETATM  22  FAD DRG    1      5.680   2.540  -1.220  1.00 20.00        F
CONECT   1   2
CONECT   2   1   3   4
CONECT   3   2
CONECT   4   2   5   6  18
CONECT   5   4
CONECT   6   4   7   8   9
CONECT   7   6
CONECT   8   6
CONECT   9   6  10  11  12
CONECT  10   9
CONECT  11   9
CONECT  12   9  13  14  15
CONECT  13  12
CONECT  14  12
CONECT  15  12  16  17  18
CONECT  16  15
CONECT  17  15
CONECT  18   4  15  19  20
CONECT  19  18
CONECT  20  18  21  22
CONECT  21  20
CONECT  22  20
END
```

CHDA Topology File:

```
!Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version 20150224
!
!
!     This file was generated by PRODRG version AA100323.0717
!     PRODRG written/copyrighted by Daan van Aalten
!     and Alexander Schuettelkopf
!
!     Questions/comments to dava@davapc1.bioch.dundee.ac.uk
!
!     When using this software in a publication, cite:
!     A. W. Schuettelkopf and D. M. F. van Aalten (2004).
!     PRODRG - a tool for high-throughput crystallography
!     of protein-ligand complexes.
!     Acta Crystallogr. D60, 1355--1363.
!
!
! *** NOTE *** IF YOU USE MORE THAN ONE PRODRG-GENERATED TOPOLOGY IN CNS,
!           PLEASE CAREFULLY READ THE FAQ AS THIS MAY CAUSE PROBLEMS
!
```

```
!
!  *** NOTE *** IF YOU WANT TO USE THIS TOPOLOGY FOR CRYSTALLOGRAPHIC
!            REFINEMENT YOU WILL NEED TO MODIFY YOUR scatter.lib. A
!            WORKING EXAMPLE CAN BE DOWNLOADED FROM
!  http://davapc1.bioch.dundee.ac.uk/programs/prodrg/stuff/scatter_1.1_mod.lib
!
!
set echo=false end
AUTOGENERATE  ANGLES=FALSE  END
mass H    1.008
mass C   12.011
mass O   15.999
mass N   14.007
MASS OBAA  15.9994
MASS CBAI  12.0110
MASS FBAC  18.9984
MASS CBAK  12.0110
MASS HBAM   1.0080
MASS CBAG  12.0110
MASS HBAI   1.0080
MASS HBAJ   1.0080
MASS CBAE  12.0110
MASS HBAE   1.0080
MASS HBAF   1.0080
MASS CBAF  12.0110
MASS HBAG   1.0080
MASS HBAH   1.0080
MASS CBAH  12.0110
MASS HBAK   1.0080
MASS HBAL   1.0080
MASS CBAL  12.0110
MASS HBAN   1.0080
MASS CBAJ  12.0110
MASS OBAB  15.9994
MASS FBAD  18.9984

Residue CHDA
GROUP
 ATOM OAA  TYPE=OBAA CHARGE=-0.689 END
 ATOM CAI  TYPE=CBAI CHARGE= 0.378 END
 ATOM FAC  TYPE=FBAC CHARGE=-0.200 END
 ATOM CAK  TYPE=CBAK CHARGE= 0.174 END
 ATOM HAM  TYPE=HBAM CHARGE= 0.009 END
 ATOM CAG  TYPE=CBAG CHARGE= 0.146 END
 ATOM HAI  TYPE=HBAI CHARGE= 0.009 END
```

```
ATOM HAJ  TYPE=HBAJ CHARGE= 0.009 END
ATOM CAE  TYPE=CBAE CHARGE= 0.146 END
ATOM HAE  TYPE=HBAE CHARGE= 0.009 END
ATOM HAF  TYPE=HBAF CHARGE= 0.009 END
ATOM CAF  TYPE=CBAF CHARGE= 0.146 END
ATOM HAG  TYPE=HBAG CHARGE= 0.009 END
ATOM HAH  TYPE=HBAH CHARGE= 0.009 END
ATOM CAH  TYPE=CBAH CHARGE= 0.146 END
ATOM HAK  TYPE=HBAK CHARGE= 0.010 END
ATOM HAL  TYPE=HBAL CHARGE= 0.009 END
ATOM CAL  TYPE=CBAL CHARGE= 0.174 END
ATOM HAN  TYPE=HBAN CHARGE= 0.009 END
ATOM CAJ  TYPE=CBAJ CHARGE= 0.377 END
ATOM OAB  TYPE=OBAB CHARGE=-0.689 END
ATOM FAD  TYPE=FBAD CHARGE=-0.200 END

BOND CAI  OAA
BOND CAI  FAC
BOND CAK  CAI
BOND CAK  HAM
BOND CAK  CAG
BOND CAK  CAL
BOND CAG  HAI
BOND CAG  HAJ
BOND CAG  CAE
BOND CAE  HAE
BOND CAE  HAF
BOND CAE  CAF
BOND CAF  HAG
BOND CAF  HAH
BOND CAF  CAH
BOND CAH  HAK
BOND CAH  HAL
BOND CAL  CAH
BOND CAL  HAN
BOND CAL  CAJ
BOND CAJ  OAB
BOND CAJ  FAD

ANGLE OAA  CAI  FAC
ANGLE OAA  CAI  CAK
ANGLE FAC  CAI  CAK
ANGLE CAI  CAK  HAM
ANGLE CAI  CAK  CAG
ANGLE CAI  CAK  CAL
```

```
ANGLE HAM  CAK  CAG
ANGLE HAM  CAK  CAL
ANGLE CAG  CAK  CAL
ANGLE CAK  CAG  HAI
ANGLE CAK  CAG  HAJ
ANGLE CAK  CAG  CAE
ANGLE HAI  CAG  HAJ
ANGLE HAI  CAG  CAE
ANGLE HAJ  CAG  CAE
ANGLE CAG  CAE  HAE
ANGLE CAG  CAE  HAF
ANGLE CAG  CAE  CAF
ANGLE HAE  CAE  HAF
ANGLE HAE  CAE  CAF
ANGLE HAF  CAE  CAF
ANGLE CAE  CAF  HAG
ANGLE CAE  CAF  HAH
ANGLE CAE  CAF  CAH
ANGLE HAG  CAF  HAH
ANGLE HAG  CAF  CAH
ANGLE HAH  CAF  CAH
ANGLE CAF  CAH  HAK
ANGLE CAF  CAH  HAL
ANGLE CAF  CAH  CAL
ANGLE HAK  CAH  HAL
ANGLE HAK  CAH  CAL
ANGLE HAL  CAH  CAL
ANGLE CAK  CAL  CAH
ANGLE CAK  CAL  HAN
ANGLE CAK  CAL  CAJ
ANGLE CAH  CAL  HAN
ANGLE CAH  CAL  CAJ
ANGLE HAN  CAL  CAJ
ANGLE CAL  CAJ  OAB
ANGLE CAL  CAJ  FAD
ANGLE OAB  CAJ  FAD

IMPROPER CAI  OAA  FAC  CAK
IMPROPER CAK  CAI  CAG  CAL
IMPROPER CAG  CAK  HAJ  HAI
IMPROPER CAE  CAG  HAF  HAE
IMPROPER CAF  CAE  HAG  HAH
IMPROPER CAH  CAF  HAK  HAL
IMPROPER CAL  CAJ  CAK  CAH
IMPROPER CAJ  CAL  OAB  FAD
```

```
 improper  CAE CAH HAG HAH
 improper  CAG CAF HAE HAF
 improper  CAF CAL HAK HAL
 improper  CAE CAK HAI HAJ
 improper  CAK CAJ CAH HAN
 improper  CAL CAI CAG HAM

 DIHEDRAL CAE  CAG  CAK  CAI
 DIHEDRAL CAJ  CAL  CAK  CAI
 DIHEDRAL CAJ  CAL  CAH  CAF

 DIHEDRAL CAK  CAG  CAE  CAF
 DIHEDRAL CAG  CAE  CAF  CAH
 DIHEDRAL CAE  CAF  CAH  CAL
 DIHEDRAL CAF  CAH  CAL  CAK
 DIHEDRAL CAH  CAL  CAK  CAG
 DIHEDRAL CAL  CAK  CAG  CAE

END { RESIdue CHDA }
set echo=true end

presidue chda

 group
   delete atom 2FAD  charge=0.0 end
   delete atom 2FAC  charge=0.0 end

 add bond 1N 2CAJ

 add angle 2CAL 2CAJ 1N
 add angle 2OAB 2CAJ 1N
 add angle 2CAJ 1N   1CA
 add angle 2CAJ 1N   1HN

 add improper 2OAB 2CAJ 1N   1CA
 add improper 1HN  1N   2CAJ 2CAL
 add improper 2CAL 2CAJ 1N   1CA

 add bond 2CAI 3N

 add angle 2CAK 2CAI 3N
 add angle 2OAA 2CAI 3N
 add angle 2CAI 3N   3CA
 add angle 2CAI 3N   3HN
```

```
 add improper 2OAA 2CAI 3N   3CA
 add improper 3HN  3N   2CAI 2CAK
 add improper 2CAK 2CAI 3N   3CA    "

end
```

CHDA Parameters File:

```
! Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version 20150224
!
!     This file was generated by PRODRG version AA100323.0717
!     PRODRG written/copyrighted by Daan van Aalten
!     and Alexander Schuettelkopf
!
!     Questions/comments to dava@davapc1.bioch.dundee.ac.uk
!
!     When using this software in a publication, cite:
!     A. W. Schuettelkopf and D. M. F. van Aalten (2004).
!     PRODRG - a tool for high-throughput crystallography
!     of protein-ligand complexes.
!     Acta Crystallogr. D60, 1355--1363.
!
!
! *** NOTE *** IF YOU USE MORE THAN ONE PRODRG-GENERATED TOPOLOGY IN CNS,
!           PLEASE CAREFULLY READ THE FAQ AS THIS MAY CAUSE PROBLEMS
!
!
! *** NOTE *** IF YOU WANT TO USE THIS TOPOLOGY FOR CRYSTALLOGRAPHIC
!           REFINEMENT YOU WILL NEED TO MODIFY YOUR scatter.lib. A
!           WORKING EXAMPLE CAN BE DOWNLOADED FROM
!  http://davapc1.bioch.dundee.ac.uk/programs/prodrg/stuff/scatter_1.1_mod.lib
!
!
set echo=off message=on end

evaluate ($pd_x = 1.0)
evaluate ($kbon = 1000)
evaluate ($kang =  500)
evaluate ($kchi =  500)
evaluate ($kback = 500)
evaluate ($kpx = $kback)
evaluate ($kpla =  500)

bond    CBAJ   NH1          $kbon    1.305
bond    NH1    CBAI         $kbon    1.305
```

```
eval ($pd_v=$pd_x* 24009.7) BOND CBAI OBAA $pd_v   1.230
eval ($pd_v=$pd_x*  7416.9) BOND CBAI FBAC $pd_v   1.348
eval ($pd_v=$pd_x* 16001.4) BOND CBAK CBAI $pd_v   1.530
eval ($pd_v=$pd_x* 13971.0) BOND CBAK HBAM $pd_v   1.090
eval ($pd_v=$pd_x* 11993.8) BOND CBAK CBAG $pd_v   1.520
eval ($pd_v=$pd_x* 11993.8) BOND CBAK CBAL $pd_v   1.520
eval ($pd_v=$pd_x* 13971.0) BOND CBAG HBAI $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CBAG HBAJ $pd_v   1.090
eval ($pd_v=$pd_x* 11993.8) BOND CBAG CBAE $pd_v   1.520
eval ($pd_v=$pd_x* 13971.0) BOND CBAE HBAE $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CBAE HBAF $pd_v   1.090
eval ($pd_v=$pd_x* 11993.8) BOND CBAE CBAF $pd_v   1.520
eval ($pd_v=$pd_x* 13971.0) BOND CBAF HBAG $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CBAF HBAH $pd_v   1.090
eval ($pd_v=$pd_x* 11993.8) BOND CBAF CBAH $pd_v   1.520
eval ($pd_v=$pd_x* 13971.0) BOND CBAH HBAK $pd_v   1.090
eval ($pd_v=$pd_x* 13971.0) BOND CBAH HBAL $pd_v   1.090
eval ($pd_v=$pd_x* 11993.8) BOND CBAL CBAH $pd_v   1.520
eval ($pd_v=$pd_x* 13971.0) BOND CBAL HBAN $pd_v   1.090
eval ($pd_v=$pd_x* 16001.4) BOND CBAL CBAJ $pd_v   1.530
eval ($pd_v=$pd_x* 24009.7) BOND CBAJ OBAB $pd_v   1.230
eval ($pd_v=$pd_x*  7416.9) BOND CBAJ FBAD $pd_v   1.348

angle CBAL CBAJ NH1          $kang    116.9
angle OBAB CBAJ NH1          $kang    122.0
angle CBAJ NH1   CT          $kang    122.0
angle CBAJ NH1    H          $kang    120.0

angle CBAK CBAI NH1          $kang    116.9
angle OBAA CBAI NH1          $kang    122.0
angle CBAI NH1   CT          $kang    122.0
angle CBAI NH1    H          $kang    120.0

eval ($pd_v=$pd_x*   500.0) ANGLE OBAA CBAI FBAC $pd_v 120.000
eval ($pd_v=$pd_x*   500.0) ANGLE OBAA CBAI CBAK $pd_v 120.000
eval ($pd_v=$pd_x*   500.0) ANGLE FBAC CBAI CBAK $pd_v 120.000
eval ($pd_v=$pd_x*   500.0) ANGLE CBAI CBAK HBAM $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAI CBAK CBAG $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAI CBAK CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAM CBAK CBAG $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAM CBAK CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAG CBAK CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAG HBAI $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAG HBAJ $pd_v 109.500
```

```
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAG CBAE $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAI CBAG HBAJ $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAI CBAG CBAE $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAJ CBAG CBAE $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAG CBAE HBAE $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAG CBAE HBAF $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAG CBAE CBAF $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAE CBAE HBAF $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAE CBAE CBAF $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAF CBAE CBAF $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAE CBAF HBAG $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAE CBAF HBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAE CBAF CBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAG CBAF HBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAG CBAF CBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAH CBAF CBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAF CBAH HBAK $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAF CBAH HBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAF CBAH CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAK CBAH HBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAK CBAH CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAL CBAH CBAL $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAL CBAH $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAL HBAN $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAK CBAL CBAJ $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAH CBAL HBAN $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAH CBAL CBAJ $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE HBAN CBAL CBAJ $pd_v 109.500
eval ($pd_v=$pd_x*   500.0) ANGLE CBAL CBAJ OBAB $pd_v 120.000
eval ($pd_v=$pd_x*   500.0) ANGLE CBAL CBAJ FBAD $pd_v 120.000
eval ($pd_v=$pd_x*   500.0) ANGLE OBAB CBAJ FBAD $pd_v 120.000

improper OBAB CBAJ NH1  CT      $kpx     0    0.0
improper H    NH1  CBAJ CBAL    $kpx     0    0.0
improper CBAL CBAJ NH1  CT      $kpx     0  180.0

improper OBAA CBAI NH1  CT      $kpx     0    0.0
improper H    NH1  CBAI CBAK    $kpx     0    0.0
improper CBAK CBAI NH1  CT      $kpx     0  180.0

eval ($pd_v=$pd_x*    80.0) IMPR CBAI OBAA FBAC CBAK $pd_v 0   0.000
eval ($pd_v=$pd_x*   160.1) IMPR CBAK CBAI CBAG CBAL $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR CBAG CBAK HBAJ HBAI $pd_v 0  35.264
eval ($pd_v=$pd_x*   160.1) IMPR CBAE CBAG HBAF HBAE $pd_v 0  -35.264
eval ($pd_v=$pd_x*   160.1) IMPR CBAF CBAE HBAG HBAH $pd_v 0  -35.264
```

```
eval ($pd_v=$pd_x*   160.1) IMPR CBAH CBAF HBAK HBAL $pd_v 0  -35.264
eval ($pd_v=$pd_x*   160.1) IMPR CBAL CBAJ CBAK CBAH $pd_v 0  35.264
eval ($pd_v=$pd_x*    80.0) IMPR CBAJ CBAL OBAB FBAD $pd_v 0   0.000

improper  CBAE CBAH HBAG HBAH $kchi 0 70
improper  CBAG CBAF HBAE HBAF $kchi 0 -70
improper  CBAF CBAL HBAK HBAL $kchi 0 70
improper  CBAE CBAK HBAI HBAJ $kchi 0 -70

improper  CBAK CBAJ CBAH HBAN $kchi 0 66.5
improper  CBAL CBAI CBAG HBAM $kchi 0 -66.5

eval ($pd_v=$pd_x*    19.6) DIHE CBAE CBAG CBAK CBAI $pd_v 3   0.000
eval ($pd_v=$pd_x*    19.6) DIHE CBAJ CBAL CBAK CBAI $pd_v 3   0.000
eval ($pd_v=$pd_x*    19.6) DIHE CBAJ CBAL CBAH CBAF $pd_v 3   0.000

eval ($pd_v=$pd_x*   19.6) DIHE CBAK CBAG CBAE CBAF $pd_v 3  0.000
eval ($pd_v=$pd_x*   19.6) DIHE CBAG CBAE CBAF CBAH $pd_v 3  0.000
eval ($pd_v=$pd_x*   19.6) DIHE CBAE CBAF CBAH CBAL $pd_v 3  0.000
eval ($pd_v=$pd_x*   19.6) DIHE CBAF CBAH CBAL CBAK $pd_v 3  0.000
eval ($pd_v=$pd_x*   19.6) DIHE CBAH CBAL CBAK CBAG $pd_v 3  0.000
eval ($pd_v=$pd_x*   19.6) DIHE CBAL CBAK CBAG CBAE $pd_v 3  0.000


NBONds
 TOLERANCE=0.5 NBXMOD=5 WMIN=1.5
 REPEL=1.0 REXPONENT=4 IREXPONENT=1 RCONST=16.0
 CTONNB=5.5 CTOFNB=6.0 CUTNB=7.0
END

NONBONDED OBAA  0.10000 2.58361 0.10000 2.31634
NONBONDED CBAI  0.10000 3.29633 0.10000 3.02906
NONBONDED FBAC  0.10000 3.56359 0.10000 3.29633
NONBONDED CBAK  0.10000 3.29633 0.10000 3.02906
NONBONDED HBAM  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAG  0.10000 3.29633 0.10000 3.02906
NONBONDED HBAI  0.10000 2.13816 0.10000 1.87089
NONBONDED HBAJ  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAE  0.10000 3.29633 0.10000 3.02906
NONBONDED HBAE  0.10000 2.13816 0.10000 1.87089
NONBONDED HBAF  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAF  0.10000 3.29633 0.10000 3.02906
NONBONDED HBAG  0.10000 2.13816 0.10000 1.87089
NONBONDED HBAH  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAH  0.10000 3.29633 0.10000 3.02906
```

```
NONBONDED HBAK  0.10000 2.13816 0.10000 1.87089
NONBONDED HBAL  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAL  0.10000 3.29633 0.10000 3.02906
NONBONDED HBAN  0.10000 2.13816 0.10000 1.87089
NONBONDED CBAJ  0.10000 3.29633 0.10000 3.02906
NONBONDED OBAB  0.10000 2.58361 0.10000 2.31634
NONBONDED FBAD  0.10000 3.56359 0.10000 3.29633

set echo=on message=on end
```

A3.4. Python Script for Generating Coordinates Files for a Parallel β-Sheet Peptide Containing DADME and CHDA Residues (e.g., Peptide 2.A)

```
# Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version
20140520
#
# generate PSF and initial pdb w/ correct covalent geometry for
# a 3-stranded parallel b-sheet peptide with sequence:
#CTN LYS GLU ARG ILE THR VAL GLY CHDA
#VAL ARG ILE PHE LYS GLN VAL PRO DADM
#LYS GLU TYR PHE ARG GLU ACE
#
#

xplor.requireVersion("2.35")

xplor.parseArguments()

seq="""
LYS GLU ARG ILE THR VAL GLY CHDA
VAL ARG ILE PHE LYS GLN VAL PRO DADM
LYS GLU TYR PHE ARG GLU ACE
"""

import protocol
protocol.initTopology(('protein','DADM.top','CHDA.top'))
protocol.initParams(('protein','DADM.par','CHDA.par'))
protocol.initRandomSeed()

import psfGen

seq1="LYS GLU ARG ILE THR VAL GLY".split()

seq1.reverse()
psfGen.seqToPSF(seq1,startResid=1,ntermPatch='',amidate_cterm=True)

for atom in AtomSel("resid 1:7"):
    atom.setResidueNum( 24-atom.residueNum() )
    pass
```

```
seq2="VAL ARG ILE PHE LYS GLN VAL PRO".split()
psfGen.seqToPSF(seq2,startResid=9,ctermPatch='',ntermPatch='')

seq3="LYS GLU TYR PHE ARG GLU ACE".split()

seq3.reverse()
psfGen.seqToPSF(seq3,startResid=18,ctermPatch='',ntermPatch='')

for atom in AtomSel("resid 18:24"):
    atom.setResidueNum( 24-atom.residueNum() )
    pass


xplor.command("""
patch LTOD
  reference=nil=( resid 16 )
end

!insert CHDA
segment
SETUP=TRUE
number 8
  chain
    sequence CHDA
    end
  end
end

patch CHDA
  reference=1=( resid 7)
  reference=2=( resid 8)
  reference=3=( resid 9)
end

!insert DADM
segment
SETUP=TRUE
number 17
  chain
    sequence DADM
    end
  end
end

patch DADM
  reference=1=( resid 16)
  reference=2=( resid 17)
  reference=3=( resid 18)
end
```

```
""")


xplor.command("write psf output=betaPep.psf end")


try:
    protocol.genExtendedStructure(verbose=1,maxFixupIters=50)
except protocol.CovalentViolation:
    pass

protocol.writePDB("betaPep.pdb")
```

A3.5. Python Script for Annealing a Parallel β-Sheet Peptide Containing DADME and CHDA Residues (e.g., Peptide 2.A)

```
#Vanessa M. Kung, Gabriel Cornilescu, Charles D. Schwieters, version
20150209
#
# Anneal a parallel beta sheet peptide with sequence:
#ACE GLU ARG PHE TYR GLU LYS DADM PRO VAL GLN
#LYS PHE ILE ARG VAL CHDA GLY VAL THR ILE
#ARG GLU LYS CTN
#


xplor.requireVersion("2.35")

(opts,args) = xplor.parseArguments(["quick"])

quick=False
for opt in opts:
    if opt[0]=="quick":
        quick=True
        pass
    pass

if quick:
    numberOfStructures=3
    pass


seq="""
ACE
GLU ARG PHE TYR GLU LYS DADM
PRO VAL GLN LYS PHE ILE ARG VAL
    GLY VAL THR ILE ARG GLU LYS
"""

import protocol
```

```
protocol.initTopology(('protein','DADM.top','CHDA.top'))
protocol.initParams(('protein','DADM.par','CHDA.par'),use_dihe=True)
protocol.initRandomSeed()


import psfGen
psfGen.seqToPSF("ACE GLU ARG PHE TYR GLU LYS",
startResid=0,ctermPatch='')

seq2="PRO VAL GLN LYS PHE ILE ARG VAL".split()

seq2.reverse()
psfGen.seqToPSF(seq2,startResid=8,ctermPatch='',ntermPatch='')

for atom in AtomSel("resid 8:15"):
    atom.setResidueNum( 23-atom.residueNum() )
    pass

psfGen.seqToPSF("GLY VAL THR ILE ARG GLU
LYS",startResid=17,ntermPatch='',amidate_cterm=True)


xplor.command("""
patch LTOD
  reference=nil=( resid 8 )
end

!insert DADM
segment
SETUP=TRUE
number 7
  chain
    sequence DADM
    end
  end
end

patch DADM
  reference=1=( resid 6)
  reference=2=( resid 7)
  reference=3=( resid 8)
end

!insert CHDA
segment
SETUP=TRUE
number 16
  chain
    sequence CHDA
    end
```

```
  end
end

patch CHDA
  reference=1=( resid 15)
  reference=2=( resid 16)
  reference=3=( resid 17)
end
""")

try:
    protocol.genExtendedStructure(verbose=1,maxFixupIters=500)
except protocol.CovalentViolation:
    pass


outFilename = "SCRIPT_STRUCTURE.pdb"
numberOfStructures=100

startFile="betaPep.pdb"


from potList import PotList
potList = PotList()


from simulationTools import MultRamp, StaticRamp, InitialParams

rampedParams=[]
highTempParams=[]

noe=PotList('noe')
potList.append(noe)
from noePotTools import create_NOEPot
for (name,scale,file) in [('all',1,"NOEs-sq.tbl"),
                          ('hb',1,"hbonds.tbl"),

                          ]:
    pot = create_NOEPot(name,file)


    pot.setScale(scale)
    noe.append(pot)
rampedParams.append( MultRamp(2,30, "noe.setScale( VALUE )") )


from xplorPot import XplorPot
protocol.initDihedrals("phi_psi_dihe.tbl",
```

```
                          )
potList.append( XplorPot('CDIH') )
highTempParams.append( StaticRamp("potList['CDIH'].setScale(10)") )
rampedParams.append( StaticRamp("potList['CDIH'].setScale(200)") )


potList['CDIH'].setThreshold( 5 )


from torsionDBPotTools import create_TorsionDBPot
torsionDBPot = create_TorsionDBPot('tDB')
potList.append( torsionDBPot )
rampedParams.append( MultRamp(.002,2,"torsionDBPot.setScale(VALUE)") )

potList.append( XplorPot('VDW') )
rampedParams.append( StaticRamp("protocol.initNBond()") )
rampedParams.append( MultRamp(0.9,0.8,
                          "xplor.command('param nbonds repel VALUE end
end')") )
rampedParams.append( MultRamp(.004,4,
                          "xplor.command('param nbonds rcon VALUE end
end')") )

highTempParams.append( StaticRamp("""protocol.initNBond(cutnb=100,
                                                        rcon=0.004,
                                                        tolerance=45,
                                                        repel=1.2,
                                                        onlyCA=1)""")
)

potList.append( XplorPot("BOND") )
potList.append( XplorPot("ANGL") )
potList['ANGL'].setThreshold( 5 )
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)")
)
potList.append( XplorPot("IMPR") )
potList['IMPR'].setThreshold( 5 )
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)")
)


potList.append( XplorPot('DIHE') )
potList['DIHE'].setThreshold( 10 )
rampedParams.append( MultRamp(0.1,1,"potList['DIHE'].setScale(VALUE)")
)

from xplorPot import XplorPot
print XplorPot('DIHE').calcEnergy()
xplor.command('print thresh=0 dihe')
```

```
from simulationTools import analyze
print analyze([XplorPot('IMPR'),XplorPot('DIHE')])



protocol.massSetup()



from ivm import IVM
dyn = IVM()
dyn.breakBond("resid 16 and name CAE","resid 16 and name CAF")
protocol.cartesianTopology(dyn,
                           sel=" or ".join( ["resid 16 and name "+a
for a in
                                             ('CAK', 'CAL', 'CAH',
                                              'CAF', 'CAE', 'CAG')]) )

protocol.torsionTopology(dyn)



minc = IVM()
protocol.cartesianTopology(minc)
minc.breakBond("resid 16 and name CAE","resid 16 and name CAF")
protocol.initMinimize(minc)



from simulationTools import AnnealIVM
init_t  = 3500.
cool = AnnealIVM(initTemp =init_t,
                 finalTemp=25,
                 tempStep =12.5,
                 ivm=dyn,
                 rampedParams = rampedParams)

def accept(potList):
    """
    return True if current structure meets acceptance criteria
    """
    if potList['noe'].violations()>5:
        return False


    if potList['CDIH'].violations()>5:
        return False
    if potList['DIHE'].violations()>5:
        return False
    if potList['BOND'].violations()>5:
```

```
        return False
    if potList['ANGL'].violations()>5:
        return False
    if potList['IMPR'].violations()>5:
        return False

    return True



def calcOneStructure(loopInfo):
    """ this function calculates a single structure, performs analysis
on the
    structure, and then writes out a pdb file, with remarks.
    """



    InitialParams( rampedParams )


    InitialParams( highTempParams )


    protocol.initDynamics(dyn,
                          potList=potList,
                          bathTemp=init_t,
                          initVelocities=1,
                          finalTime=10,
                          numSteps=5000,
                          printInterval=100)

    dyn.setETolerance( init_t/100 )
    dyn.run()


    InitialParams( rampedParams )



    protocol.initDynamics(dyn,
                          potList=potList,
                          numSteps=100,
                          finalTime=.2 ,
                          printInterval=100)



    cool.run()
```

```
    protocol.initMinimize(dyn,
                          printInterval=50)
    dyn.run()


    protocol.initDynamics(minc,
                          potList=potList,
                          numSteps=100,
                          finalTime=.4 ,
                          printInterval=100)


    protocol.initMinimize(minc,
                          potList=potList,
                          dEPred=10)
    minc.run()


    loopInfo.writeStructure(potList)
    pass

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
              pdbTemplate=outFilename,
              structLoopAction=calcOneStructure,
              genViolationStats=1,
              averagePotList=potList,

averageSortPots=[potList['BOND'],potList['ANGL'],potList['DIHE'],potLi
st['IMPR'],
                               noe,potList['CDIH']],

              averageTopFraction=0.1,

              averageContext=FinalParams(rampedParams),
              averageFilename="SCRIPT_ave.pdb",
              averageFitSel="name CA",
              averageCompSel="not resname ANI and not name H*"
).run()
```

Appendix 4. Studying Peptide Conformations with Cold, Gas-Phase Ion Spectroscopy

A4.1. Cold, Gas-Phase Ion Spectroscopy

Gas phase (vibrational, electronic, or rotational) spectroscopy can be used to study the conformations of isolated biomolecules, enabling the examination of intrinsic conformational preferences in the absence of solvent effects or other inter-molecular forces.[1] At cryogenic temperatures, single conformations of isolated biomolecules can be examined. Timothy S. Zwier, Scott A. McLuckey, and colleagues have been developing a novel, tandem mass spectrometry based instrument,[2] with an octopole cold ion trap. This instrument is enabling Nicole L. Burke, Andrew F. DeBlase, and Timothy S. Zwier to study the gas phase, cryogenic temperature conformations of antiparallel β-hairpins from our laboratory.

A4.2. Synthesis and Purification of Peptides

The primary sequences of the three peptides in this study are depicted in **Figure A4.1.** Juan F. Espinosa and Samuel H. Gellman designed the peptides.[3] Peptides **A4.A** and **A4.B** were reported to have antiparallel β-hairpin structures on NMR studies in acidic aqueous solution (1.6 mM peptide, 100 mM NaOAc-d_3, 9:1 H_2O:D_2O, pH 3.8) at 3 °C, while the structure of peptide **A4.C** was interpreted to be a random coil.[3] The macrocyclic peptide **A4.A** was designed to be a more stable β-hairpin than peptide **A4.B**, as strand fraying is disallowed by ^{D}Pro-Gly linkers at both ends of **A4.A**, whereas peptide **A4.B** is constrained at only one end by a β-hairpin-promoting ^{D}Pro-Gly linker.

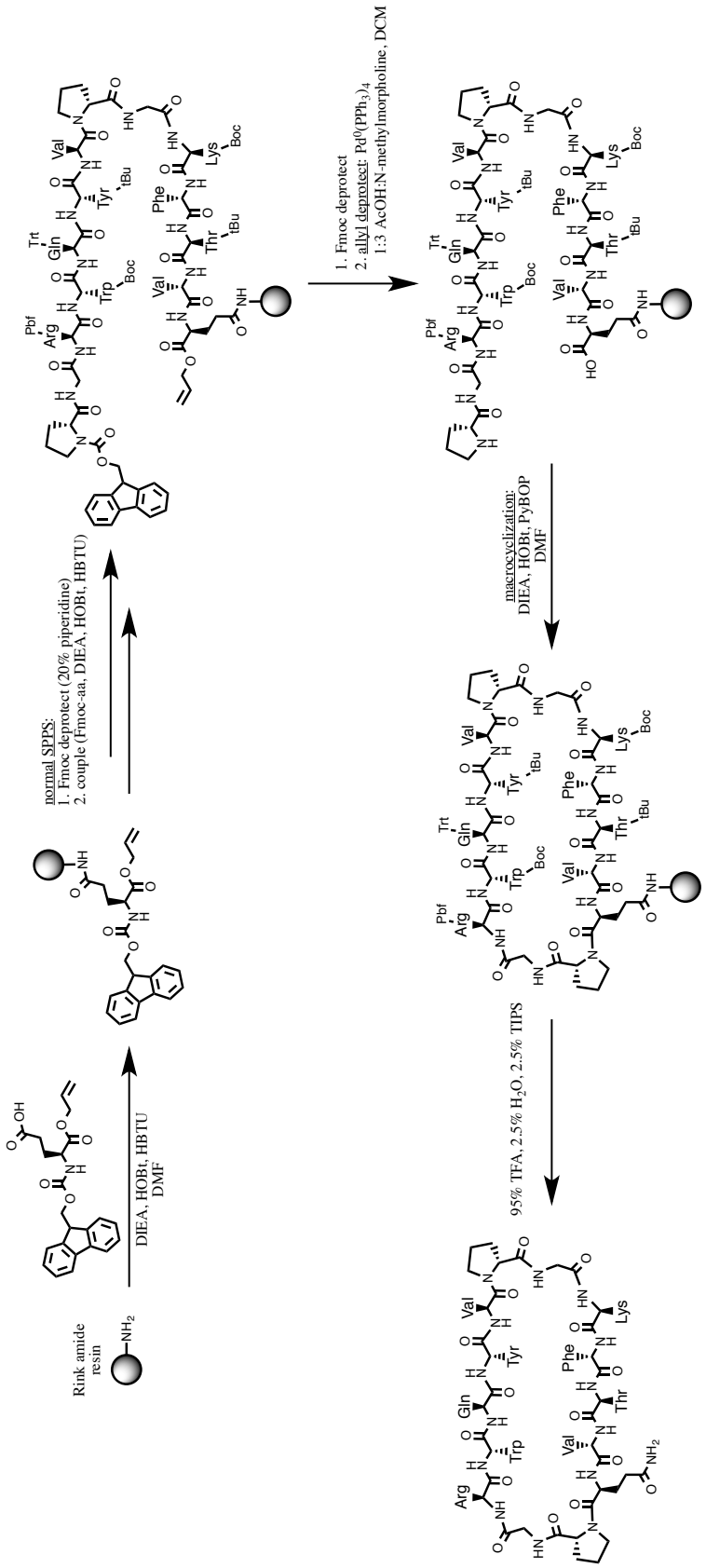

Figure A4.2. Synthetic route to peptide **A4.A**.

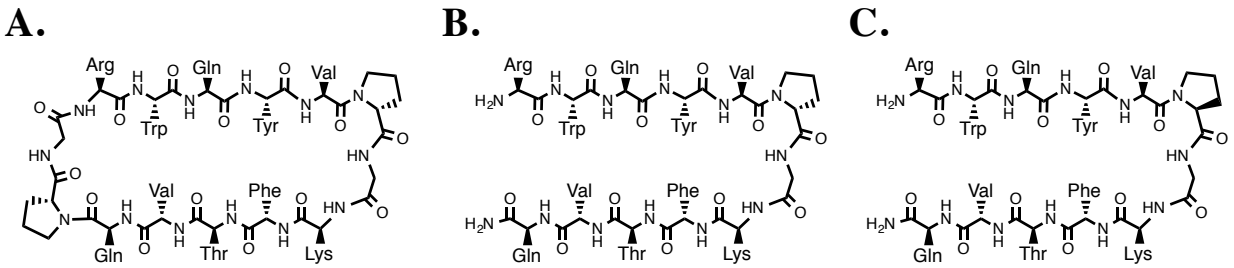

Figure A4.1. Peptides (A) **A4.A**, (B) **A4.B**, and (C) **A4.C**, previously reported to have (A) macrocyclic antiparallel β-hairpin, (B) antiparallel β-hairpin, and (C) random coil secondary structures.[3]

We synthesized the peptides via standard solid phase (SPPS) methods. The slightly more unusual synthetic route to an antiparallel macrocycle (peptide **A4.A**) is depicted in **Figure A4.2**. Peptides were purified by reversed-phase high-performance liquid chromatography (HPLC), using a C18 preparatory column, and a mobile phase of CH_3CN:H_2O:CF_3COOH. Peptides were determined to be to >99% pure by reversed-phase ultra performance liquid chromatography (UPLC), as shown in **Figures A4.3**. Peptide masses were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), as shown in **Figures A4.4**. MALDI-TOF MS showed the following m/z values for the $[M+H^+]^+$ species: peptide **A4.A** calculated = 1644.87, observed = 1645.4; peptide **A4.B** calculated = 1507.82, observed = 1508.2; peptide **A4.C** calculated = 1507.82, observed = 1508.1. No peptide degradation was observable by UPLC or MALDI-TOF MS, when peptides were stored dry at -20 °C, for >3 years.

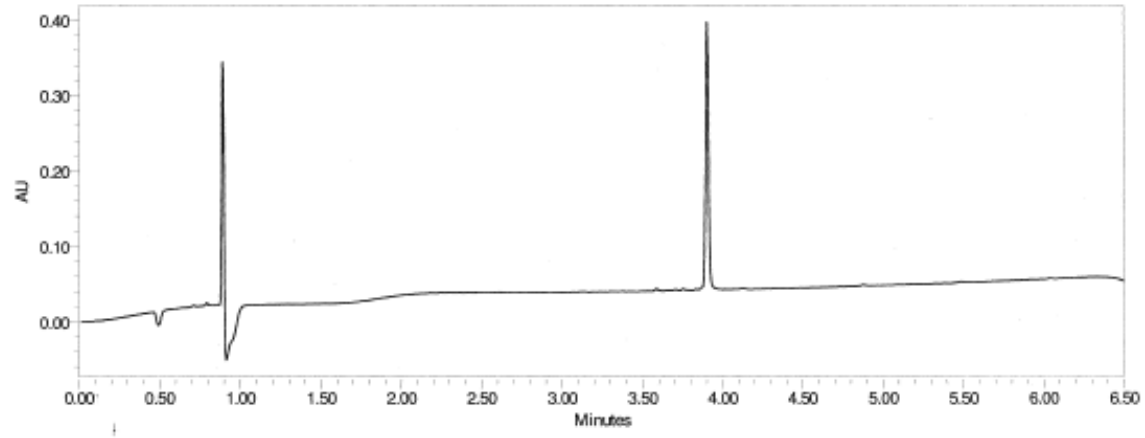

Figure A4.3A. UPLC of peptide **A4.A**.

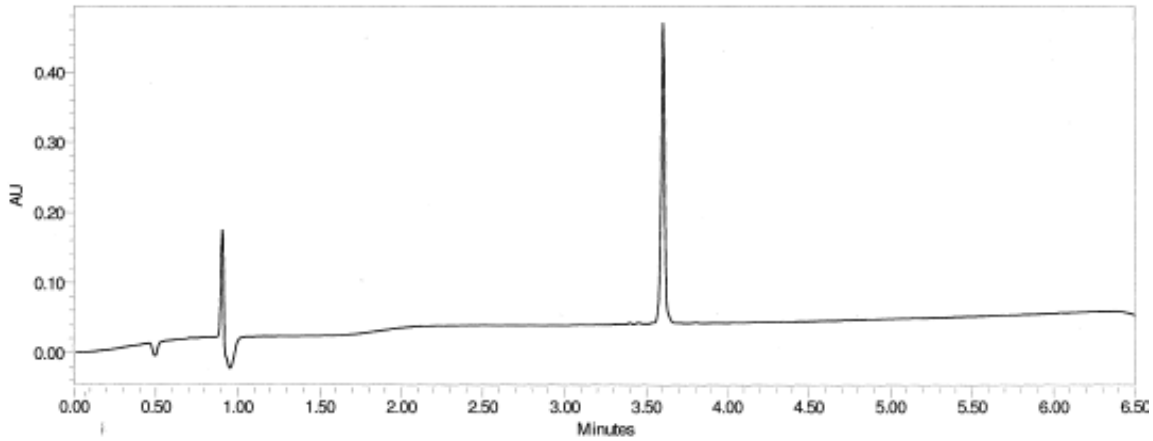

Figure A4.3B. UPLC of peptide **A4.B**.

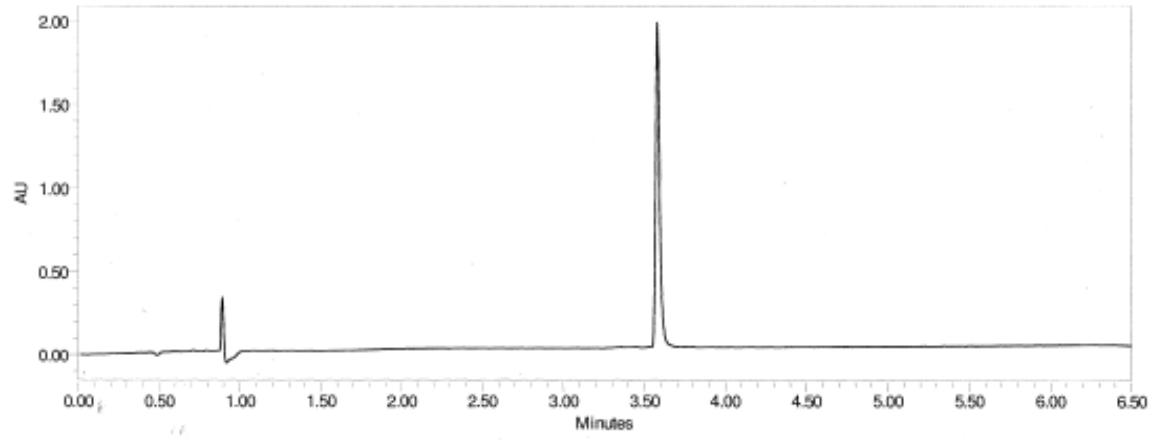

Figure A4.3C. UPLC of peptide **A4.C**.

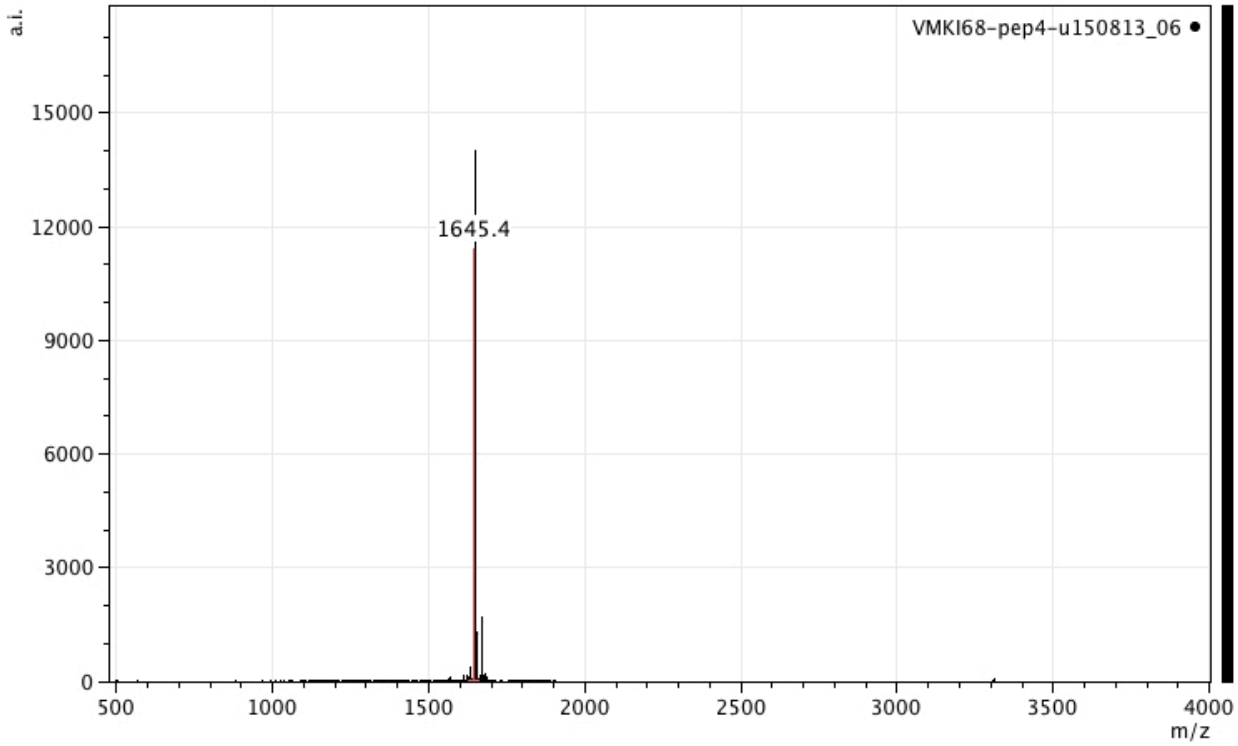

Figure A4.4A. MALDI-TOF MS of peptide **A4.A**.

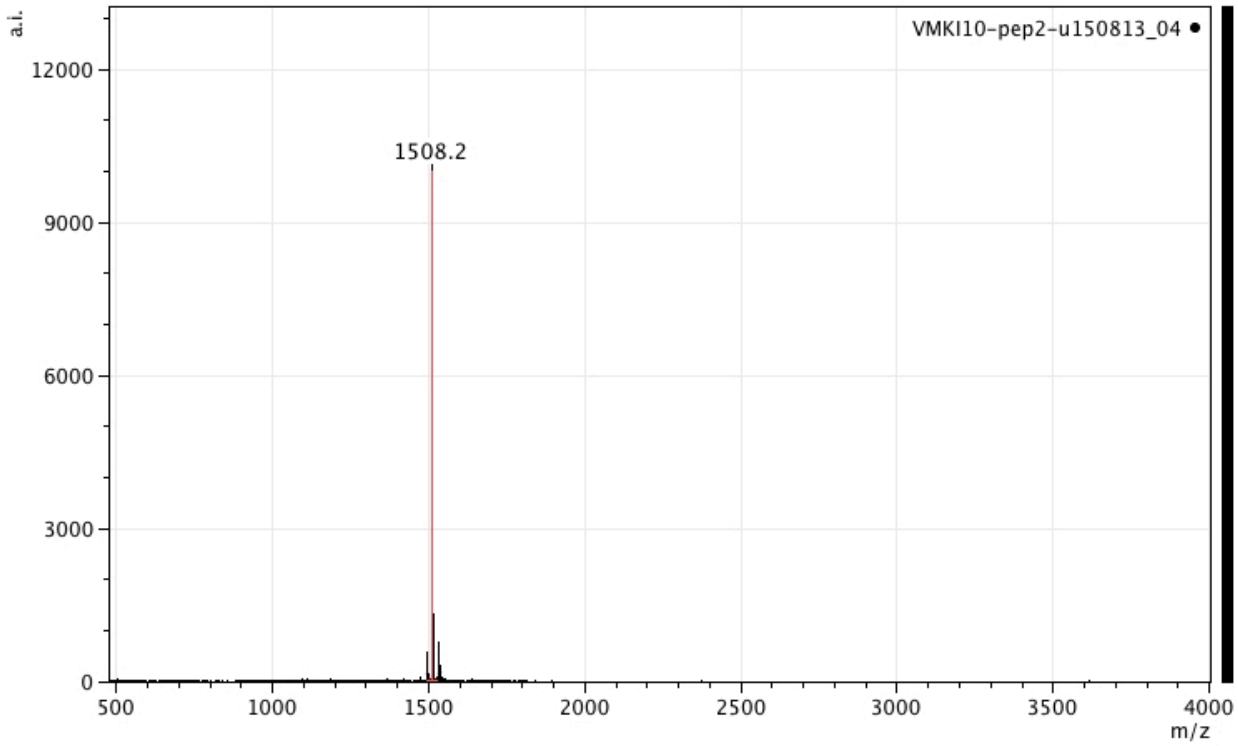

Figure A4.4B. MALDI-TOF MS of peptide **A4.B**.

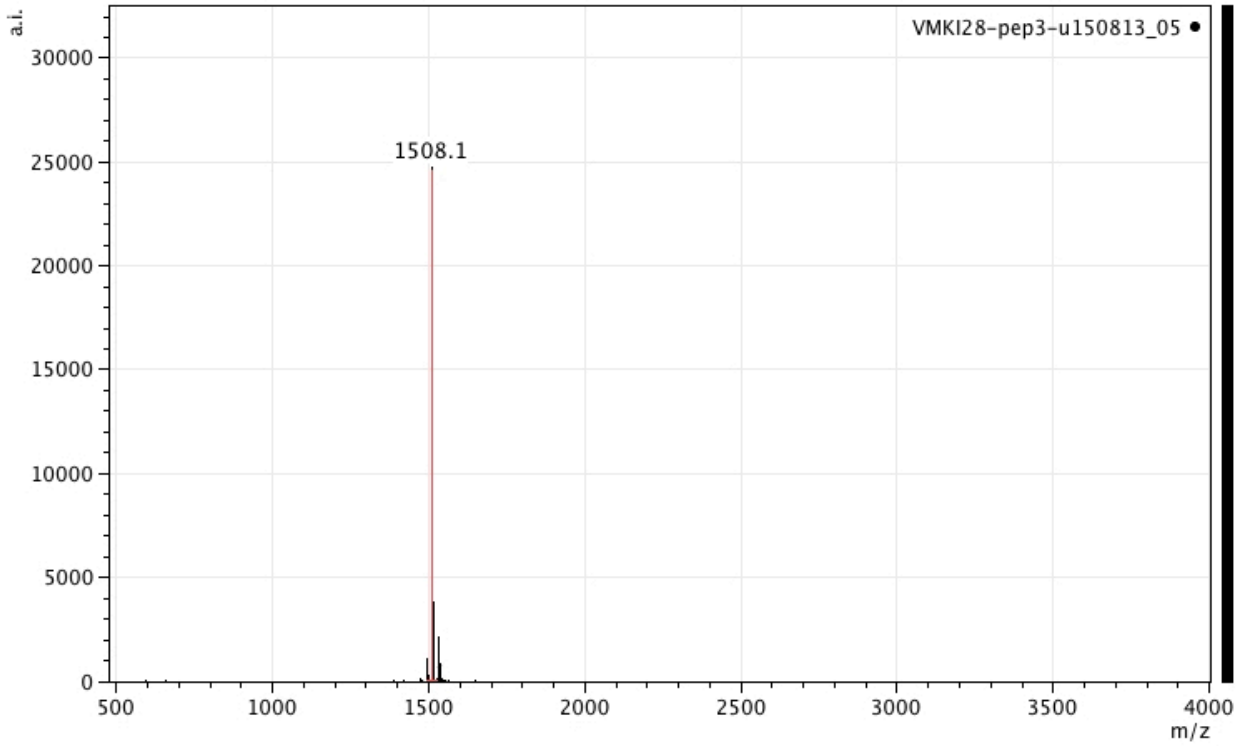

Figure A4.4C. MALDI-TOF MS of peptide **A4.C**.

A4.3. Preliminary Data

DeBlase obtained preliminary UV photofragmentation mass spectra of the antiparallel β-hairpin peptides, using the novel spectrometer that is being developed. Many fragments were observed in the spectrum of peptide **A4.B**, suggesting that the $[M+2H^+]^+$ species is fragile (**Figure A4.6**). In contrast, the spectrum of peptide **A4.A** was much cleaner, with some small molecule (e.g., H_2O) losses (**Figure A4.5**), suggesting that more energy is required to fragment the macrocyclic β-hairpin than the "linear" β-hairpin. Zwier and coworkers are currently redesigning the laser system of their spectrometer,[4] in an effort to boost photofragment yields, such that molecules like peptide **A4.A** can be studied. Once the spectrometer has been optimized, it will be used to characterize individual peptide conformations by infrared (IR) spectroscopy.

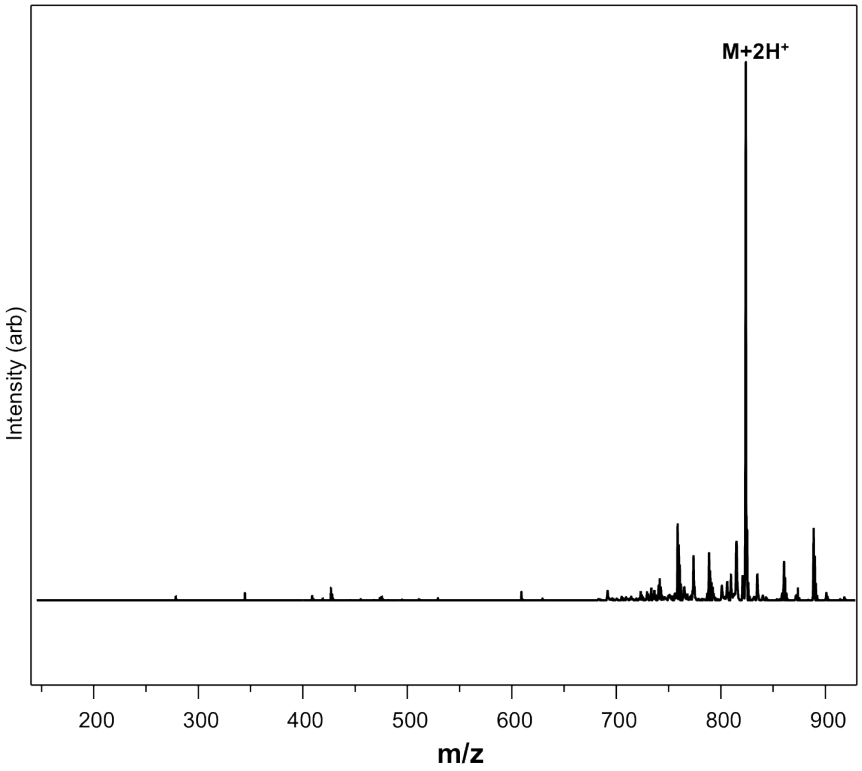

Figure A4.5. UV photofragmentation mass spectrum of peptide **A4.A**.

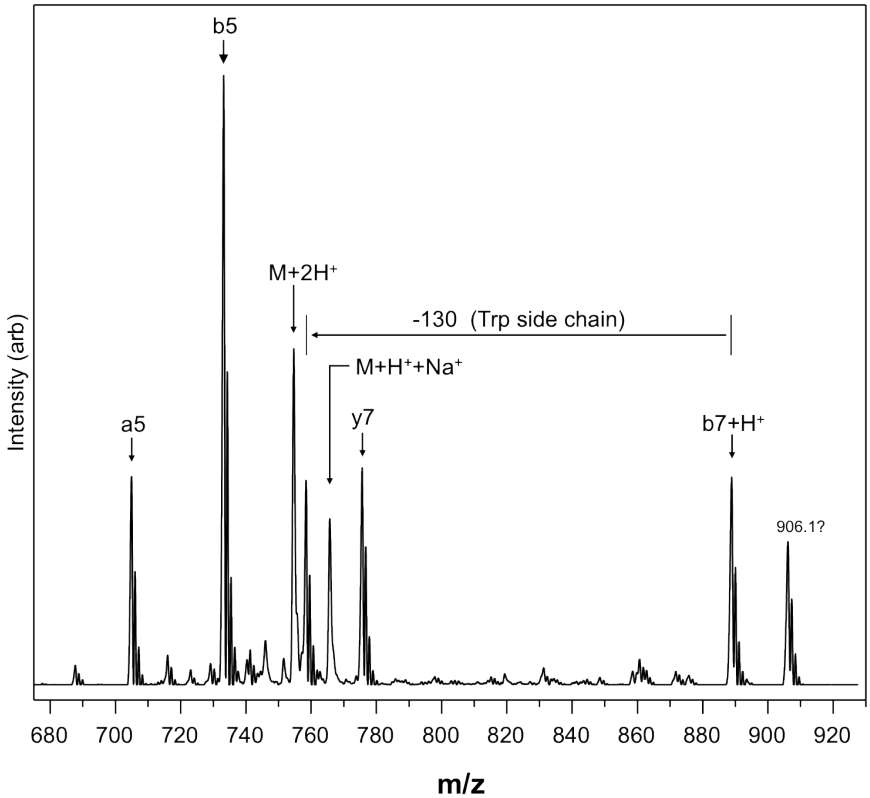

Figure A4.6. UV photofragmentation mass spectrum of peptide **A4.B**.

A4.4. References

(1) de Vries, M. S.; Hobza, P. *Annu Rev Phys Chem* **2007**, *58*, 585.

(2) Redwine, J. G.; Davis, Z. A.; Burke, N. L.; Oglesbee, R. A.; McLuckey, S. A.; Zwier, T. S. *Int J Mass Spectrom* **2013**, *348*, 9.

(3) Espinosa, J. F.; Gellman, S. H. *Angew Chem Int Ed* **2000**, *39*, 2330.

(4) Guidi, M.; Lorenz, U. J.; Papadopoulos, G.; Boyarkin, O. V.; Rizzo, T. R. *J Phys Chem A* **2009**, *113*, 797.

Appendix 5. Succinyl-Glycyl Linker Orientation Guidelines

A5.1. Succinyl-Glycyl Linker Orientation Guidelines

Through studies of macrocyclic peptides intended to fold into two-stranded parallel β-sheets, Aaron M. Almeida and Samuel H. Gellman concluded that the orientation of a designed succinyl-glycyl (succinyl-Gly) linker affects whether or not the macrocycle has parallel β-sheet or random coil structure.[1] They reported that if the succinyl-Gly linker is oriented such that the intended hydrogen-bond network would include a hydrogen-bond to the succinyl residue, then β-sheet structure is promoted (**Figure A5.1A**); whereas if the intended hydrogen-bond network would include a hydrogen-bond to the Gly residue, then β-sheet structure is not promoted (**Figure A5.1B**).

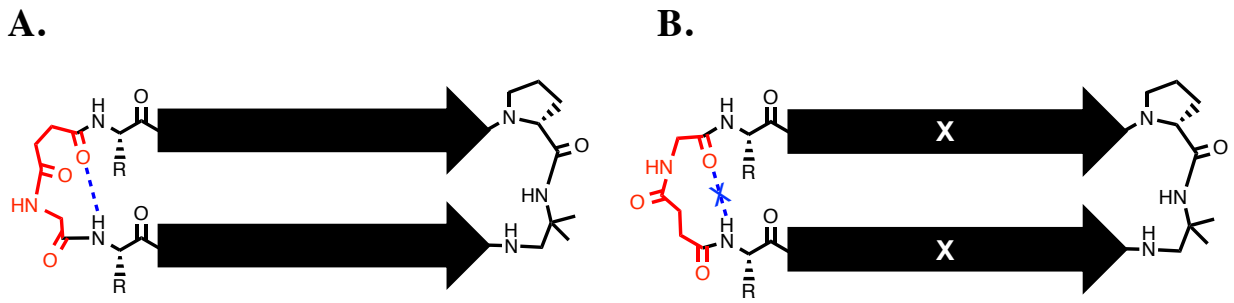

Figure A5.1. Orientation of succinyl-Gly linker (red) determines whether the (A) succinyl residue or (B) Gly residue would be a hydrogen bond (blue) acceptor in the hydrogen bond network of an intended parallel β-sheet. Almeida and Gellman reported observing parallel β-sheet structure with orientation (A), but not orientation (B).

A dependence of parallel β-sheet structure on the orientation of the succinyl-Gly linker could originate from a variety of factors. For instance, here we speculate that the succinyl residue in a hydrogen-bonded position (**Figure A5.1A**) might be able to engage in a three-centered interaction (bifurcated hydrogen bond),[2] as depicted in **Figure A5.2**. Such an interaction would decrease the flexibility of the succinyl-Gly linker, which would decrease β-strand fraying at the N-termini of the peptide.

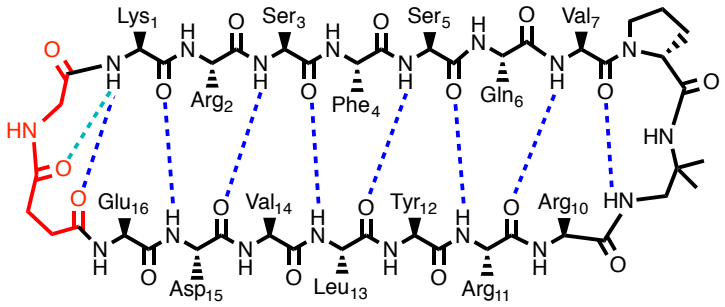

Figure A5.2. Design of peptide **A5.A**, with succinyl-Gly linker shown in red, intended hydrogen-bonds shown in blue, and hypothesized three-centered interaction shown in light blue.

A5.2. Peptide Synthesis

Almeida and Gellman designed peptide **A5.A**, which has the succinyl-Gly linker in the orientation hypothesized to promote parallel β-sheet structure, and asked us to synthesize and characterize this peptide for their study. We synthesized peptide **A5.A** via the route depicted in **Figure A5.3**. Peptide identity was confirmed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (**Figure A5.4**), with the following m/z value for the $[M+H^+]^+$ species: peptide **A5.A** calculated = 2090.14, observed = 2090.5. Purity was determined by analytical HPLC to be >95%.

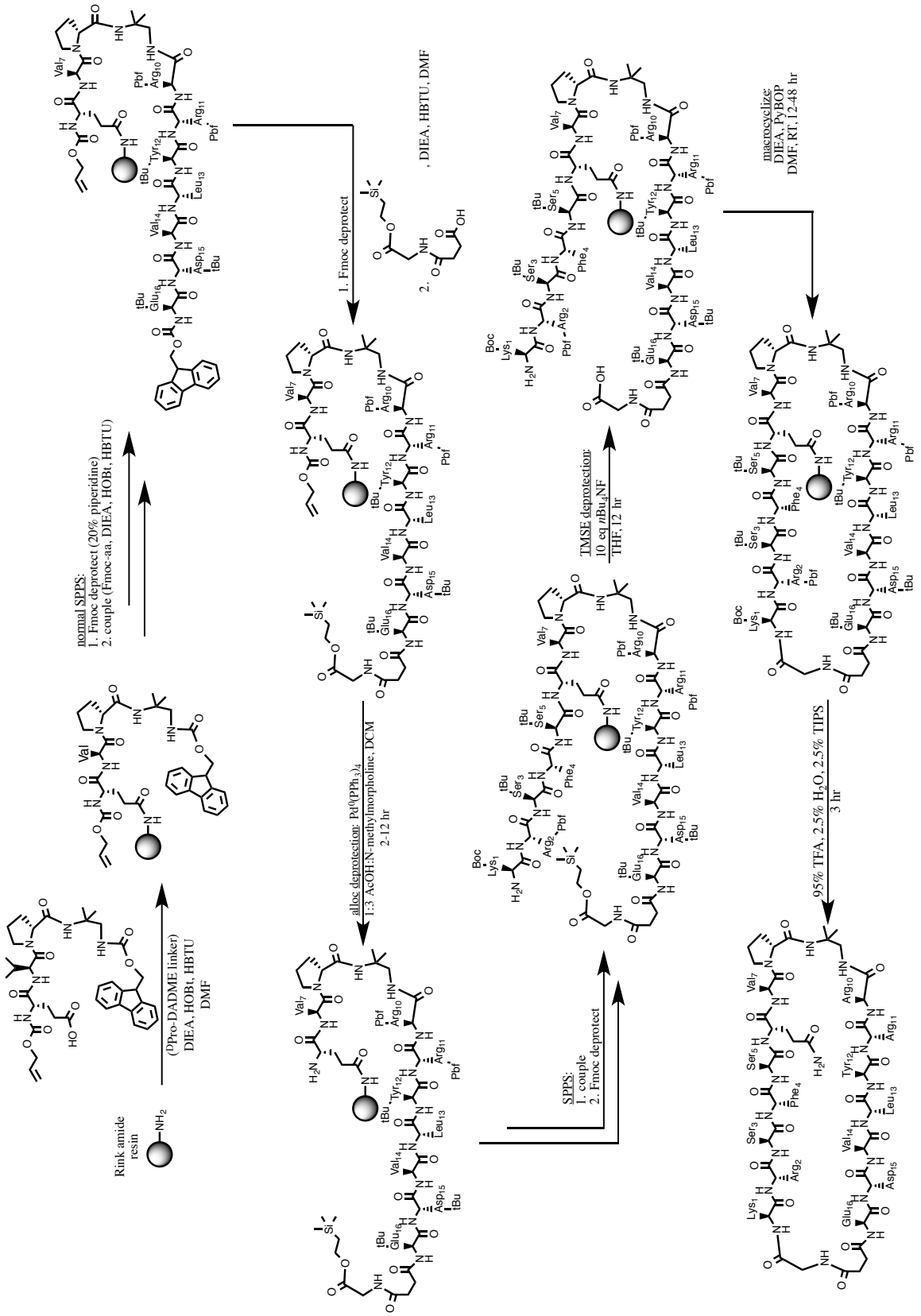

Figure A5.3. Synthetic route to peptide **A5.A**.

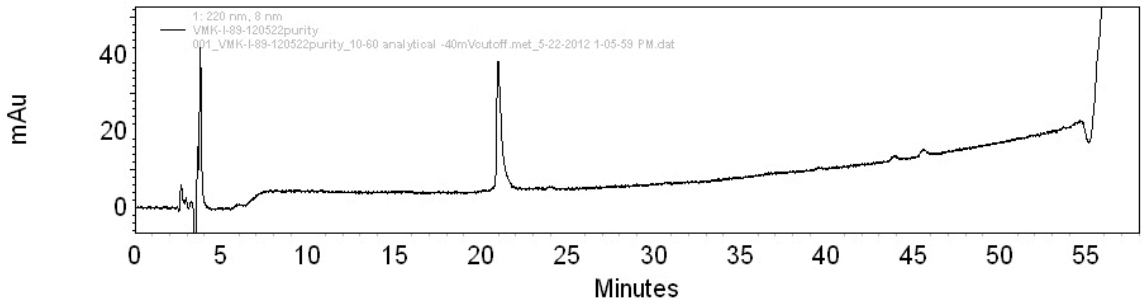

Figure A5.4. Analytical HPLC of peptide **A5.A**.

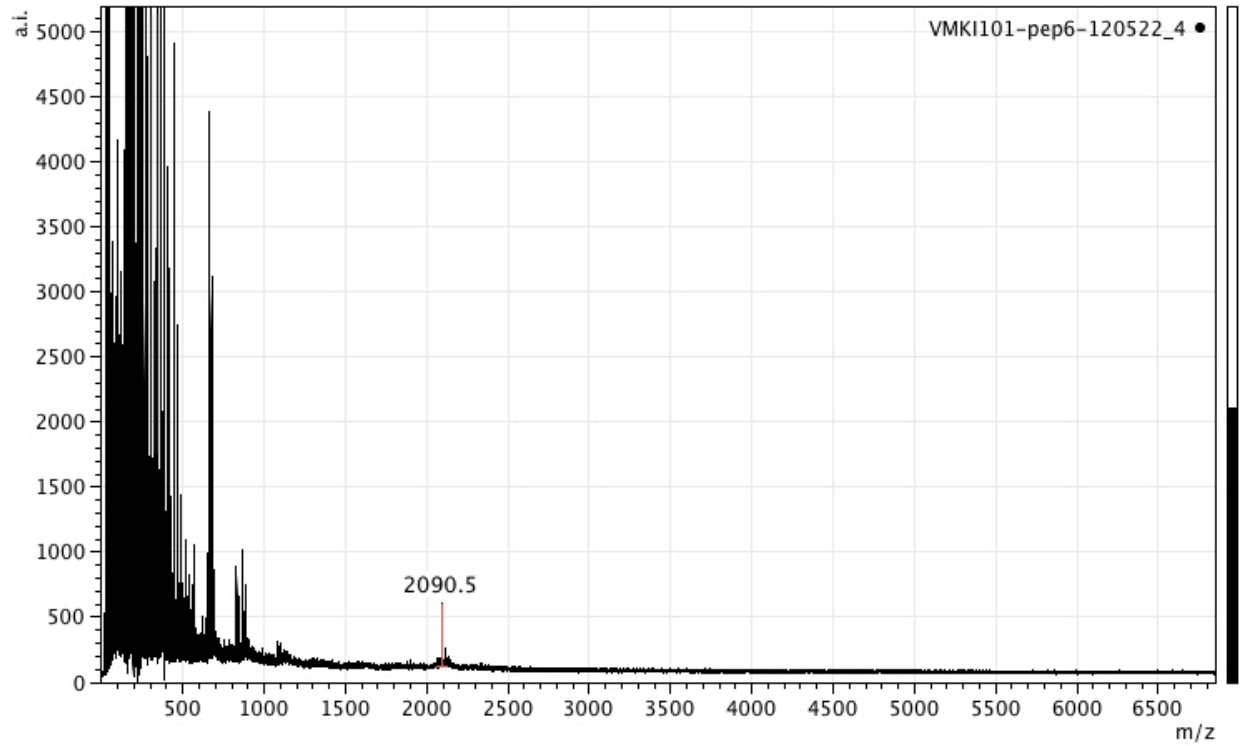

Figure A5.5. MALDI-TOF MS of peptide **A5.A**.

A5.3. Assessing Peptide Folding by NMR

We performed 2D COSY, TOCSY, and ROESY experiments on peptide **A5.A**, using the conditions and methods that Almeida and Gellman reported using for other NMR experiments in this study.[1] We observed few inter-residue ROEs in peptide **A5.A**, suggesting a lack of β-sheet structure. This finding would not be consistent with the trend reported by Almeida and Gellman.

A5.4. References

(1) Almeida, A. M., Doctoral Disseration Thesis, University of Wisconsin - Madison, 2012.

(2) Rozas, I.; Alkorta, I.; Elguero, J. *J Phys Chem A* **1998**, *102*, 9925.

Appendix 6. Peptides Intended to Inhibit Fibronectin Matrix Assembly

A6.1. Inhibition of Fibronectin Matrix Assembly

Fibronectin (FN) is an extracellular glycoprotein that mediates various cellular interactions with the extracellular matrix (ECM).[1] FN matrix assembly is the process whereby fibronectin fibrils are deposited into the ECM.[2] One function of fibronectin is in wound healing: in blood clots, fibronectin binds to fibrin, an interaction mediated in part by ^{4}FNI-^{5}FNI (a pair of modules of fibronectin).[3] The ^{4}FNI-^{5}FNI pair is also thought to mediate interactions between fibronectin and FUD, a domain of a bacterial protein (*Streptococcus pyogenes* F1 adhesin) that can prevent FN matrix assembly (**Figure A6.1**).[4,5] Sequence alignments of FUD with the fibronectin-binding region of another bacterial protein, BBK32 (from *Borrelia burgdorferi*), enabled the determination of a segment of Bbk32 that likely also binds ^{4}FNI-^{5}FNI.[6]

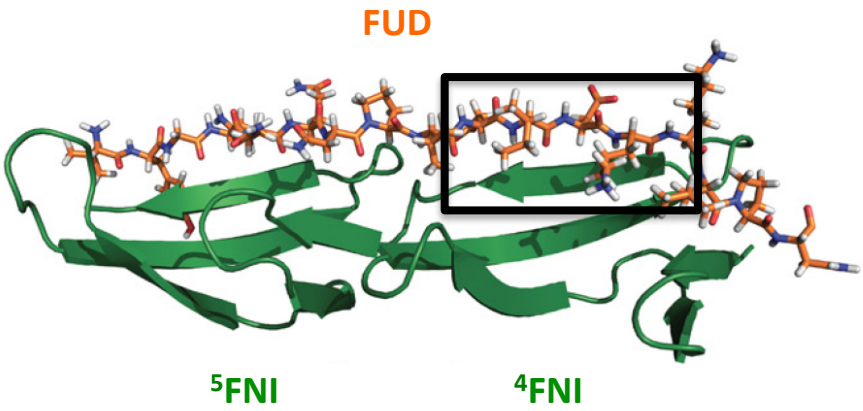

Figure A6.1. Putative structure of the FUD domain of *Streptococcus pyogenes* F1 adhesin (orange) bound to the ^{4}FNI-^{5}FNI module pair of fibronectin (green). Note the antiparallel β-sheet interactions between FUD and ^{4}FNI (box). Figure adapted from Mosher and coworkers,[4] with permission from the American Society for Biochemistry and Molecular Biology (ASBMB).

A6.2. Preliminary Peptide Design

Bianca R. Tomasini-Johansson, of the Hans W. Sollinger Laboratory, and Samuel H. Gellman proposed that we make antiparallel β-hairpin macrocycles intended to

competitively inhibit fibronectin (specifically ^{4}FNI) assembly into matrices. An agent that can inhibit FN matrix assembly might have interesting applications, for example as an anti-fibrosis therapeutic. With Wenjiang "Bill" Ma and Deane F. Mosher, we designed the peptides depicted in **Figure A6.2**. The sequences of the intended β-strands of the peptides were derived from the sequences at the interfaces of FUD-^{4}FNI or Bbk32-^{4}FNI interactions (**Figure A6.1** box). The non-natural linker ^{D}Pro-Gly was employed to promote antiparallel β-sheet conformations.[7-9] The intended β-strands of peptides **A6.A** and **A6.B** are six residues in length, as this is the length of the β-strand of ^{4}FNI that is thought to hydrogen bond with FUD. However, it has been observed that antiparallel β-sheet propensity in cyclic peptides varies periodically with peptide lengths, with lower β-sheet content in intended strands comprised of even numbers of residues.[10] We therefore lengthened the intended β-strands by one residue (we chose Thr, because it promotes β-sheet structure and water-solubility), for the design of peptide **A6.C**. We hypothesized that these peptides, particularly peptide **A6.C**, might be able to bind to ^{4}FNI, in a manner analogous to the binding of FUD or Bbk32, and inhibit FN matrix assembly.

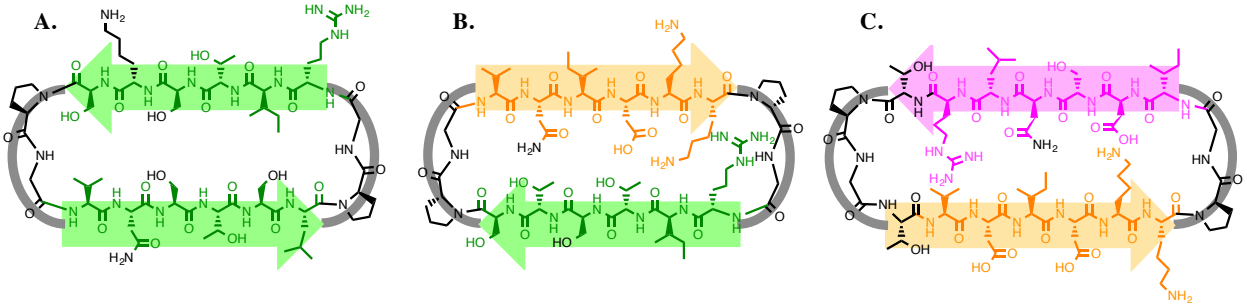

Figure A6.2. Design of potential inhibitors of FN matrix assembly, peptides (A) **A6.A**, (B) **A6.B**, and (C) **A6.C**, with intended β-strands color-coded by the protein structures from which they were derived: ^{4}FNI (green), FUD (orange), and Bbk32 (pink).

A6.3. Synthesis of Peptides

See **Appendix 4** (**Figure A4.2**) for the synthetic route to an antiparallel β-hairpin.

Peptides were determined to be to >95% pure (**Figures A6.3**), except for peptide **A6.B**, which was only 67% pure. Peptide masses were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), which showed the following m/z values for the $[M+H^+]^+$ species: peptide **A6.A** calculated = 1582.86, observed = 1583.8; peptide **A6.B** calculated = 1651.92, observed = 1652.8; peptide **A6.C** calculated = 1908.02, observed = 1908.9 (**Figures A6.4**).

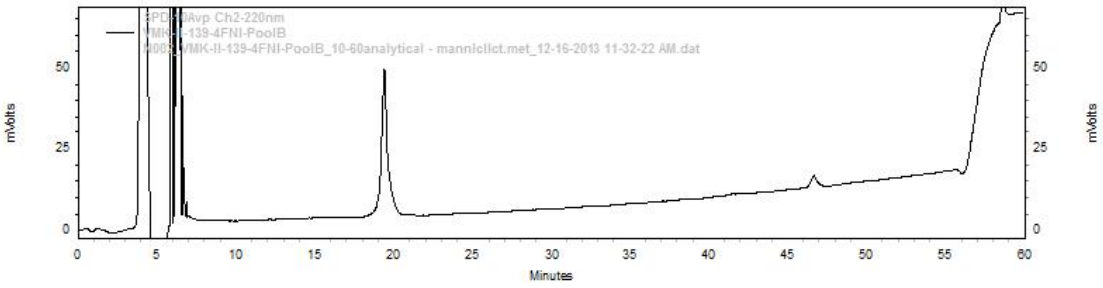

Figure A6.3A. Analytical HPLC of peptide **A6.A**.

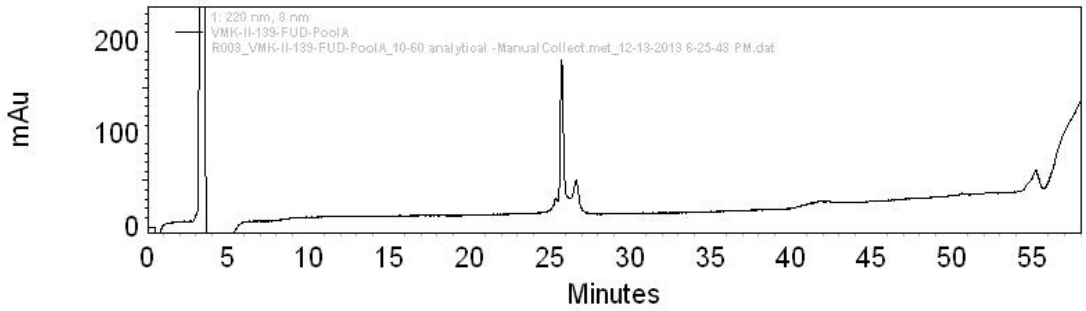

Figure A6.3B. Analytical HPLC of peptide **A6.B**.

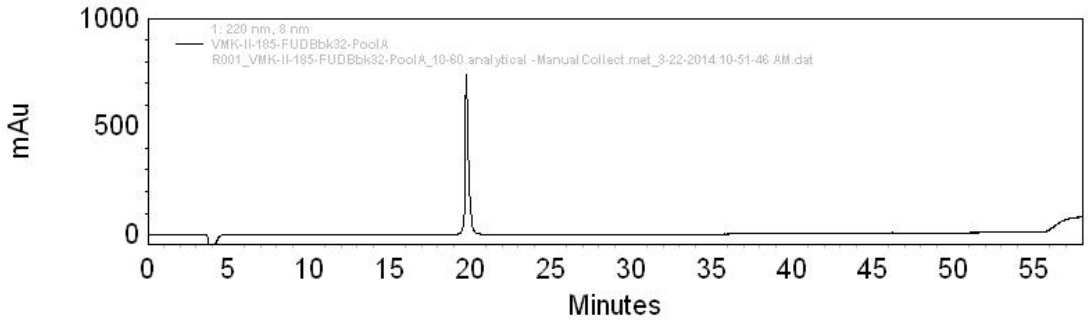

Figure A6.3C. Analytical HPLC of peptide **A6.C**.

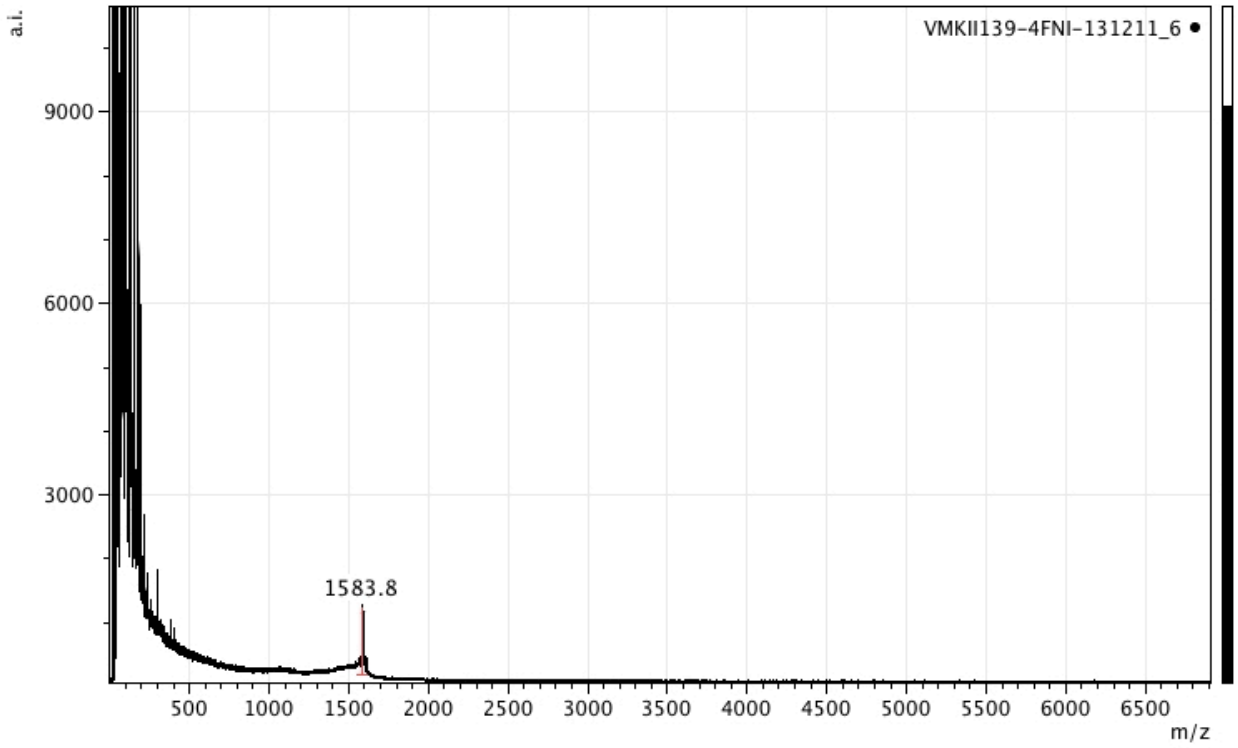

Figure A6.4A. MALDI-TOF MS of peptide **A6.A**.

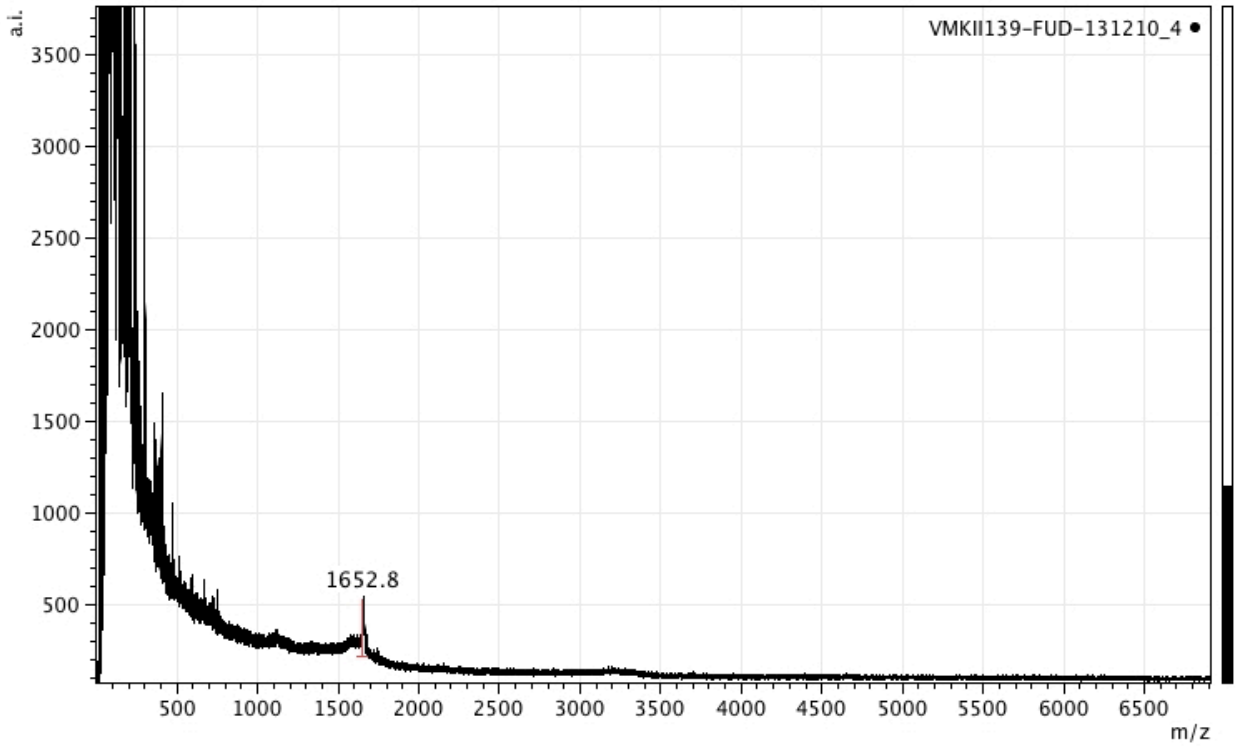

Figure A6.4B. MALDI-TOF MS of peptide **A6.B**.

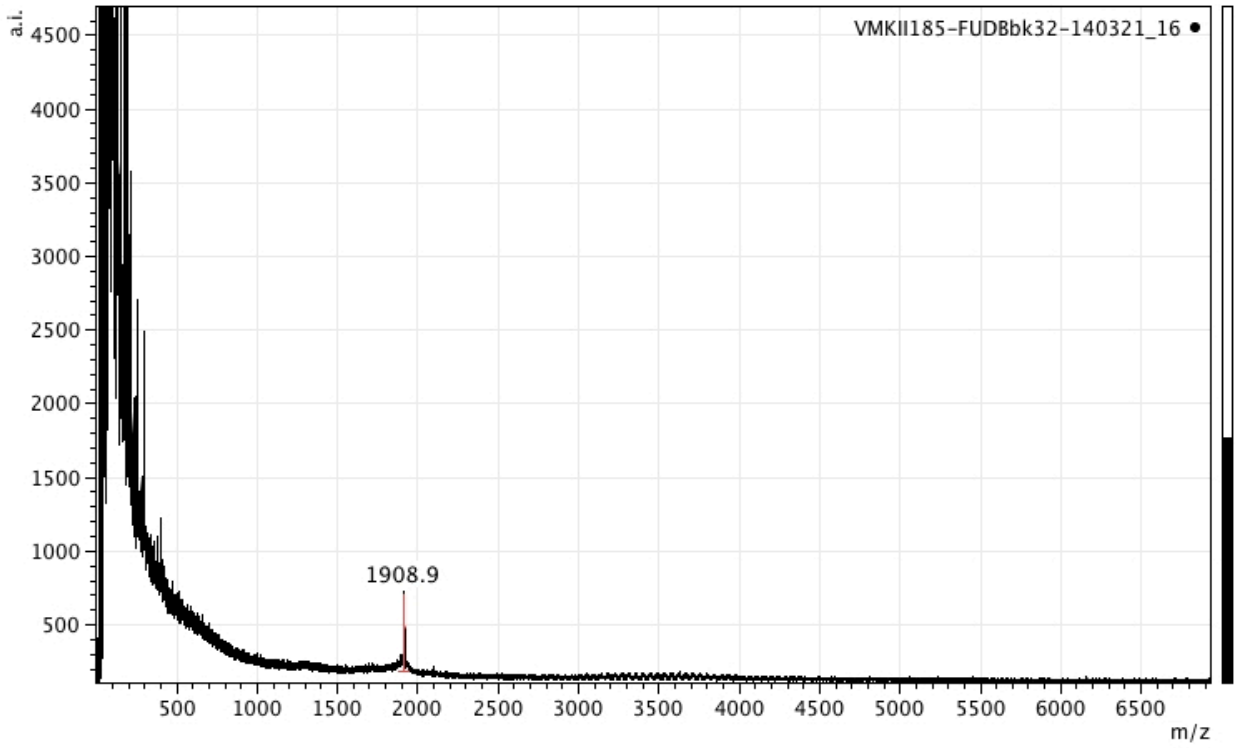

Figure A6.4C. MALDI-TOF MS of peptide **A6.C**.

A6.4. Preliminary Results

Fluorescence polarization (FP) competition assays, assessing the affinity of the peptides for FN, relative to the affinity of Alexa-labeled Bbk32 for FN, were performed by Ma, according to previously reported methods.[11] These assays showed that there was no inhibition of Alexa-Bbk32 binding to FN by peptides **A6.B** or **A6.C**, and little to no inhibition of this binding by peptide **A6.A** (**Figure A6.5**). Assays of Alexa-labeled fibronectin binding to fibroblast (AH1F) cell surfaces were performed by Tomasini-Johansson, according to previously reported methods.[12] These assays showed that peptides **A6.A**, **A6.B**, and **A6.C** did not inhibit Alexa-FN binding to the cell surfaces (**Figure A6.6**). In summary, the *in vitro* assays show that the designed peptides fail to bind fibronectin and fail to inhibit fibronectin matrix assembly, at useful levels. A next step in this work could be to perform structural studies of the designed peptides, to determine the extent to which they adopt the

intended antiparallel β-sheet structures. If necessary, the peptides could then be redesigned to incorporate more residues or residue pairs that are known to promote antiparallel β-sheet structure.[13,14] However, such a redesign could decrease the sequence similarity of the peptides to the bacterial or ^{4}FNI sequences that they mimic, and consequently might further decrease the affinities of the peptides for ^{4}FNI.

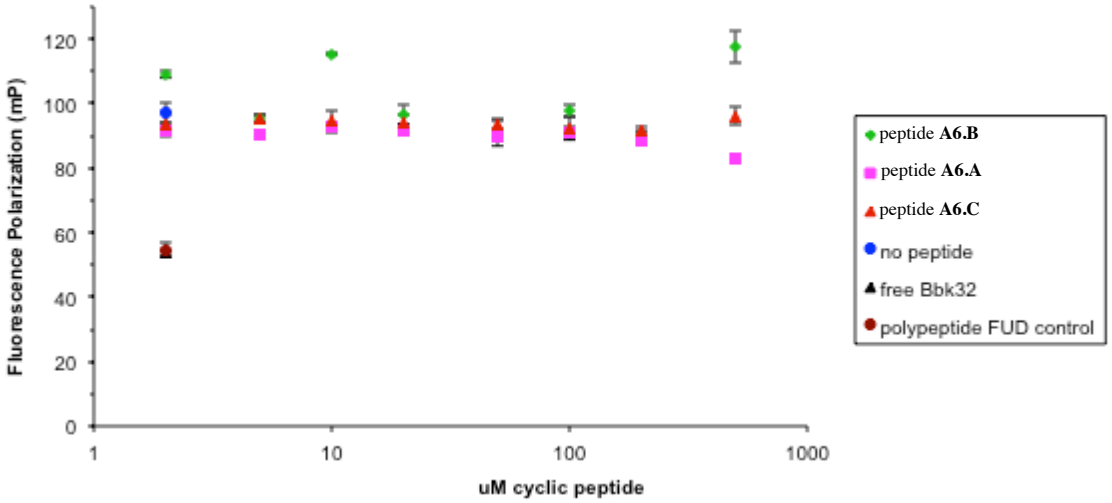

Figure A6.5. Competition fluorescence polarization (FP) assay, for Alexa-Bbk32 binding to FN. Peptides **A6.A** (pink square), **A6.B** (green diamond), and **A6.C** (red triangle), like the "no peptide" negative control (blue circle), show little to no ability to compete off Alexa-Bbk32. Positive controls, FUD (brown circle) and Bbk32 (black triangle), are also shown. N = 3. Figure by Ma.

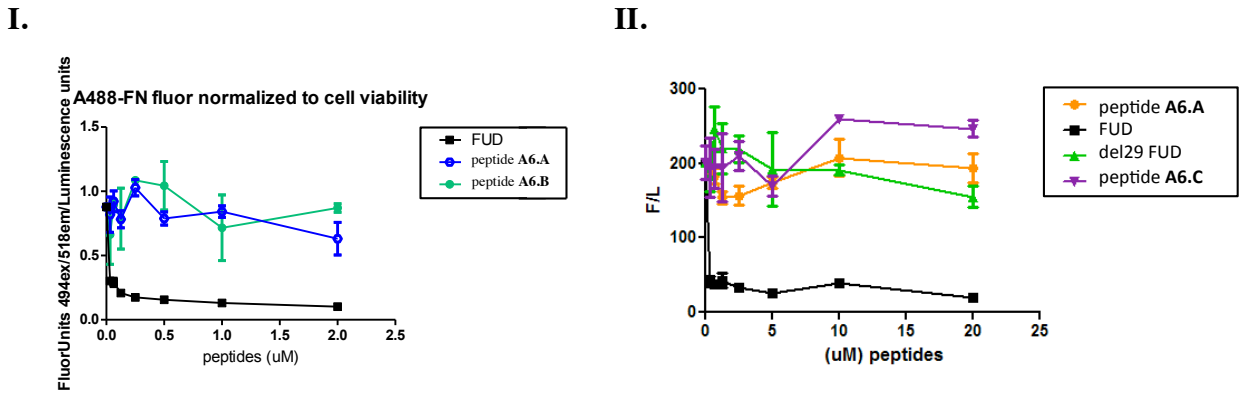

Figure A6.6. Assay of inhibition of Alexa-FN binding to fibroblast (AH1F) cell surfaces. Peptides **A6.A** (I. blue circle, II. orange circle), **A6.B** (I. green circle), and **A6.C** (II. purple triangle), like the negative control FUD mutant (II. green triangle), did not inhibit Alexa-FN binding to the cell surfaces. A positive control, FUD, is also shown. Figure by Tomasini-Johansson.

A6.5. References

(1) Pankov, R.; Yamada, K. M. *J Cell Sci* **2002**, *115*, 3861.

(2) Singh, P.; Carraher, C.; Schwarzbauer, J. E. *Annu Rev Cell Dev Biol* **2010**, *26*, 397.

(3) Potts, J. R.; Campbell, I. D. *Matrix Biol* **1996**, *15*, 313.

(4) Maurer, L. M.; Tomasini-Johansson, B. R.; Ma, W.; Annis, D. S.; Eickstaedt, N. L.; Ensenberger, M. G.; Satyshur, K. A.; Mosher, D. F. *J Biol Chem* **2010**, *285*, 41087.

(5) Bingham, R. J.; Rudino-Pinera, E.; Meenan, N. A.; Schwarz-Linek, U.; Turkenburg, J. P.; Hook, M.; Garman, E. F.; Potts, J. R. *Proc Natl Acad Sci USA* **2008**, *105*, 12254.

(6) Harris, G.; Ma, W.; Maurer, L. M.; Potts, J. R.; Mosher, D. F. *J Biol Chem* **2014**, *289*, 22490.

(7) Haque, T. S.; Little, J. C.; Gellman, S. H. *J Am Chem Soc* **1996**, *118*, 6975.

(8) Stanger, H. E.; Gellman, S. H. *J Am Chem Soc* **1998**, *120*, 4236.

(9) Karle, I. L.; Awasthi, S. K.; Balaram, P. *Proc Natl Acad Sci USA* **1996**, *93*, 8189.

(10) Gibbs, A. C.; Kondejewski, L. H.; Gronwald, W.; Nip, A. M.; Hodges, R. S.; Sykes, B. D.; Wishart, D. S. *Nat Struct Biol* **1998**, *5*, 284.

(11) Ma, W.; Ma, H.; Fogerty, F. J.; Mosher, D. F. *J Biol Chem* **2015**, *290*, 4866.

(12) Tomasini-Johansson, B. R.; Johnson, I. A.; Hoffmann, F. M.; Mosher, D. F. *Matrix Biol* **2012**, *31*, 360.

(13) Wouters, M. A.; Curmi, P. M. G. *Proteins* **1995**, *22*, 119.

(14) Hutchinson, E. G.; Sessions, R. B.; Thornton, J. M.; Woolfson, D. N. *Protein Sci* **1998**, *7*, 2287.