

**Design of Unnatural Oligomers with Protein-like Tertiary Structure**

by

**Zachary Eric Reinert**

B.S. Biochemistry, University of Maryland Baltimore County, 2010

Submitted to the Graduate Faculty of  
the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH  
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Zachary Eric Reinert

It was defended on

November 12, 2015

and approved by

Dr. Paul Floreancig, Professor, Department of Chemistry

Dr. Stephen Weber, Professor, Department of Chemistry

Dr. Ronald Wetzell, Professor, Department of Structural Biology

Committee Chair: Dr. W. Seth Horne, Associate Professor, Department of Chemistry

Copyright © by Zachary Reinert

2015

# DESIGN OF UNNATURAL OLIGOMERS WITH PROTEIN-LIKE TERTIARY STRUCTURE

Zachary Reinert, PhD

University of Pittsburgh, 2015

Proteins play key roles in biological processes that are highly dependent on their three-dimensional fold. Given the therapeutic relevance of many proteins, significant research effort has pursued the development of unnatural oligomers with protein-like folds. However, as the complexity of the target fold pushes beyond secondary structure, the difficulty of recapitulating the native protein fold becomes considerably high. There remains an unmet need for methods that enable tertiary structure mimicry of proteins by unnatural oligomers. Accessing complex protein-like folds on protease-resistant backbones would yield improved therapeutics with high target specificity and sustained biological effects *in vivo*.

The goal of the current work was to generate design strategies for tertiary protein structure mimicry. We selected GB1, a protein that adopts a compact tertiary fold, as a system for backbone modification. We systematically replaced residues in the secondary structures of GB1 and measured the resulting changes to folding thermodynamics and structure. Combination of separate modifications into one protein led to a mutant that showed evidence for tertiary folding despite having an ~ 20% unnatural backbone sequence. Furthermore, grafting the combined backbone alterations onto a side-chain sequence that encodes for a more stable and identical tertiary fold resulted in a significant stabilization of the folded state. The observations supported a general design hypothesis that proteins have two mutually orthogonal design sequences: 1) backbone and 2) side-chain.

The detailed effects of unnatural residues on protein folding thermodynamics were also examined and revealed several interesting trends. A series of  $\alpha \rightarrow \beta^3$  substitutions were implemented in GB1. As  $\beta^3$ -residues have more flexible backbones than  $\alpha$ -residues, we investigated rigidification of the backbone through either cyclization or  $C_\alpha$ -methylation, which somewhat or significantly stabilized the folded state of the protein.

We studied other types of unnatural residues in the context of a small hairpin peptide.  $\gamma^{\text{cyc}}$ -Residues were found to stabilize the hairpin secondary structure greater than the natural backbone. However when applied to GB1, the same strategy was detrimental to folding. Optimization of the location of unnatural residues resulted in a restoration of near wild-type folded stability.

Overall the developed strategies should be applicable to larger, more complex protein folds.

## TABLE OF CONTENTS

<b>1.0</b>	<b>INTRODUCTION TO PROTEIN STRUCTURE MIMICRY .....</b>	<b>1</b>
<b>1.1</b>	<b>BACKBONE MODIFICATION OF PROTEINS .....</b>	<b>1</b>
<b>1.2</b>	<b>HIGHER ORDER STRUCTURAL MIMICRY.....</b>	<b>14</b>
<b>1.3</b>	<b>GOALS .....</b>	<b>18</b>
<b>2.0</b>	<b>DESIGN OF FOLDAMERS WITH PROTEIN-LIKE FOLDING BEHAVIOR</b>	<b>20</b>
<b>2.1</b>	<b>MODEL SYSTEM GB1 BACKGROUND.....</b>	<b>21</b>
<b>2.2</b>	<b>SECONDARY STRUCTURE BACKBONE MODIFICATIONS .....</b>	<b>22</b>
<b>2.2.1</b>	<b>Loops.....</b>	<b>22</b>
<b>2.2.2</b>	<b>Helix .....</b>	<b>36</b>
<b>2.2.3</b>	<b>Sheet .....</b>	<b>41</b>
<b>2.2.4</b>	<b>Turns.....</b>	<b>48</b>
<b>2.3</b>	<b>COMBINED BACKBONE MUTATIONS .....</b>	<b>54</b>
<b>2.3.1</b>	<b>“All” Mutant .....</b>	<b>54</b>
<b>2.3.2</b>	<b>“NuG-All” Mutant.....</b>	<b>56</b>
<b>2.4</b>	<b>CONCLUSIONS AND OUTLOOK.....</b>	<b>59</b>
<b>2.5</b>	<b>EXPERIMENTAL.....</b>	<b>61</b>
<b>2.5.1</b>	<b>General Information.....</b>	<b>61</b>
<b>2.5.2</b>	<b>Protein Synthesis.....</b>	<b>61</b>

2.5.2.1	General Considerations for Fmoc-SPPS of GB1.....	62
2.5.2.2	Optimized Fmoc-SPPS Methods.....	65
2.5.3	Protein Purification and Characterization.....	67
2.5.4	Competition ELISA.....	70
2.5.5	Circular Dichroism Spectroscopy.....	71
2.5.6	Gel Permeation Chromatography.....	72
2.5.7	Molecular Dynamics.....	72
2.5.8	Crystallization of Proteins .....	74
3.0	FOLDING THERMODYNAMICS OF BACKBONE-MODIFIED PROTEINS	77
3.1	HELIX SUBSTITUTION STRATEGIES.....	78
3.1.1	$\alpha \rightarrow \beta^3$ Substitution.....	78
3.1.2	$\beta^3 \rightarrow \beta^2$ Substitution .....	86
3.1.3	$\beta^3 \rightarrow \beta^{\text{cyc}}$ Substitution .....	92
3.1.4	$\beta^3 \rightarrow C_\alpha$ Methyl Substitution .....	96
3.2	CONCLUSIONS .....	101
3.3	EXPERIMENTAL.....	104
3.3.1	General Information.....	104
3.3.2	Synthesis of Fmoc- $\beta^2$ -Asn(Dmcp)-OH .....	105
3.3.3	Protein Synthesis and Purification.....	109
3.3.4	Circular Dichroism.....	111
3.3.5	Crystallization of Proteins .....	116
4.0	IMPROVEMENTS TO SHEET MIMICRY STRATEGIES .....	119
4.1	N-METHYL MODEL HAIRPIN INVESTIGATIONS.....	120

<b>4.2</b>	<b>SECOND GENERATION GB1 SHEET MUTANTS .....</b>	<b>125</b>
<b>4.3</b>	<b>CONCLUSIONS AND FUTURE DIRECTIONS.....</b>	<b>132</b>
<b>4.4</b>	<b>EXPERIMENTAL.....</b>	<b>133</b>
<b>4.4.1</b>	<b>General Information.....</b>	<b>133</b>
<b>4.4.2</b>	<b>Peptide Synthesis .....</b>	<b>134</b>
<b>4.4.3</b>	<b>Protein Synthesis.....</b>	<b>134</b>
<b>4.4.4</b>	<b>Protein Purification and Characterization.....</b>	<b>134</b>
<b>4.4.5</b>	<b>NMR Sample Preparation and Analysis .....</b>	<b>136</b>
<b>4.4.6</b>	<b>Circular Dichroism.....</b>	<b>137</b>
<b>5.0</b>	<b>OUTLOOK.....</b>	<b>138</b>
	<b>REFERENCES.....</b>	<b>147</b>



## LIST OF TABLES

Table 1. Thermodynamics of Folding for WT GB1 and PEG Mutants as Measured by CD. ....	25
Table 2. Melting Temperatures and Changes to Free Energy of Folding for Helix Mutants 6-7. 39	
Table 3. Comparison of Helical Parameters of $\alpha/\beta$ Helices and Wild-type GB1 Helix. ....	40
Table 4. Summary of Folding Thermodynamics for Wild-type GB1 and Mutants.....	56
Table 5. Summary of Folding Thermodynamics for Wild-type GB1 and All Mutants.....	59
Table 6. MALDI-TOF MS Data for Proteins 1-14, and 16. ....	70
Table 7. Crystallization Conditions for Proteins 5-7 and 13. ....	74
Table 8. X-ray Diffraction Data and Refinement Statistics for Proteins 5-7, and 13. ....	76
Table 9. Thermodynamic Parameters for the Folding Transitions of 1 and 5-7 at 298 K. ....	81
Table 10. Thermodynamic Parameters for the Folding Transitions of 6 and 17-19 at 298 K. ....	91
Table 11. Thermodynamic Parameters for the Folding Transitions of 6 and 20-23 at 298 K. ....	96
Table 12. Thermodynamics of Folding for Proteins 6, 21, 25, and 26 at 298 K. ....	100
Table 13. Proposed Physical Bases for Observed Thermodynamic Changes to Folding Enthalpy and Entropy in Backbone-Modified GB1 Mutants Relative to $\alpha$ -Residues.....	103
Table 14. MALDI-TOF MS Data for Proteins 17-24. ....	111
Table 15. Crystallization Conditions for Proteins 17, 18, 21, 24, and 25.....	116

Table 16. X-ray Diffraction Data and Refinement Statistics for Proteins 17, 18, 21, 24, and 25. .....	118
Table 17. Folding Thermodynamics of Hairpin Peptides 33-37 from NMR Measurements. <sup>a</sup> ...	124
Table 18. Folding Thermodynamics of Hairpin Peptides 33, 38, and 39 from NMR Measurements. <sup>a</sup> .....	126
Table 19. Folding Thermodynamics of Proteins 1, 10, and 40-42 from CD Measurements.....	129
Table 20. Folding Thermodynamics of Proteins 1, 10, and 40-43 from CD Measurements.....	131
Table 21. MALDI-TOF MS Data for Peptides 34-39 and GB1 Mutants 40-43.....	135

## LIST OF FIGURES

Figure 1. Selection of unnatural residues that have been explored as replacements for $\alpha$ -residues (yellow) in folded proteins.....	2
Figure 2. Overview of NCL.....	3
Figure 3. Overview of expression of proteins with unnatural amino acids (UAA) using engineered aminoacyl tRNA synthetases (aa-RS).....	4
Figure 4. Summary of main consequences of amide to ester substitution in proteins.....	5
Figure 5. Close-up of amide in OMTKY3 (yellow, PDB 1CHO) <sup>14</sup> that was replaced with an ester via total chemical synthesis. Purple structure is chymotrypsin with active site serine residues shown. Dotted lines indicate polar contacts necessary for recognition of OMTKY3 inhibitor. ....	6
Figure 6. Thioamides can quench intrinsic fluorophores in proteins and provide a reliable readout of distance between the fluorescent donor and the acceptor thioamide. Graph on the right shows folded structure of Villin Hp35 (left) as a function of temperature monitored by circular dichorism measurements. The measured distance between the donor Cnf <sub>35</sub> and acceptor thioleucine (Leu') correlated well with conformation of the protein. Figure used with permission from reference 22. Copyright 2010 American Chemical Society.....	7
Figure 7. The crystal structure of racemic plectasin (PDB: 3E7R). <sup>34</sup> .....	8

Figure 8. Structure of thermolysin C-terminal subdomain (PDB 1TRL).<sup>36</sup> Alanine residues that were systematically replaced with Aib are colored green in both monomers of the dimeric protein. Stick models of Ala-residues are shown only in one monomer for clarity. .... 10

Figure 9. Schematic depicting how *N*-methyl substitution removes a backbone hydrogen bond donor. .... 11

Figure 10. Crystal structure of a synthetic peptide (yellow) containing a peptoid (pink) residue bound to the SH3 domain of Sem5 (gray surface) (PDB 2SEM).<sup>46</sup> The cartoon highlights the peptoid residue in the sequence. .... 12

Figure 11. Crystal structure of an analogue of a tetrameric coiled-coil protein bearing 1,2,3-triazoles (PDB 1UF9).<sup>53</sup> Box shows a zoom of a hydrogen bond between a backbone amide and a triazole nitrogen. .... 13

Figure 12. Structure of  $\beta$ -peptide bundle that catalyzes hydrolysis of an aromatic ester substrate. Figure used with permission from reference 74. Copyright 2014 American Chemical Society. . 16

Figure 13. Cartoon of zinc-dependent helix bundle fold of  $\alpha$ -peptoid oligomer previously reported.<sup>76</sup> Stars indicated fluorophores that participated in FRET only when in close proximity. Metal binding side-chains of the peptoid helix bound zinc and stabilized bundle formation. .... 17

Figure 14. Structure of GB1 (PDB 2QMT). .... 21

Figure 15. Secondary structure of GB1 and sequences of wild-type 1 and PEG mutants 2-4. .... 23

Figure 16. CD scans at 4 °C and thermal melts of 40  $\mu$ M proteins 1-4 in 20 mM phosphate buffered water pH 7. .... 24

Figure 17. GPC chromatograms of 40  $\mu$ M proteins 1-4 in 150 mM sodium chloride, 20 mM phosphate, pH 7 . GPC was performed on a Superdex-75 column. .... 26

Figure 18. Competition ELISA and IC<sub>50</sub> values for proteins 1-4. .... 27

Figure 19. The crystal structure of GB2 (yellow; C-terminal loop colored magenta) bound to F <sub>c</sub> domain of IgG (gray surface) (PDB: 1FCC). <sup>99</sup> .....	28
Figure 20. MD simulations of proteins 1-4. A-D) Snapshots from MD simulations of proteins A) 1, B) 2, C) 3 and D) 4 with RMSD plots for backbone atoms. E-H) Close-up views of the PEG segments in proteins E) 2, F) 3, and G-H) 4 (G and H depict N-terminal and C-terminal loops, respectively). Figure used with permission from reference 91. Copyright 2012 Wiley.....	29
Figure 21. RMSD plots for the backbone atoms in the N-terminal loop of A) protein 1 and B) mutant 2. Analysis of the dynamics of the loop dihedrals in C) protein 1 and D) mutant 2. For each dihedral in the residue 19-21 loop (key to angle nomenclature shown in the structure), plots are shown for time-dependent rotation about the bond and the frequency distribution of dihedral values over the course of the 1 μs simulation. Dihedrals are plotted either 0° to 360° or -180° to 180° for clarity. Figure used with permission from reference 91. Copyright 2012 Wiley. ....	32
Figure 22. Secondary structure of GB1 and sequence of mutant 5. β <sup>3</sup> -residues are highlighted blue and bolded. R is the side-chain group of the corresponding α-residue.....	33
Figure 23. CD scans at 25 °C and thermal melts of proteins 1 and 5 in 20 mM phosphate buffered water pH 7. ....	34
Figure 24. A) Protein 5 (yellow) overlaid with wild-type GB1 (purple, PDB: 2QMT). B, C) Close ups of C-terminal loop polar contacts (yellow dashed lines) in wild-type GB1 (B) and mutant 5 (C). β <sup>3</sup> -Residues are colored cyan. ....	36
Figure 25. Sequences of GB1 helix mutants 6 and 7. R is the side-chain group of the corresponding α-residue.....	37
Figure 26. CD scans at 25 °C and thermal melts of 40 μM wild-type 1 and mutants 6 and 7 in 20 mM phosphate buffered water pH 7. ....	38

Figure 27. Overlay of wild-type GB1 structure (2QMT) with refined crystal structures of 6 (yellow) and 7 (gray). $\beta^3$ -Residues are colored cyan.....	39
Figure 28. A) Overlay of helix backbone and side-chains in crystal structures of 6 (yellow) and 7 (gray). B) Tertiary interaction in the structure of 6 of side-chain from $\beta^3$ -Lys <sub>31</sub> with side-chain of Trp <sub>43</sub> .....	41
Figure 29. Change in side-chain display in hairpin peptide as a result of 1:1 $\alpha/\beta$ substitution. $\beta$ -Residues are colored cyan and side-chain atoms of key hydrophobic core residues are shown as gray spheres. Data from reference 119. ....	42
Figure 30. Dipeptide substitution designs for hairpin mimicry in mixed $\alpha/\beta$ -peptides. R is the side-chain group(s) of the corresponding $\alpha$ -residue(s). Adapted with permission from reference 120.....	44
Figure 31. Sequences of GB1 sheet mutants 8 and 9. R is the side-chain group of the corresponding $\alpha$ -residue.....	45
Figure 32. CD scans at 25 °C and thermal melts of 40 $\mu$ M GB1 sheet mutants 8 and 9 in 20 mM phosphate buffered water, pH 7.....	46
Figure 33. Sequence of <i>N</i> -Methyl GB1 sheet mutant 10. R is the side-chain group of the corresponding $\alpha$ -residue.....	47
Figure 34. CD scans at 25 °C and thermal melts of 40 $\mu$ M wild-type 1 and mutant 10 in 20 mM phosphate buffered water pH 7.....	48
Figure 35. Example of a canonical four-residue $\beta$ -loop, $i - i + 3$ . $\beta$ -Turn residues are colored magenta. Side-chains are omitted for clarity. Dotted black lines indicate interstrand hydrogen bonds. $\beta$ -Loop adapted from GB1 (PDB: 2QMT).....	49
Figure 36. Examples of several $\beta$ -turn mimics studied in the literature. ....	50

Figure 37. Sequences of turn mutants 11-13. ....	51
Figure 38. CD scans and melts of 40 $\mu$ M turn mutants 11-13 in phosphate buffered water pH 7. .....	52
Figure 39. Close-up of A) N-terminal and B) C-terminal turns from 13 (yellow) overlaid with those from wild-type GB1 (magenta) (2QMT).....	53
Figure 40. Sequences of wild-type GB1 1 and mutants used to design All mutant 14. R is the side-chain group of the corresponding $\alpha$ -residue.....	55
Figure 41. CD scans at 25 $^{\circ}$ C and thermal melts of 40 $\mu$ M wild-type 1 and mutant 14 in 20 mM phosphate buffered water pH 7.....	56
Figure 42. Sequences of wild-type protein 1, all mutant 14, NuG 15, and NuG-All mutant 16. Underlined residues indicate side-chain sequence alterations relative to 1. R is the side-chain group of the corresponding $\alpha$ -residue. ....	57
Figure 43. CD scans at 25 $^{\circ}$ C and thermal melts of 40 $\mu$ M wild-type 1 and mutants 14 and 16 in 20 mM phosphate buffered water pH 7. ....	59
Figure 44. Crude HPLC traces from GB1 microwave-assisted SPPS before and after optimization. ....	62
Figure 45. Location of pseudoproline dipeptide residues in synthesis of wild-type 1 sequence. The oxazolidine-based backbone is removed under acidic cleavage conditions to yield two $\alpha$ -residues. ....	64
Figure 46. Crude HPLC trace from GB1 SPPS by automated room temperature methods. ....	65
Figure 47. Analytical HPLC chromatograms of purified proteins 1-14, and 16. ....	69
Figure 48. Global fit of thermal and guanidinium chloride denaturation of 40 $\mu$ M 1 in 20 mM phosphate buffer, pH 7.....	80

Figure 49. Changes in thermodynamic parameters of folding for mutants 5-7 relative to wild-type GB1 (1). .....	81
Figure 50. Putative $n \rightarrow \pi^*$ interactions in helices of wild-type GB1 and 6. Dotted lines indicate orbital interactions between $n_p$ electrons of carbonyl oxygens and $\pi^*$ orbital of neighboring carbonyl carbons. ....	83
Figure 51. Hypothesized differences between denatured ensembles of native and $\beta^3$ -containing proteins. Gray spheres represent hydrophobic side-chains. ....	85
Figure 52. Sequences of synthesized helix mutants 6 and 17-19 and biologically expressed mutant 20. Underlined residues indicate point mutations relative to synthesized wild-type 1. R is the side-chain group of the corresponding $\alpha$ -residue. ....	87
Figure 53. Proposed intraresidue $n \rightarrow \pi^*$ interaction possible in $\beta^2$ -Asn. ....	88
Figure 54. A) Overlay of helices from crystal structures of 6 (yellow), 17 (green), and 18 (gray). B) Close-up of $\beta^3/\beta^2$ Ala <sub>24</sub> in 6 and 17. C) Close-up of $\beta^3/\beta^2$ -Asn <sub>35</sub> in 6 and 18. Dotted line highlights a putative intraresidue orbital interaction in $\beta^2$ -Asn <sub>35</sub> that may contribute favorably to folded stability of protein. ....	89
Figure 55. Changes in thermodynamic parameters of folding for mutants 17-19 relative to 6. ....	91
Figure 56. Sequences of mutants 6 and 21-24. R is the side-chain group of the corresponding $\alpha$ -residue. ....	93
Figure 57. A) Overlay of helices from crystal structures of 6 (yellow), 21 (green), and 24 (ivory). B) Close-up of helices in 6 and 21. C) Close-up of loop in 5 (orange) and 24. ....	94
Figure 58. Changes in thermodynamic parameters of folding for mutants 21-24 relative to helix mutant 6. ....	95



Figure 59. Sequences of C <sub>α</sub> -methylated GB1 analogues 25 and 26. R is the side-chain group of the corresponding α-residue.....	97
Figure 60. A) Overlay of helices from 6 (yellow) and 25 (gray). B) Close-up of Ala/Aib <sub>24</sub> in 6/25. C) Close-up of ACPC/Aib <sub>35</sub> in 21 (orange)/26. Aib residues are colored green in all structures.....	98
Figure 61. Changes in thermodynamic parameters of folding for mutants 21, 25, and 26 relative to helix mutant 6. ....	100
Figure 62. Analytical HPLC chromatograms of purified proteins 17-26. ....	111
Figure 63. Fits of combined thermal and chemical melts of proteins 1, 5-7, 17-26. ....	116
Figure 64. Sequences of model hairpin 33 and N-methyl mutants 34-37. An NMR solution structure of GB1m2A <sup>120</sup> with the key hydrophobic core residues colored purple is shown on the right.....	121
Figure 65. Rationale of <i>trans</i> amide identification based on NOE contacts observed in 2D NMR experiments. Black arrows indicate key observed NOE contacts.....	122
Figure 66. Cartoon of how a <i>cis</i> amide could severely destabilize a hairpin fold by displacing a critical hydrophobic side-chain from the core. ....	123
Figure 67. Sequences of GB1m2A hairpin 33 and mutants 38 and 39 with NMR solution structures of 33 (gray) and 38 (yellow) overlaid. Data from reference 198. ....	125
Figure 68. Solution structure of 39 <sub>cyc</sub> from NMR analyses. γ <sup>4</sup> -Residues are colored pink. ....	127
Figure 69. Sequences of wild-type 1, N-methyl sheet mutant 10, and sheet mutants 40-42. Underlined residues indicate side-chain changes relative to 1. ....	128
Figure 70. CD scans and thermal melts of wild-type 1 and sheet mutants 10 and 40-42 in 20 mM phosphate buffered water pH 7. ....	129

Figure 71. Model based on wild-type structure (2QMT) with location of unnatural residues in GB1 sheet mutants 41 and 43 colored green. ....	130
Figure 72. Sequences of wild-type 1 and sheet mutants 41 and 43. ....	130
Figure 73. CD scans and thermal melts of wild-type 1 and sheet mutants 41 and 43 in 20 mM phosphate buffered water pH 7. ....	131
Figure 74. Analytical HPLC chromatograms of purified proteins 40-43. ....	135

## LIST OF SCHEMES

Scheme 3.1 Synthesis of Fmoc- $\beta^2$ -Asn(Dmcp)-OH .....	105
---	-----

## LIST OF EQUATIONS

Equation 2.1 .....	72
Equation 3.1 .....	112
Equation 3.2 .....	112
Equation 3.3 .....	113
Equation 3.4 .....	113

## LIST OF ABBREVIATIONS

CNS	Crystallography & NMR System
COSY	Correlation spectroscopy
DCM	Dichloromethane
DIEA	N,N-diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
EDT	1,2-ethanedithiol
FCC	Flash Column Chromatography
Fmoc-OSu	9-fluorenylmethyl N-succinimidyl carbonate
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5- b]pyridinium-3-oxid hexafluorophosphate
HCTU	2-(6-Chloro - 1H - benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography

HRMS	High-resolution mass spectrometry
MALDI-TOF	Matrix assisted laser desorption ionization – time of flight
MD	Molecular dynamics
MS	Mass spectrometry
NMP	<i>N</i> -methyl-2-pyrrolidone
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NMR	Nuclear magnetic resonance
PyAOP	(7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RMSD	Root-mean-square-deviation
TOCSY	Total correlation spectroscopy
THF	Tetrahydrofuran
TIS	Triisopropylsilane
TFA	Trifluoroacetic acid

## ACKNOWLEDGEMENTS

There are many people who have contributed to my development as a scientist and person throughout my PhD journey whom I would like to thank. First and foremost, I must thank my advisor, Seth Horne. Seth has been extremely patient with me. I distinctively remember starting my first peptide synthesis and having Seth check in on me. When he found me, I was weighing amino acids out on the balance and it looked like a scene from Scarface. He promptly showed me a better way to weigh out powders. Every interaction I have had with him has improved my vocabulary and ability to articulate my research. The lessons I've learned from group meetings and practice talks were all invaluable and I thank Seth for his guidance throughout the years. Seth was also kind enough to write many letters of recommendation and allowed me to pursue a summer internship in Germany; the experience was invaluable both scientifically and personally.

I thank Drs. Steve Weber, Ronald Wetzel, and Paul Floreancig for serving as my PhD committee members and Drs. Xinyu Liu and Kabirul Islam for serving as my proposal committee. I'd also like to thank Dr. Alex Deiters for mentoring my original research proposal. Although we both knew the idea was never going to work, I appreciated your method of teaching me how to write the different sections of the proposal and offering constructive feedback.

I have to thank the mass spec, NMR, machine, glass, and electronics shops at Pitt. I always was pleased by the service and advice I received and will certainly miss these facilities.

Having spent five years in the same lab, I've made friends who have provided scientific and emotional support during my stay. In particular I would like to thank George Lengyel, Conor Haney, Kaylyn Oshaben, Halina Werner, and Kelly George. Early on, George would provide encouragement ("you're going to fail"), organize impromptu lab outings to Hem's, host hockey nights, and generally be a source of comforting bitterness/realism along with Conor and Halina. Conor, George, and Halina provided a great deal of guidance in my synthetic endeavors as well as written documents. I thank Kaylyn for making and sharing delicious food and providing a number of pleasant distractions from unyielding science (box robot, decorating the lab with Pens gear). Kelly was truly the first graduate student I mentored extensively and helped me develop teaching skills. Other members of the lab including Nate Tavenor, Haley Grimm, Hajira Hotiana, and Chino Cabalteja have been enjoyable coworkers and I especially wish the younger "generations" continued success in their research. Brian Griffith was a fantastic undergraduate researcher who remained optimistic even in the face of terrible organic syntheses. Tim Cunningham trained me in bacterial expression of proteins and I hope to continue to improve my biology skillset.

I thank my family and friends for being there for me. My Mom, Mickey, and sisters have all visited me several times during my stay in Pittsburgh and I loved seeing them. The DeCarlos kindly welcomed me into their family and they are certainly my home away from home. I especially thank my love, Katlyn. You have been so supportive these past three years, whether it was surprise visits at work/home, surprise visits bearing food, or just listening to my complaints about work and life. Thank you, I love you, and I can't wait to start our life together in California (at least it's not Germany).



A special thank you to my Mom and Mickey for the financial support, moving help, apartment hunting, phone calls, visits, and care packages. Mom, you are the best mother anybody could ever ask for. I have always been inspired by your work ethic and desire to always try to do more than expected. You have always had my back. Thank you, I love you, and I dedicate this work to you and Nana.

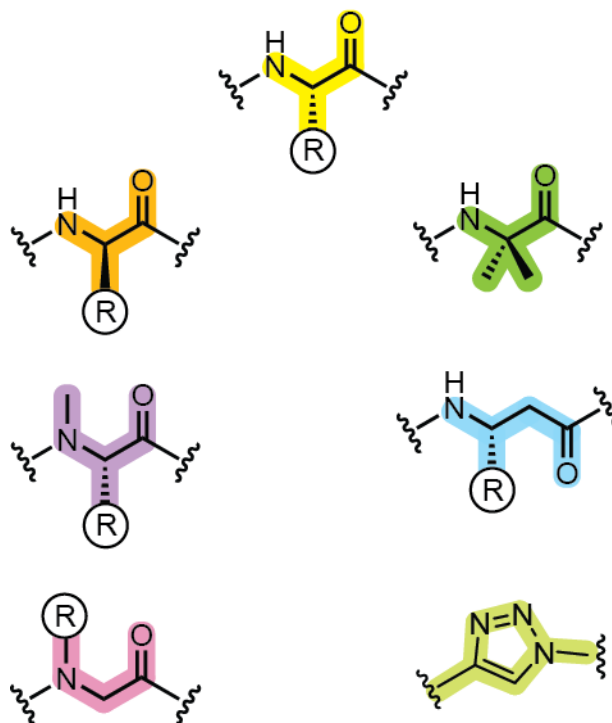
## 1.0 INTRODUCTION TO PROTEIN STRUCTURE MIMICRY

A portion of the background discussed in this chapter was reviewed in:

Z.E. Reinert, W.S. Horne. "Protein backbone engineering as a strategy to advance foldamers toward the frontier of protein-like tertiary structure." *Org. Biomol. Chem.* **2014**, *12*, 8796-8802.

### 1.1 BACKBONE MODIFICATION OF PROTEINS

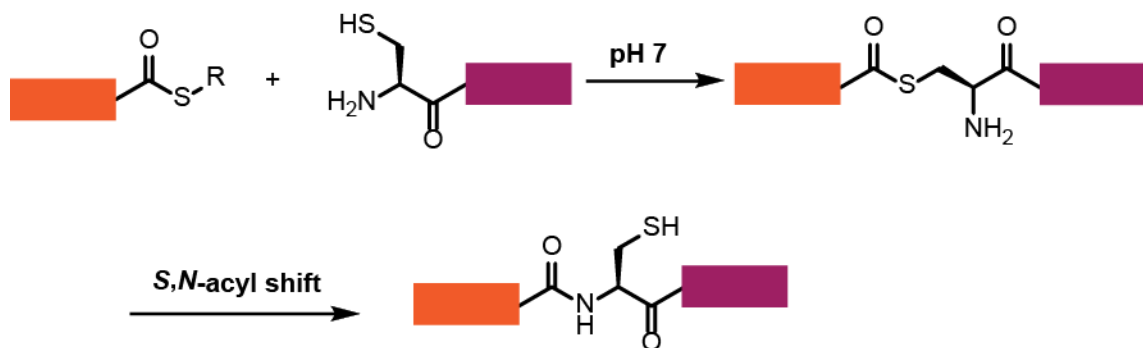
Proteins, polymers of  $\alpha$ -amino acids, perform an incredible number of functions and tasks in living systems including catalysis, signal transduction, immune response, and replication of genetic material. The vast majority of proteins' biological roles are determined by their three dimensional structure. Many intricate folds have been identified through a variety of high-resolution methods. Given the wide range of utility and function available to proteins, it is interesting to probe how unique polymers of  $\alpha$ -amino acids are in their ability to form such complex structures. Stated another way, can the sequence-encoded folds of proteins be recapitulated on non-native backbones?



**Figure 1.** Selection of unnatural residues that have been explored as replacements for  $\alpha$ -residues (yellow) in folded proteins.

Early work exploring protein backbone alterations was driven in part by two significant technological advances in protein synthesis that enabled the construction of large proteins bearing noncanonical amino acids (Figure 1): native<sup>1</sup> or expressed protein chemical ligation<sup>2</sup> and heterologous expression of proteins containing unnatural amino acids.<sup>3</sup> Native chemical ligation (NCL) is based on the reaction of two peptide fragments: one bearing an N-terminal Cys residue and the other with a C-terminal thioester (Figure 2). Thioester exchange from attack of the N-terminal thiol on the thioester carbonyl carbon forms a peptidyl acyl thioester that readily undergoes an *S,N*-acyl shift to yield a native peptide bond. The C-terminal thioester may be

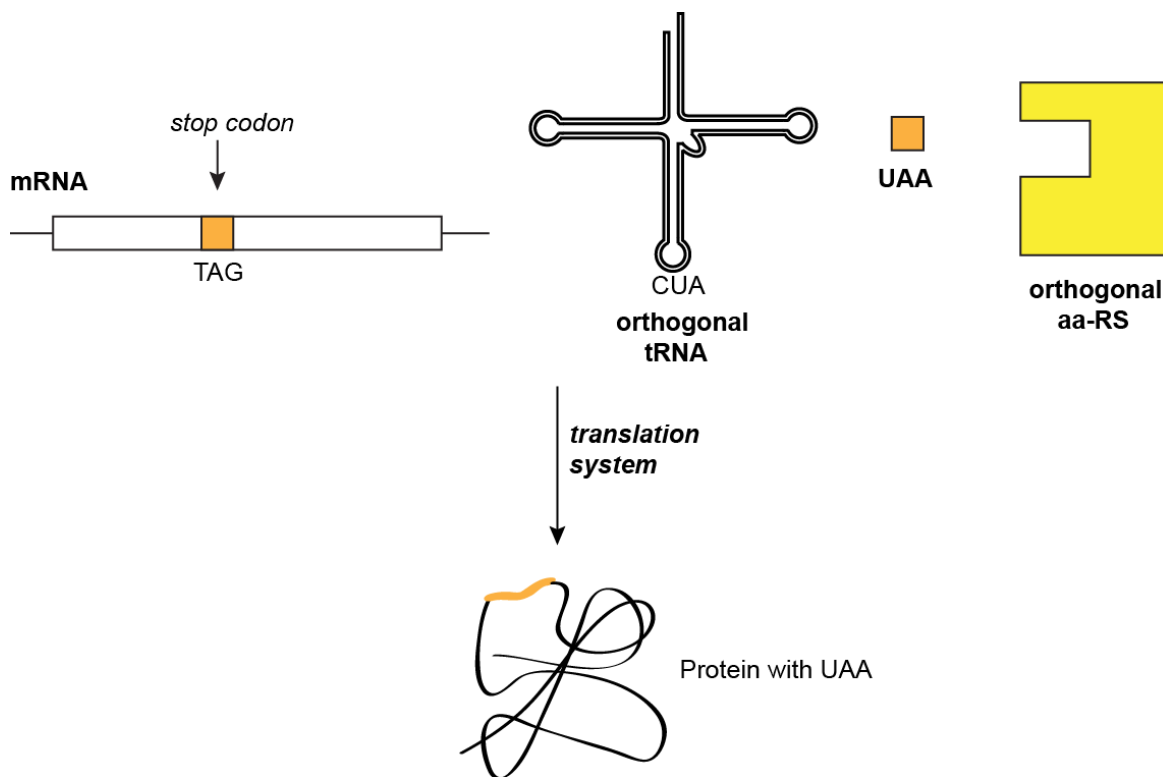
generated through SPPS or from biological expression through thiol-mediated intein cleavage as in the case of expressed protein ligation.<sup>2, 4-5</sup>



**Figure 2.** Overview of NCL.

Expression of proteins coding for unnatural amino acids is made possible through nonsense suppression and orthogonal acylated-tRNAs (Figure 3). In short, a stop codon is incorporated at a position for an unnatural residue in a protein-coding gene. A tRNA bearing the anticodon of the stop codon is charged with the desired unnatural amino acid using an engineered aminoacyl-tRNA synthetase. The exogenous synthetase and tRNA must be orthogonal to the host cell machinery to efficiently synthesize the target protein. During translation of the mRNA transcript, the charged tRNA is recognized by the ribosome and ultimately the unnatural protein is produced. This approach has been used for numerous applications in proteins, including introduction of biophysical probes,<sup>6-7</sup> backbone mutations,<sup>3, 8-9</sup> and reactive and/or fluorescent groups.<sup>7, 10-11</sup> Although many noncanonical amino acids can be

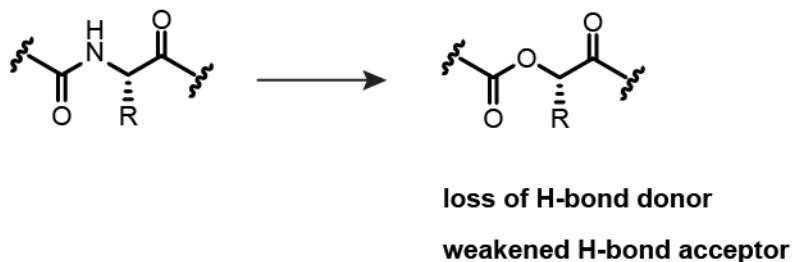
incorporated using nonsense suppression,<sup>12</sup> total chemical synthesis remains the method of choice for production of highly backbone-modified proteins.



**Figure 3.** Overview of expression of proteins with unnatural amino acids (UAA) using engineered aminoacyl tRNA synthetases (aa-RS).

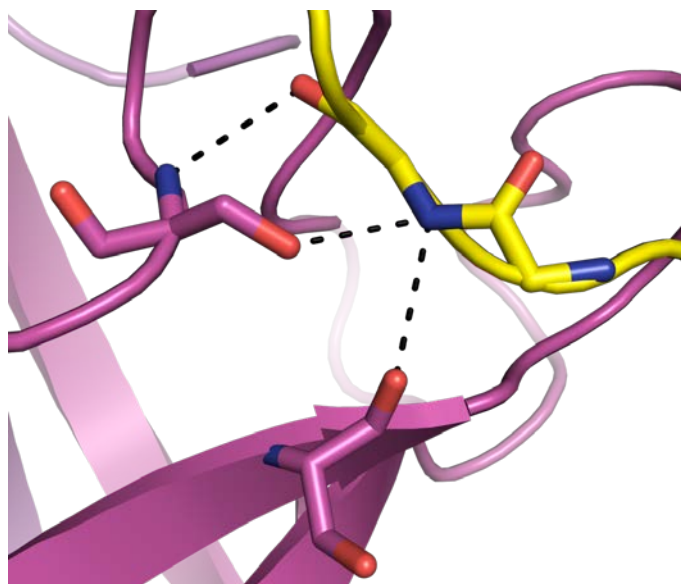
Backbone engineering of proteins was first applied to investigate the thermodynamic contribution of backbone hydrogen bonds to folding. While known to be energetically important, probing individual hydrogen bonds was experimentally challenging due to the need for methods that minimally disrupted the folded structure. One solution is to replace an  $\alpha$ -amino acid in a protein with the  $\alpha$ -hydroxy acid analogue. This has the effect of exchanging the amide N-H for

an oxygen, which cannot act as a H-bond donor, and weakens the neighboring carbonyl oxygen's ability to act as an acceptor (Figure 4).<sup>13</sup>



**Figure 4.** Summary of main consequences of amide to ester substitution in proteins.

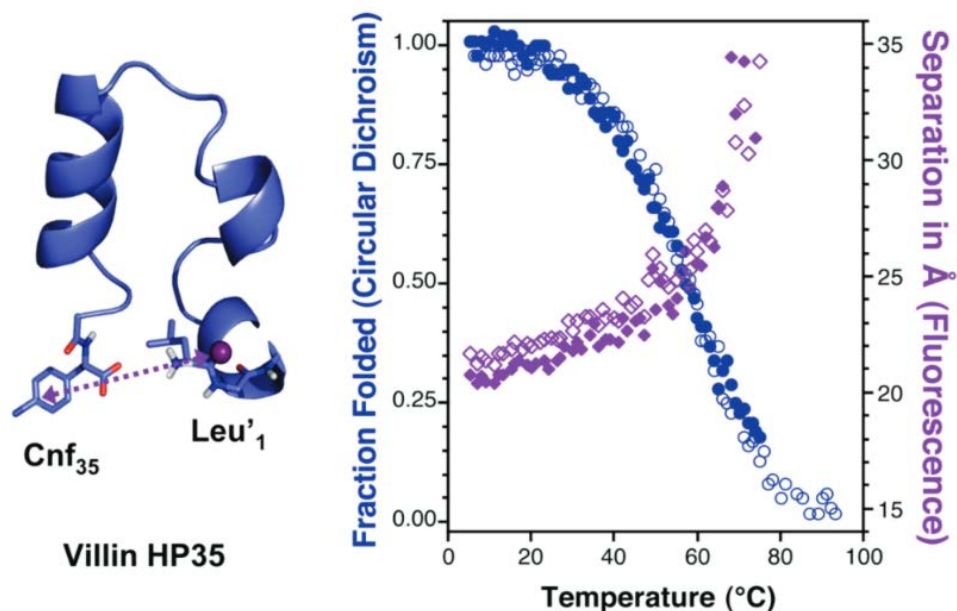
Enabled by the aforementioned technologies, an isolated amide to ester substitution was initially performed in turkey ovomucoid third domain (OMTKY3) via total chemical synthesis.<sup>14</sup> OMTKY3 is known to inhibit serine proteases and is recognized by chymotrypsin at Leu<sub>18</sub> primarily through hydrogen bonds (Figure 5). When the amide of Leu<sub>18</sub> in OMTKY3 was replaced with an ester, the resulting depsipeptide had reduced affinity for the protease. The thermodynamic contribution of the replaced hydrogen bond towards binding was found to be ~1.5 kcal/mol. Several other seminal studies have since added to the estimated folding energy values for backbone hydrogen bonds in a variety of contexts.<sup>14-19</sup> These efforts have revealed that hydrogen bonds in hydrophobic environments contribute the greatest free energy to protein folding.



**Figure 5.** Close-up of amide in OMTKY3 (yellow, PDB 1CHO)<sup>14</sup> that was replaced with an ester via total chemical synthesis. Purple structure is chymotrypsin with active site serine residues shown.

Dotted lines indicate polar contacts necessary for recognition of OMTKY3 inhibitor.

Another isosteric replacement for amides is substitution of the carbonyl oxygen for sulfur. Similar to esters, thioamides have been employed to provide information on the strength of hydrogen bonds in protein-protein interactions.<sup>20</sup> The electronic properties of thioamides also enable them to act as minimalistic quenchers for both natural<sup>21</sup> and synthetic<sup>22</sup> fluorescent moieties in proteins (Figure 6). This feature has been applied to study the folding mechanisms of proteins known to aggregate in neurological disorders.<sup>23-24</sup> In addition, the sulfur atom enhances the ability of thioamides relative to amides to participate in carbonyl-based stabilizing orbital interactions in proteins.<sup>25</sup>

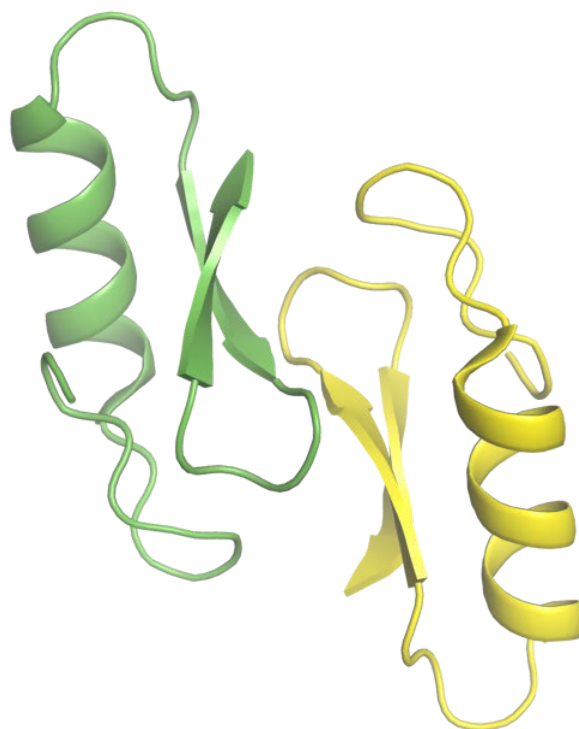


**Figure 6.** Thioamides can quench intrinsic fluorophores in proteins and provide a reliable readout of distance between the fluorescent donor and the acceptor thioamide. Graph on the right shows folded structure of Villin Hp35 (left) as a function of temperature monitored by circular dichroism measurements. The measured distance between the donor Cnf<sub>35</sub> and acceptor thioleucine (Leu') correlated well with conformation of the protein. Figure used with permission from reference 22. Copyright 2010 American Chemical Society.

Proteins in nature are almost exclusively composed of L- $\alpha$ -amino acids, however, synthetic peptides are not subject to this limitation. Enantiomers of proteins constructed from entirely D-amino acids have been reported.<sup>26-28</sup> Intuitively from first principles, D-proteins were found to adopt mirror-image folds and, in the case of enzymes, mirror image substrate specificity. Aside from providing support for fundamental paradigms of chirality, D-proteins have found use in the crystallization of proteins (Figure 7).<sup>29-30</sup> Briefly, because proteins are intrinsically chiral, they can only crystallize in 65 unique space groups. Racemic protein



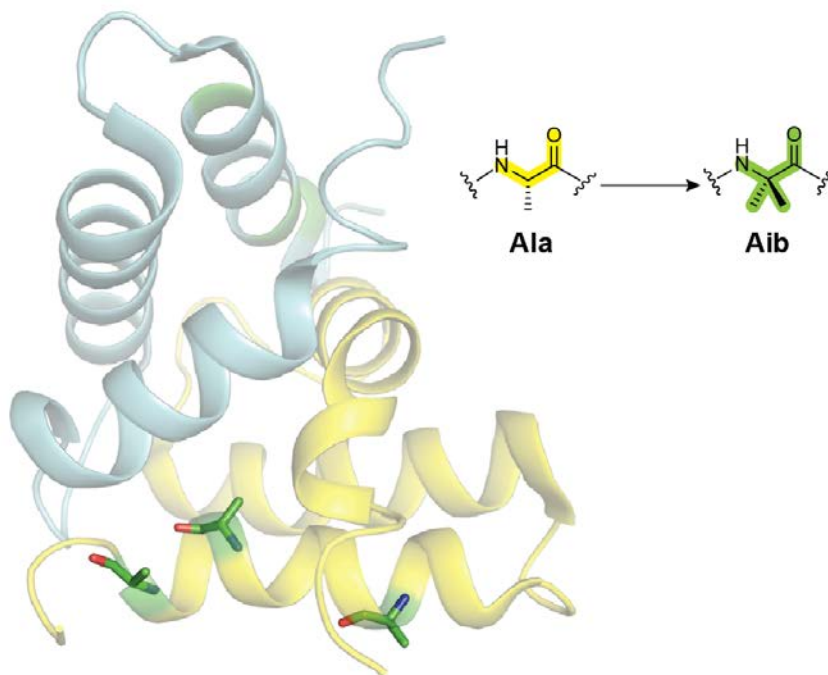
mixtures, on the other hand, can crystallize in a larger number of achiral space groups that are predicted to be much more favorable to typical protein conformations.<sup>31</sup> Several high-resolution structures of proteins that had been previously elusive or refractory towards standard crystallization techniques have been obtained using this method.<sup>32-33</sup>



**Figure 7.** The crystal structure of racemic plectasin (PDB: 3E7R).<sup>34</sup>

The above work largely focused on modification of amides to explore the role of hydrogen bonds in protein folding. Other researchers have turned to replacing  $\alpha$ -amino acids in proteins as a means of improving upon folding and/or function of natural protein scaffolds. A

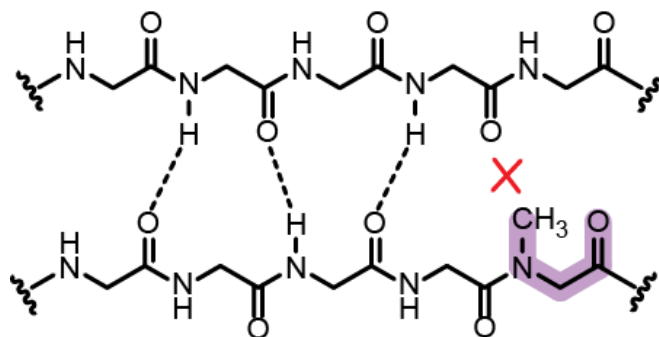
well-known class of amino acids for stabilizing helical secondary structure is C $\alpha$ -methyl- $\alpha$ -residues.<sup>35-36</sup> Methylation of C $\alpha$  significantly reduces the backbone conformational freedom resulting in an increased preference for helical folds. Aminoisobutyric acid (Aib) is a naturally occurring C $\alpha$ -methyl amino acid that was one of the first noncanonical amino acids to be incorporated in a protein through heterologous expression.<sup>3</sup> Prior work modified the backbone of a flexible flap region of HIV-1 protease with Aib.<sup>37</sup> The modified enzyme served as tool to obtain structural and kinetic insights into the catalytic mechanism. In a separate study, a series of Ala $\rightarrow$ Aib mutations were performed in a helix-forming subdomain of thermolysin (Figure 8).<sup>36</sup> An increase in folded stability relative to the wild-type protein was observed in all but one Aib substitution. Replacement of a flexible residue near the termini of the helix of thermolysin with Aib was unfavorable towards folding, which was attributed to a decrease in the entropy of the folded state.



**Figure 8.** Structure of thermolysin C-terminal subdomain (PDB 1TRL).<sup>36</sup> Alanine residues that were systematically replaced with Aib are colored green in both monomers of the dimeric protein. Stick models of Ala-residues are shown only in one monomer for clarity.

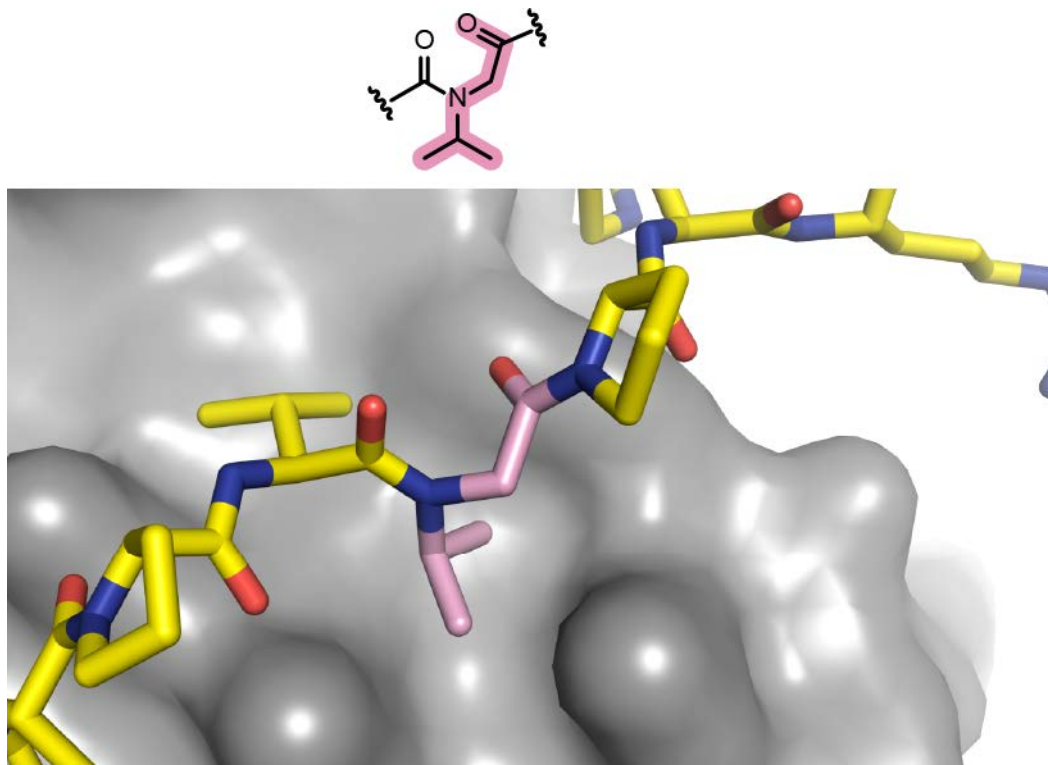
Similar to C $_{\alpha}$ -methylation, alkylation of the amide nitrogen has been explored as a protein backbone modification. *N*-methyl analogues of  $\alpha$ -amino acids are naturally occurring in cyclic peptides derived from microbes.<sup>38</sup> Methylation of the backbone nitrogen in proteins removes a hydrogen bond donor, a feature that has found utility in studying multimeric proteins. Interleukin-8 is a secreted protein known to self-associate primarily through a  $\beta$ -sheet formed from two strands of each monomer.<sup>39</sup> *N*-methyl residues were used to disrupt this interface by blocking an interstrand backbone hydrogen bond (Figure 9).<sup>40</sup> Interestingly, the resulting monomeric interleukin had biological activity identical to that of the wild-type. In addition to methyl groups, entire side-chains have been placed onto the amide nitrogen. Referred to as  $\alpha$ -

peptoids, these *N*-alkyl glycine derivatives have an achiral backbone.<sup>41-42</sup> Bulky chiral side-chains can be installed to promote right-handed helix formation.<sup>41, 43</sup> One notable property of peptoids is that they readily cross cellular membranes, making them excellent alternatives to natural peptides for studying intracellular protein-ligand interactions (Figure 10).<sup>42, 44-47</sup>



**Figure 9.** Schematic depicting how *N*-methyl substitution removes a backbone hydrogen bond donor.

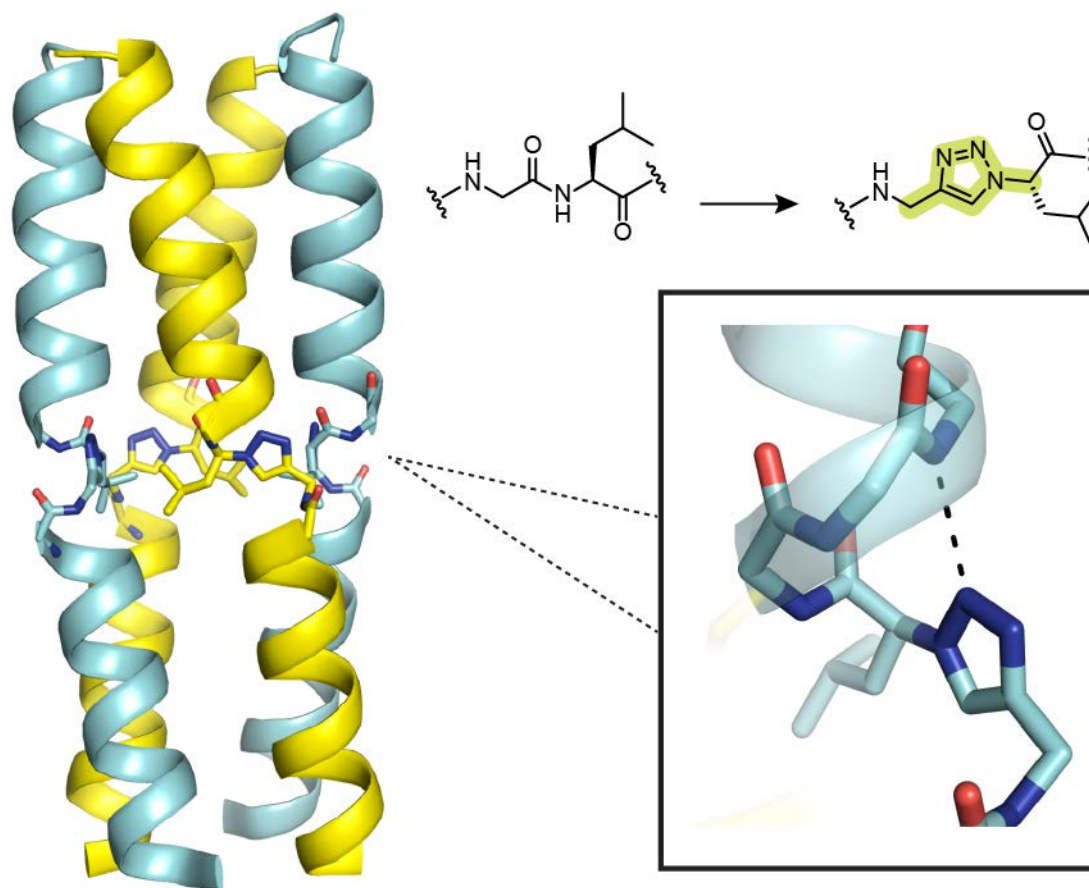
Insertion of a methylene unit into the backbone of an  $\alpha$ -amino acid generates a  $\beta$ -amino acid, and oligomers of  $\beta$ -amino acids are known as  $\beta$ -peptides.  $\beta$ -Peptides can adopt a diverse array of helical folds in solution beyond those seen with natural sequences.<sup>48-49</sup> Given their similarity to  $\alpha$ -peptides, entire secondary structures of natural proteins have been replaced with  $\beta$ -analogues.<sup>50-51</sup> In one example, the backbone of a helix from interleukin-8 was replaced by a  $\beta$ -peptide bearing the same side-chains using expressed protein ligation.<sup>50</sup> The hybrid  $\alpha/\beta$ -oligomer displayed diminished biological activity compared to the wild-type protein.



**Figure 10.** Crystal structure of a synthetic peptide (yellow) containing a peptoid (pink) residue bound to the SH3 domain of Sem5 (gray surface) (PDB 2SEM).<sup>46</sup> The cartoon highlights the peptoid residue in the sequence.

Moving away from methylated analogues of  $\alpha$ -amino acids, other exotic functional groups have been explored in the context of proteins. 1,2,3-triazoles, which have similar dipole moments as amides,<sup>52</sup> have been incorporated in a tetrameric  $\alpha$ -helical coiled-coil peptide (Figure 11).<sup>53</sup> The heterocycle-containing mutant retained the fold of the native peptide despite four central  $\alpha \rightarrow$  triazole substitutions. The role of *cis/trans* amide isomerization in protein folding has also been explored by triazoles. A prolinyl amide in ribonuclease A was replaced with one of two regioisomers of a 1,2,3-triazole. The 1,5-disubstituted triazole was found to be superior to the 1,4-disubstituted triazole at mimicking the native *cis* amide.<sup>54</sup> Other unnatural linkages used

to explore *cis/trans* isomerization include substituted alkenes,<sup>55-56</sup> however, the large difference in polarity and size between an amide and an alkene destabilized the protein fold considerably.



**Figure 11.** Crystal structure of an analogue of a tetrameric coiled-coil protein bearing 1,2,3-triazoles (PDB 1UF9).<sup>53</sup> Box shows a zoom of a hydrogen bond between a backbone amide and a triazole nitrogen.

## 1.2 HIGHER ORDER STRUCTURAL MIMICRY

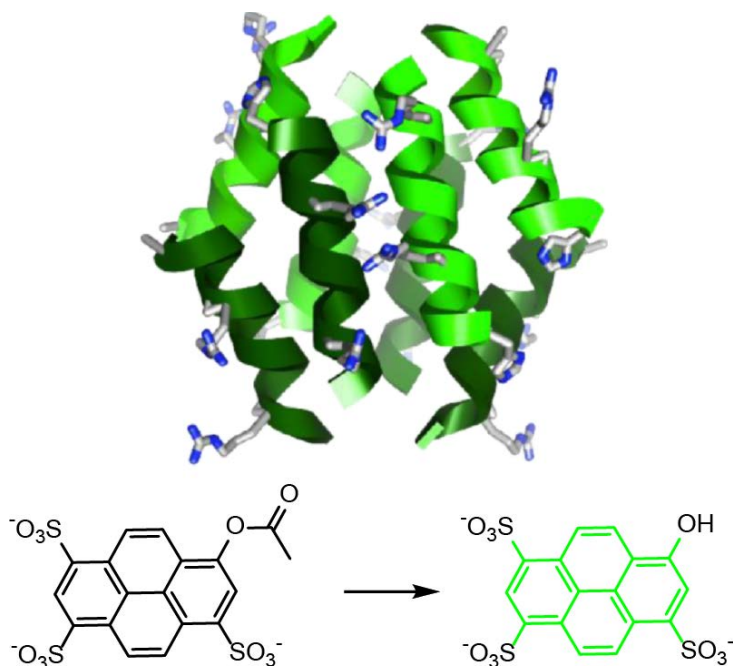
The aforementioned literature describes attempts to understand and improve upon the folding and function of natural proteins. In all examples, a very small subset of the total residues in a protein were modified. These studies are pivotal and reveal new insights into protein folding. They also raise the interesting possibility of significantly modifying the protein backbone. Increasing the ratio of modified:natural residues to a critical threshold results in a highly-unnatural oligomer that may exhibit folding behavior akin to the native sequence. Fundamentally, expanding the diversity of both chemical and structural motifs beyond that found in nature showcases the ability of non-native oligomers to adopt complex folds. From a practical standpoint, bioactive polymers that display protein-like folds on non peptidic backbones could find use in therapeutic applications. The proteolytic susceptibility of  $\alpha$ -peptides often precludes their use as drugs. In contrast, unnatural oligomers are not easily recognized by proteases and have longer half-lives in serum, prolonging their biological effects.<sup>57</sup>

Significant research efforts have explored the design and synthesis of foldamers, non-native polymers that have strong folding propensities.<sup>58</sup> Many common protein secondary structures such as helices<sup>58-60</sup> and turns<sup>61-69</sup> have been displayed by foldamers. However, few examples of higher order structural mimicry exist in the literature. In one report, blending  $\alpha$  and  $\beta$ -residues into a natural oligomeric coiled-coil produced mixed  $\alpha/\beta$ -peptides that retained the quaternary structure of the parent protein.<sup>70</sup> Extending this methodology to therapeutic peptides, the authors replaced several  $\alpha$ -residues in a helical sequence derived from HIV gp41 known to disrupt virus-cell fusion with  $\beta$ -residues.<sup>71</sup> The modified helices did not significantly alter the ability of the peptide to block cell-cell fusion except in cases where unnatural residues were incorporated near the N-terminus. It was hypothesized that the flexible  $\beta$ -residues were

entropically destabilizing towards helix formation. Systematic rigidification of the  $\alpha/\beta$ -backbone through the use of cyclic  $\beta$ -residues markedly improved the biological activity of the hybrid peptides. In addition, the unnatural backbone oligomers were one to two orders of magnitude more resistant to degradation by a promiscuous protease relative to the all  $\alpha$ -peptide.

Homooligomers of  $\beta$ -amino acids can also form helices that self-associate. A *de novo* designed  $\beta$ -dodecamer peptide was found to adopt a homooctameric helix bundle.<sup>72</sup> The folding of the  $\beta$ -bundle was highly dependent on concentration and required large amounts of oligomer to achieve a 90 % folded population. Motivated by this nonideal folding equilibrium, the authors computationally designed a variant of the  $\beta$ -peptide and linked two separate peptide chains together.<sup>73</sup> The second generation oligomer assembled into a tetramer with a higher affinity for self-association. Later revisions to the sequence produced a  $\beta$ -oligomer with mild esterase activity (Figure 12).<sup>74</sup>

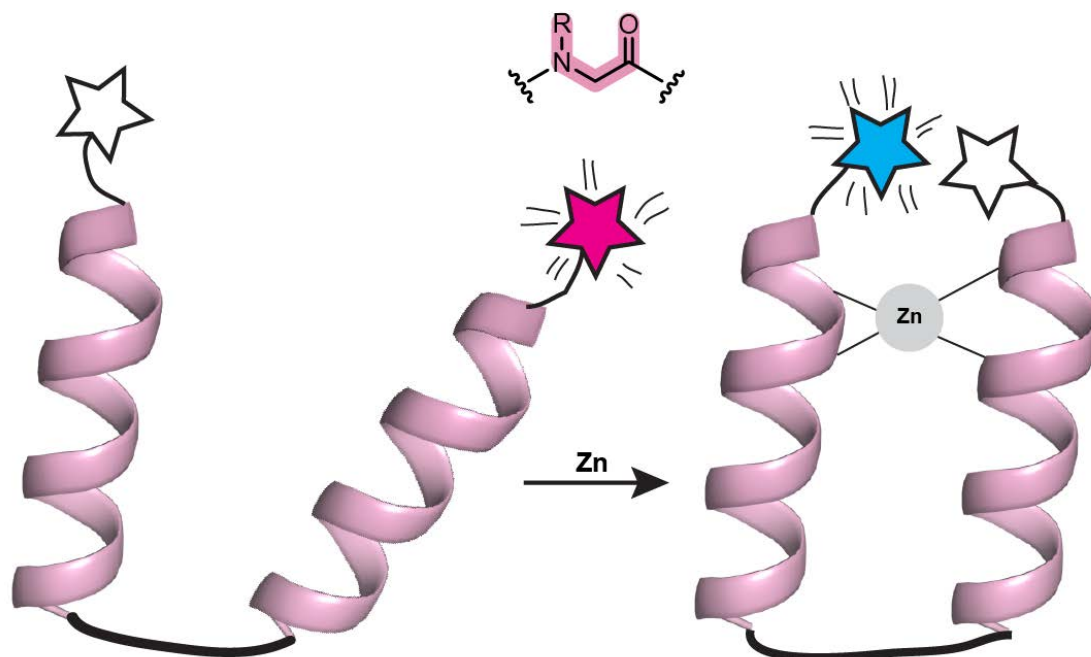




**Figure 12.** Structure of  $\beta$ -peptide bundle that catalyzes hydrolysis of an aromatic ester substrate.

Figure used with permission from reference 74. Copyright 2014 American Chemical Society.

Helix-turn-helix motifs have been recreated on unnatural backbones bearing  $\beta$ -residues.<sup>75</sup> as well as  $\alpha$ -peptoids.<sup>76</sup> In the latter case, *de novo* peptoid sequences that code for small covalently linked helices were substituted with imidazole and thiol side-chains resembling those of histidine and cysteine residues. In natural proteins, these functional groups can coordinate metal ions such as zinc. Using FRET, it was shown that the peptoid helix bundle had a more compact fold in the presence of zinc, supporting the design hypothesis (Figure 13).



**Figure 13.** Cartoon of zinc-dependent helix bundle fold of  $\alpha$ -peptoid oligomer previously reported.<sup>76</sup> Stars indicated fluorophores that participated in FRET only when in close proximity. Metal binding side-chains of the peptoid helix bound zinc and stabilized bundle formation.

Quaternary assemblies of helices have been explored on unnatural backbones. Recently, it was shown that urea-based polymers could fold and self-assemble into either helical bundles or channel-like structures in water.<sup>77</sup> These dodecamers of urea-based residues were synthesized using solid-phase methods and bore proteinaceous side-chains. Fine control of folding was achieved through changes in the primary sequence. Near atomic resolution crystal structures provided detailed insights into the relationship between sequence and structure.

### 1.3 GOALS

The few examples of tertiary structure mimicry underscore the challenge of pushing foldamer design beyond secondary structures. Accessing complex protein-like folds on protease resistant backbones could yield improved therapeutics with high target specificity and sustained half-lives in serum. The goal of the current work is to develop design strategies for protein tertiary structure mimicry. As tertiary structure is a combination of multiple secondary structure units, methods for loops, helices, sheets, and turns will be developed or improved upon from literature precedent. Provided that suitable design rules exist, many intricate higher order folds of proteins should be attainable by foldamers. Additionally, the impact of unnatural residues on protein folding thermodynamics is poorly understood. Exploring such effects could inform on design considerations for backbone modified proteins. The modifications of each secondary unit that are most thermodynamically favorable towards folding will be combined into one sequence to generate a foldameric derivative of the native protein. Ideally, this highly modified variant would possess a fold identical to that of the wild-type on a heterogeneous backbone oligomer.

Underlying our design strategy is the hypothesis that every protein has two orthogonal aspects: 1) side-chain and 2) backbone sequences. Alterations to one sequence would have identical effects towards folding on either a natural or unnatural backbone, provided that both oligomers maintain the same fold. Traditionally changes to proteins have focused largely on side-chain modifications due to technological limitations for longer sequences. Given the aforementioned improved methodologies for the synthesis of highly unnatural analogues of proteins, we envision that modulation of side-chain and backbone sequences in tandem could produce biomimetics with favorable properties. As an example, if backbone modification decreased the stability of the folded state, side-chain mutations may be implemented to

compensate for the loss of folding free energy. This dual-sequence design strategy should enable the design of oligomers with highly unnatural backbones that maintain the same three dimensional structure as the native protein.

## 2.0 DESIGN OF FOLDAMERS WITH PROTEIN-LIKE FOLDING BEHAVIOR

The results in this chapter have been published in:

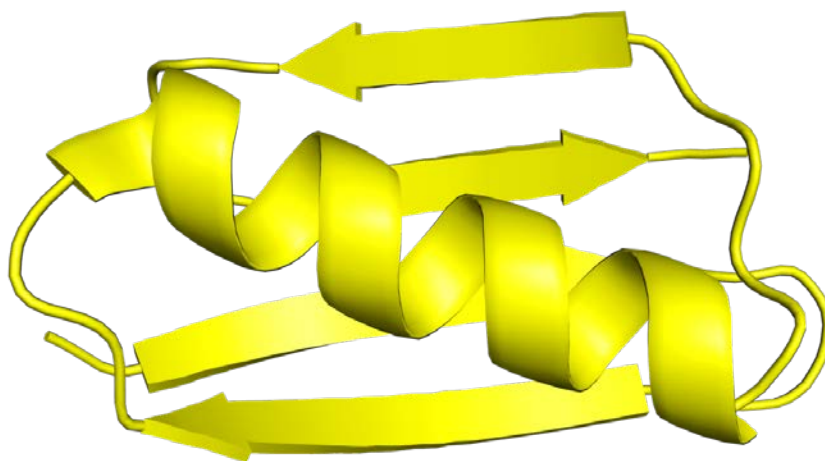
1. Z.E. Reinert, E.D. Musselman, A.H. Elcock, W.S. Horne. "A PEG-based oligomer as a backbone replacement for surface-exposed loops in a protein tertiary structure." *ChemBioChem*. **2012**, *13*, 1107-111.
2. Z.E. Reinert, G.A. Lengyel, W.S. Horne. "Protein-like tertiary folding behavior from heterogeneous backbones." *Journal of the American Chemical Society*. **2013**, *135*, 12528-12531.

The crystallography data collection and structure refinement were performed by W. Seth Horne. Molecular dynamics simulations were performed by Eli Musselman and Adrian Elcock (University of Iowa).

From Chapter 1, it is clear that mimicry of protein tertiary structure has not been widely explored despite the promise of significant fundamental and practical applications. We hypothesized that if suitable guidelines existed, a target protein of interest could be used a blueprint for the design of a foldameric derivative that would both retain the biological function of the native protein and be highly resistant to proteolytic degradation. We therefore set out to develop strategies for modifying the backbone of proteins with higher order structure.

## 2.1 MODEL SYSTEM GB1 BACKGROUND

We wanted to identify a model protein that could serve as a platform for examining various design strategies in the context of a tertiary structure. The Immunoglobulin binding domain 1 of protein G from *Streptococcus* (GB1, Figure 14) was selected as a model system for several reasons: 1) the entire protein can be synthesized using solid-phase peptide synthesis (SPPS), 2) the structure of GB1 has been well characterized by both high-resolution X-ray<sup>78</sup> and NMR<sup>79</sup> methods, 3) GB1 folds into a compact tertiary structure that includes every common secondary structure (loops, helix, sheet, turns), and 4) it has been previously shown that GB1 is tolerant of both backbone and side-chain modifications.<sup>80-84</sup> While the tertiary fold of GB1 is relatively simple compared to the folds of many larger proteins, it is at the frontier of complexity for foldamer targets. We hypothesized that design strategies developed for GB1 would be applicable to more complex proteins.



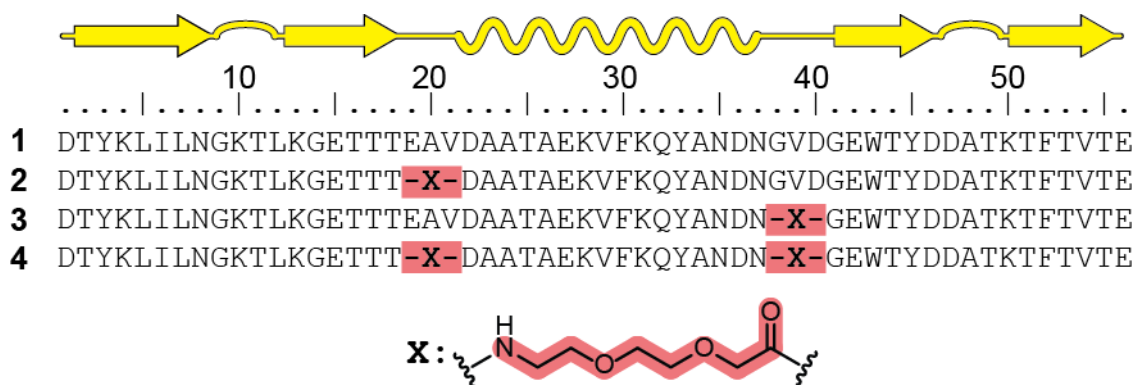
**Figure 14.** Structure of GB1 (PDB 2QMT).

## 2.2 SECONDARY STRUCTURE BACKBONE MODIFICATIONS

### 2.2.1 Loops

We first examined loop alteration in GB1 as loops are relatively unstructured regions of proteins and would be expected to be most tolerant of backbone modifications. Given this feature of loops, we hypothesized that enhancing the conformational freedom beyond what is possible in natural backbones may be well accommodated. Segments based on poly (ethylene) glycol (PEG) linkers were investigated as replacements for the loops of GB1. PEG is a water-soluble polymer that is often used as a spacer to link proteins to fluorophores or other chemical entities including nanoparticles.<sup>85</sup> PEGylation, the attachment of PEG to the side-chains of residues, has been shown to improve physiological stability of proteins,<sup>86-87</sup> this has led to several FDA-approved drugs based on PEG-modified proteins.<sup>88</sup> Few studies have explored PEG as a backbone replacement. Prior work found that certain ethylene glycol analogues of a natural 10-residue peptide can be effectively recognized by a thioesterase enzyme that is involved in cyclization of the antibiotic tyrocidine A.<sup>89</sup> The most active PEG-based substrates for the thioesterase were those that did not replace residues crucial to cyclization. In another study, a PEG-based oligomer served as a dipeptide replacement in 20-residue conotoxins stabilized by three disulfide bridges.<sup>90</sup> The resulting hybrid peptides displayed improved analgesic properties *in vivo*.

We envisioned that the enhanced conformational freedom of PEG-based loops could serve as a chemical tool to locally enhance backbone dynamics. There are two loops in GB1 (**1**) and each is approximately three residues in length. We thus selected a PEG-derived amino acid (**X**) that had the same number of backbone atoms as a tripeptide (Figure 15). We synthesized three mutants of GB1 where either the N-terminal (**2**), C-terminal (**3**), or both loops (**4**) were replaced with **X**.<sup>91</sup>

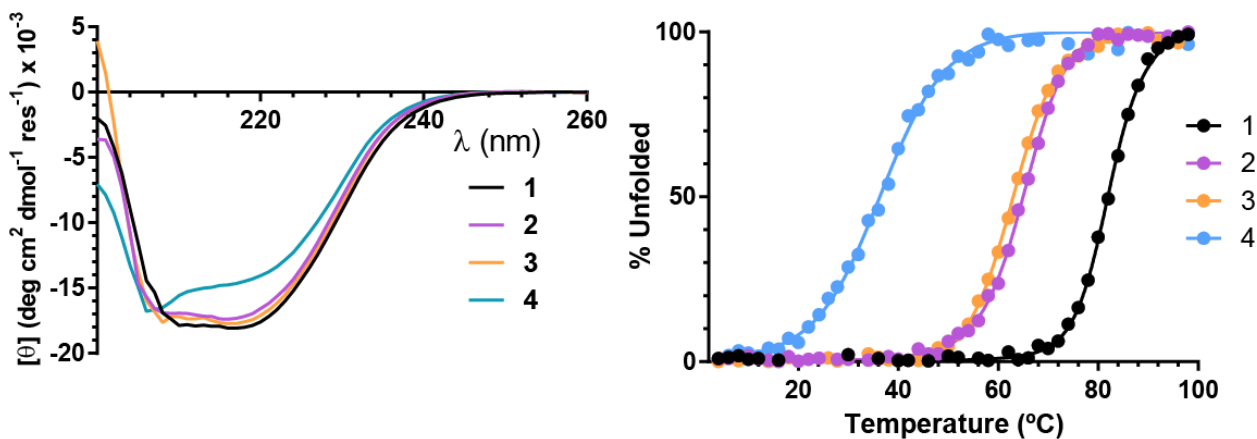


**Figure 15.** Secondary structure of GB1 and sequences of wild-type **1** and PEG mutants **2-4**.

Circular dichroism (CD) spectroscopy can provide information about the secondary structures present in a protein or peptide in solution.<sup>92</sup> Strong minima at 208 nm and 222 nm are characteristic of  $\alpha$ -helices whereas  $\beta$ -sheets will often show a single minimum at 215 nm. CD scans were performed on 40  $\mu$ M protein samples in 20 mM phosphate buffer pH 7 at 4  $^{\circ}$ C (Figure 16). The CD spectrum of **1** agreed with previously reported scans of wild-type GB1 from bacterial expression.<sup>93</sup> The single PEG mutants **2** and **3** both showed CD spectra similar to that



of protein **1** in both shape and magnitude. Mutant **4** had a slightly reduced molar ellipticity at 220 nm suggesting a less native-like fold.



**Figure 16.** CD scans at 4 °C and thermal melts of 40  $\mu\text{M}$  proteins **1-4** in 20 mM phosphate buffered water pH 7.

CD thermal melts can provide thermodynamic information on protein folding equilibria. A strong characteristic signal of the protein from the CD spectrum is monitored as the temperature of the solution is increased to unfold the protein. The resulting unfolding curve may be fit to obtain the melting temperature of the protein ( $T_m$ ), defined as the midpoint of the unfolding transition. Thermal denaturation of proteins **1-4** revealed highly cooperative and two-state unfolding transitions evidenced by the sigmoidal shape of the melting transition (Figure 16). Such cooperative folding behavior is a characteristic of well-folded tertiary structures in general<sup>94</sup> and specifically GB1<sup>95</sup>. Thermal stabilities of mutants **2** and **3** were less than that of wild-type **1** with an observed  $\Delta T_m$  of  $\sim 18$  °C (Table 1).  $\Delta T_m$  can be related to  $\Delta\Delta G_{\text{fold}}$ , the

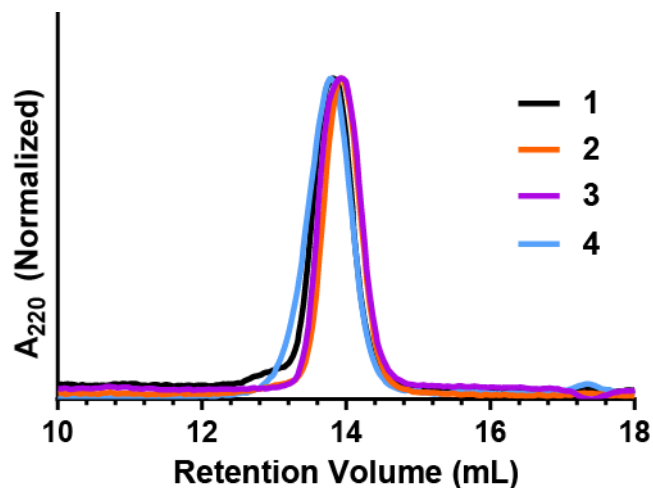
difference relative to wild-type in the change of free energy of folding (Equation **2.1**, see experimental).<sup>96-97</sup> The  $\Delta\Delta G_{\text{fold}}$  values for each mutant show that each residue **X** substitution destabilized the protein by  $\sim 3$  kcal/mol. Mutant **4** showed a  $\Delta\Delta G_{\text{fold}}$  ( $\sim 8$  kcal/mol) greater than what would be predicted from the additive loss of stability based on data obtained for single PEG mutants ( $\sim 6$  kcal/mol). This result suggested synergistic destabilization of the  $\alpha$ -helix by the adjacent PEG segments.

**Table 1.** Thermodynamics of Folding for WT GB1 and PEG Mutants as Measured by CD.

Sequence	$T_m$ (°C)	$\Delta H_{\text{fold}}$ (kcal/mol)	$\Delta T_m$ (°C) <sup>a</sup>	$\Delta\Delta G_{\text{fold}}$ (kcal/mol) <sup>a</sup>
<b>1</b>	$81.4 \pm 0.1$	$-58.4 \pm 1^{95}$	-	-
<b>2</b>	$64.7 \pm 0.1$	-	-16.7	$2.8 \pm 0.1$
<b>3</b>	$61.6 \pm 0.1$	-	-19.8	$3.3 \pm 0.1$
<b>4</b>	$32.4 \pm 0.2$	-	-49.8	$8.1 \pm 0.2$

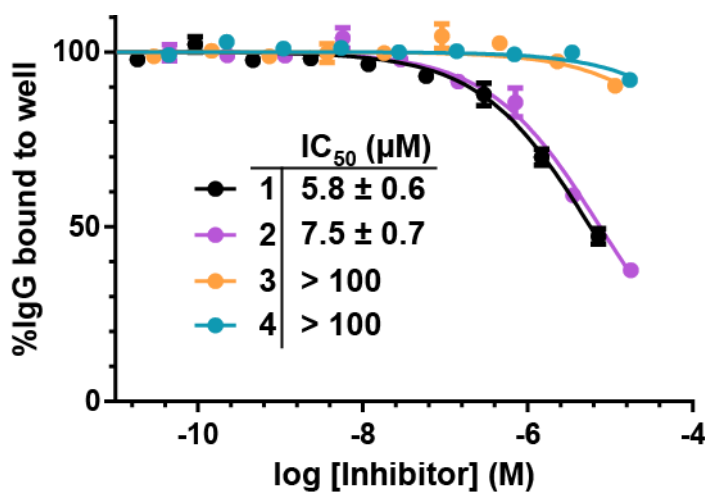
*a.* Relative to **1**.

Prior work has shown that various point mutations in GB1 can lead to changes in oligomerization state, including formation of domain-swapped dimers and tetramers.<sup>80</sup> We performed Gel Permeation Chromatography (GPC) experiments to probe for non-native oligomerization states in mutants **2-4**. In a GPC experiment, proteins are separated on the basis of their three dimensional size under non-denaturing conditions with larger proteins eluting earlier than smaller peptides. GPC chromatograms (Figure 17) for 40  $\mu\text{M}$  samples of proteins **1-4** in phosphate solution pH 7 had identical retention volumes, supporting a monomeric fold.



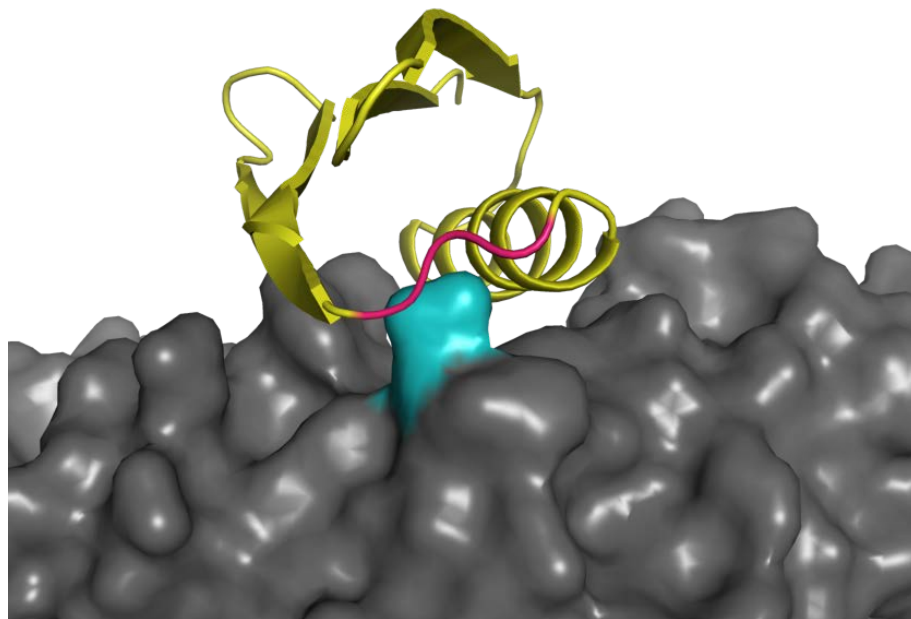
**Figure 17.** GPC chromatograms of 40  $\mu$ M proteins **1-4** in 150 mM sodium chloride, 20 mM phosphate, pH 7 . GPC was performed on a Superdex-75 column.

The strength of protein-protein binding interactions depend critically on the folded structure of the protein. Protein G binds to immunoglobulin G (IgG) as a means of coating the bacterium with host antibodies to evade the host’s immune system. Each domain of protein G, including GB1, is capable of binding IgG antibodies at the F<sub>c</sub> or F<sub>ab</sub> region.<sup>98-99</sup> We reasoned that the strength of IgG binding affinity should be indicative of how well each mutant’s folded structure matched that of the wild-type. Enzyme Linked Immunosorbent Assay (ELISA) experiments were performed for proteins **1-4** to compare binding affinities for IgG conjugated to horseradish peroxidase (HRP) (Figure 18). Varying concentrations of the GB1 mutants were incubated with a fixed concentration of IgG/horseradish peroxidase (HRP) conjugates in microtiter plates coated with protein G. The plates were washed and developed with tetramethylbenzidine (TMB) and 1 M HCl to obtain absorbance of oxidized TMB, indicative of the amount of IgG bound to the wells.



**Figure 18.** Competition ELISA and IC<sub>50</sub> values for proteins **1-4**.

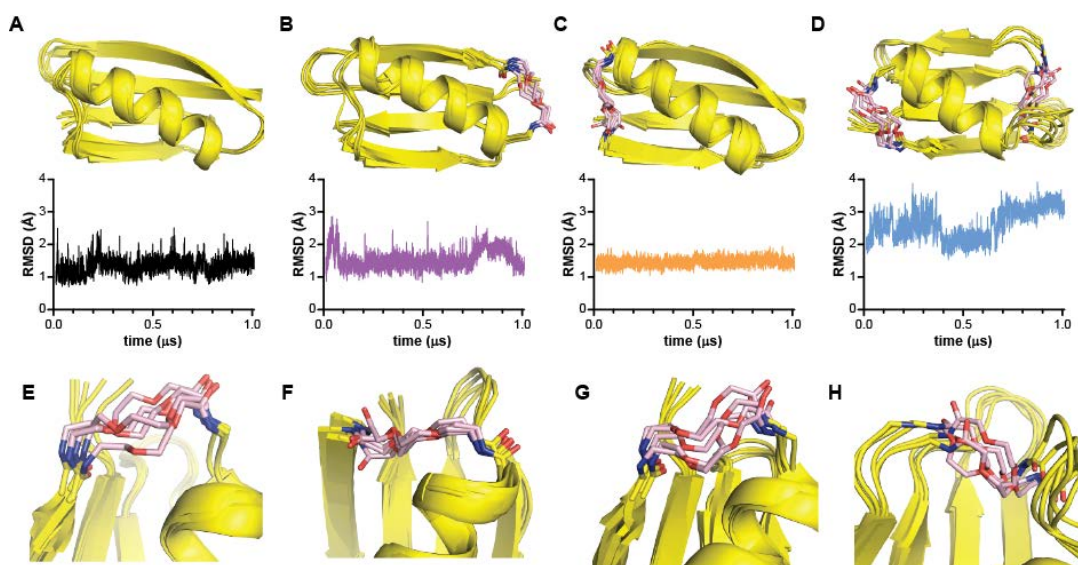
Mutant **2** bound IgG with an affinity indistinguishable from wild-type **1** suggesting similarly folded structures. Conversely, mutants **3** and **4** did not show a measureable affinity for IgG. A crystal structure of the homologous protein GB2 bound to the F<sub>c</sub> domain of IgG shows direct contacts between the C-terminal loop and Asn<sub>434</sub> of the receptor<sup>99</sup> (Figure 19). Furthermore, an 11-residue peptide corresponding to this domain has been shown to effectively compete with protein G among other full-length proteins for binding to IgG, albeit weakly.<sup>100</sup> Based on these precedents, it is not surprising that modification of the C-terminal loop resulted in loss of IgG affinity. Thus, loss of binding affinity of **3** and **4** to IgG does not imply a significant difference in folded structure from proteins **1** and **2**.



**Figure 19.** The crystal structure of GB2 (yellow; C-terminal loop colored magenta) bound to F<sub>c</sub> domain of IgG (gray surface) (PDB: 1FCC).<sup>99</sup>

We sought to obtain high-resolution structures of mutants **2-4** but were unable to crystallize the proteins in numerous buffers. To obtain data bearing on folded structure in the modified backbones, our collaborators (Eli Musselman and Adrian Elcock, University of Iowa) performed all-atom molecular dynamics (MD) simulations based on a crystal structure of the wild-type.<sup>78</sup> Each simulation lasted 1  $\mu$ s in explicit solvent and was performed using standard techniques (see experimental). All proteins remained folded during the simulation (Figure 20). Both single-PEG mutants **2** and **3** retained folds similar to that of the wild-type structure with RMSD values of  $\sim$ 1-2 Å for backbone atoms relative to the initial structure, suggesting **2** and **3** have similar folded structures to **1** which agrees with the CD data and supports the presence of a

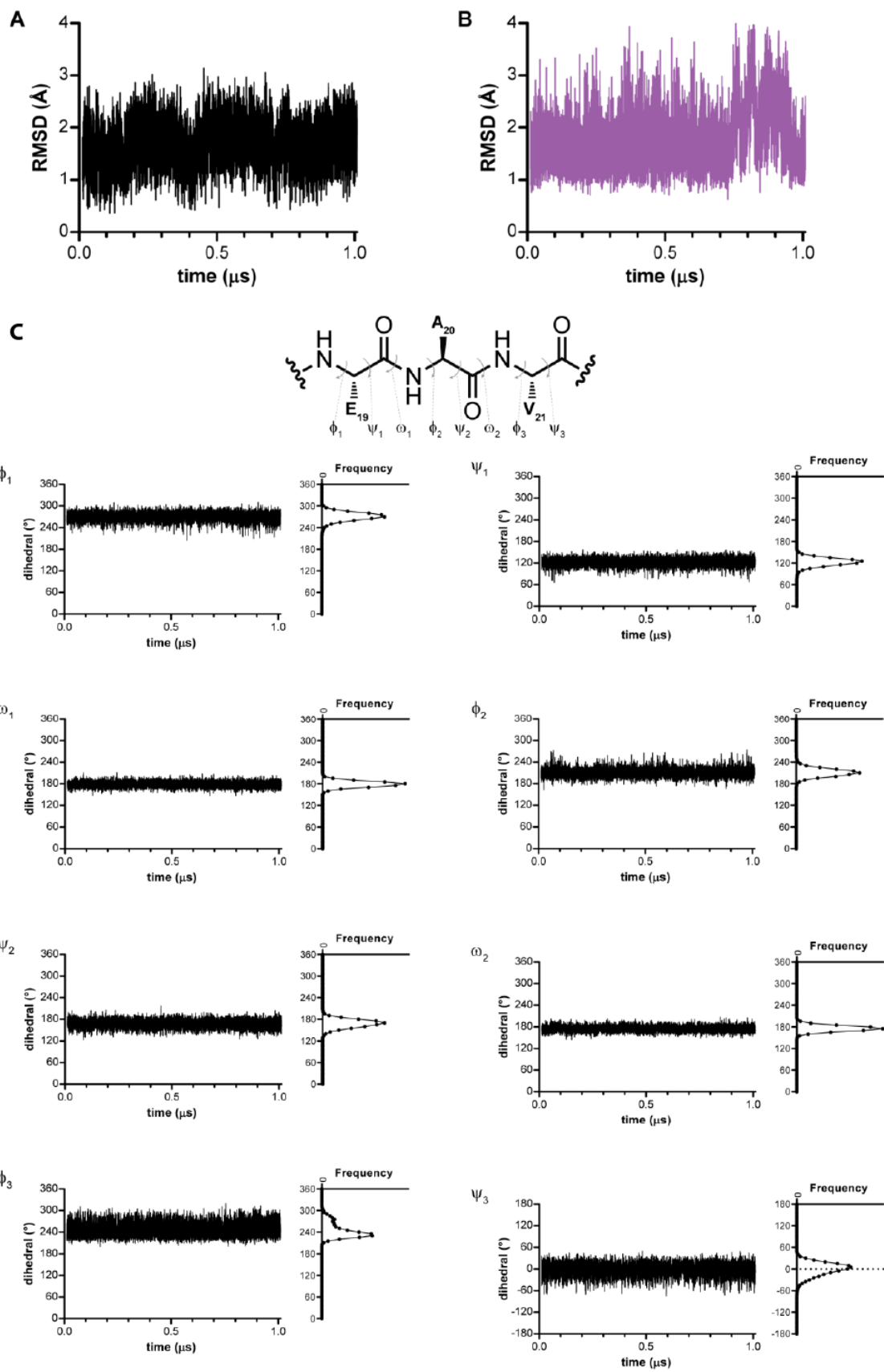
native-like tertiary fold. Double-PEG mutant **4** displayed partial unfolding of the  $\beta$ -sheet and unwinding of the  $\alpha$ -helix, consistent with the conclusions from CD scan and thermal melt data.



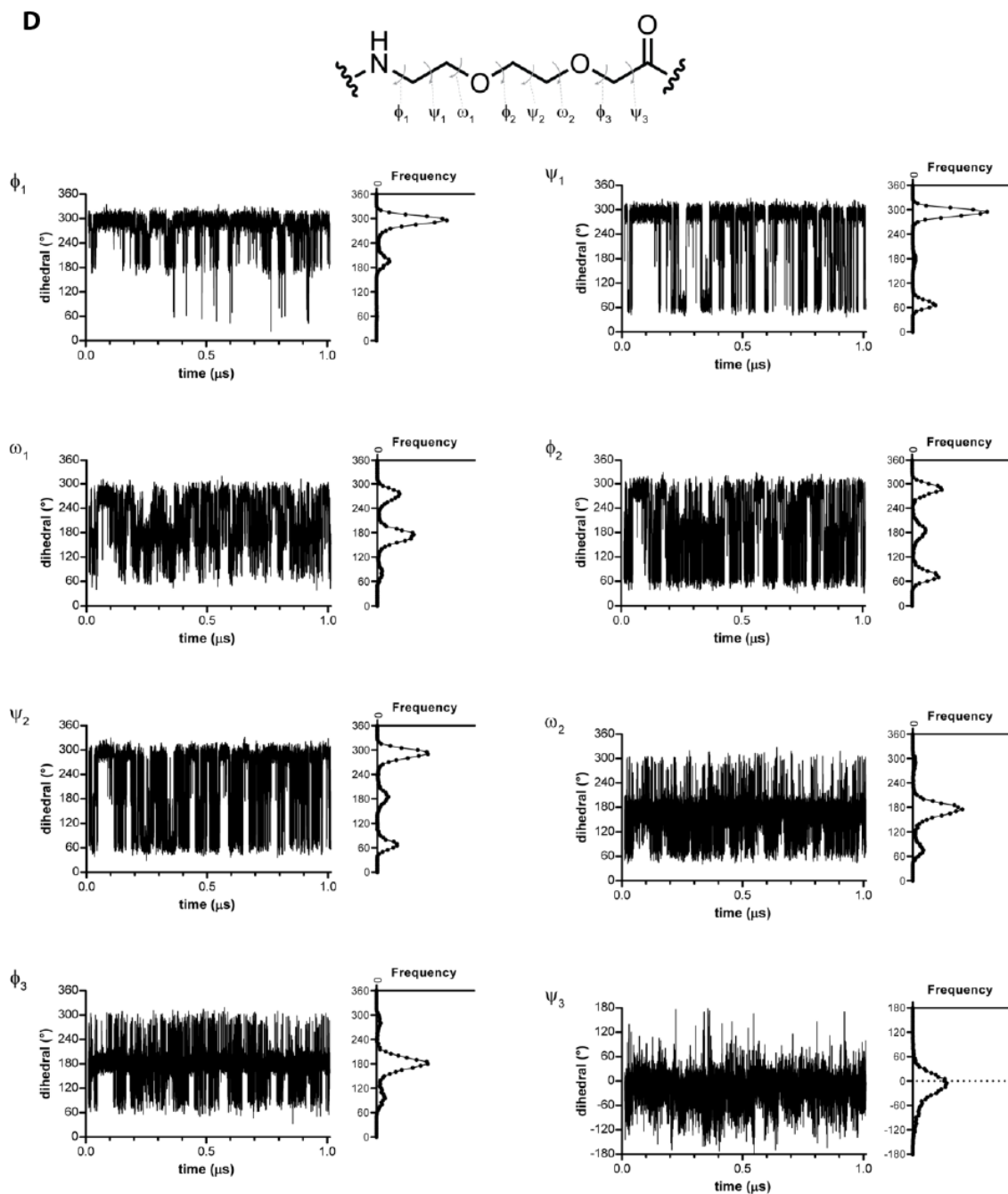
**Figure 20.** MD simulations of proteins **1-4**. **A-D)** Snapshots from MD simulations of proteins **A) 1, B) 2, C) 3** and **D) 4** with RMSD plots for backbone atoms. **E-H)** Close-up views of the PEG segments in proteins **E) 2, F) 3**, and **G-H) 4** (**G** and **H** depict N-terminal and C-terminal loops, respectively). Figure used with permission from reference 91. Copyright 2012 Wiley.

Due to the inherent flexibility of the PEG-substitution, we were interested in the effects of PEG on protein backbone dynamics. Additional analyses were performed on the conformation of residue **X** in mutant **2** and the corresponding N-terminal loop in protein **1**. Despite differences in chemical structure, the overall positions of the backbones of both proteins were not significantly different as shown by their similar RMSD plots (Figure 21). Instead, the PEG backbone showed much more conformational freedom than the natural backbone as evidenced

by the dihedral angles preferences of the loop residues of proteins **1** and **2** (Figure 21). In wild-type **1**, the dihedral angles did not appreciably vary during the simulation, consistent with the low RMSD values of the backbone atoms observed. Residue **X** in mutant **2** exhibited a much wider range of bond dihedrals than the natural backbone, although certain angles were preferred. The RMSD plots and the dihedral angle conformational preferences together suggest that PEG enhances the torsional dynamics of the backbone atoms while retaining the native positions of the atoms. PEG-modified backbones could provide a useful tool to probe the role of backbone torsional freedom in protein-ligand interactions where recent work suggests that a defined degree of disorder can be important in determining protein function.<sup>101-102</sup>



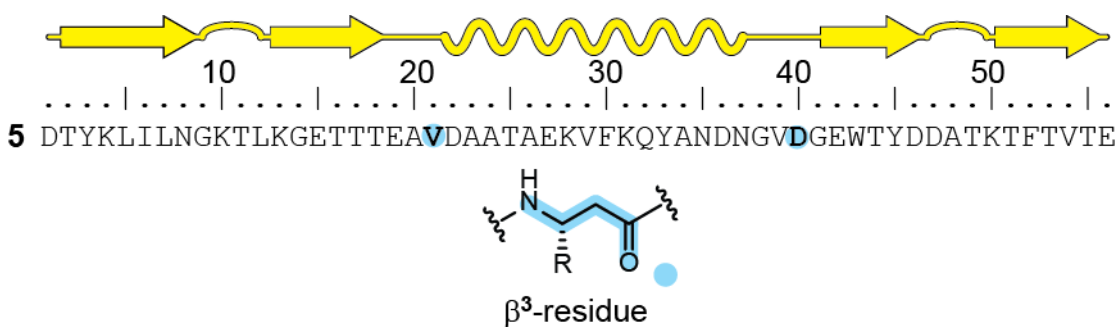




**Figure 21.** RMSD plots for the backbone atoms in the N-terminal loop of A) protein 1 and B) mutant 2. Analysis of the dynamics of the loop dihedrals in C) protein 1 and D) mutant 2. For each dihedral in the residue 19-21 loop (key to angle nomenclature shown in the structure), plots are shown for time-dependent rotation about the bond and the frequency distribution of dihedral values over the course

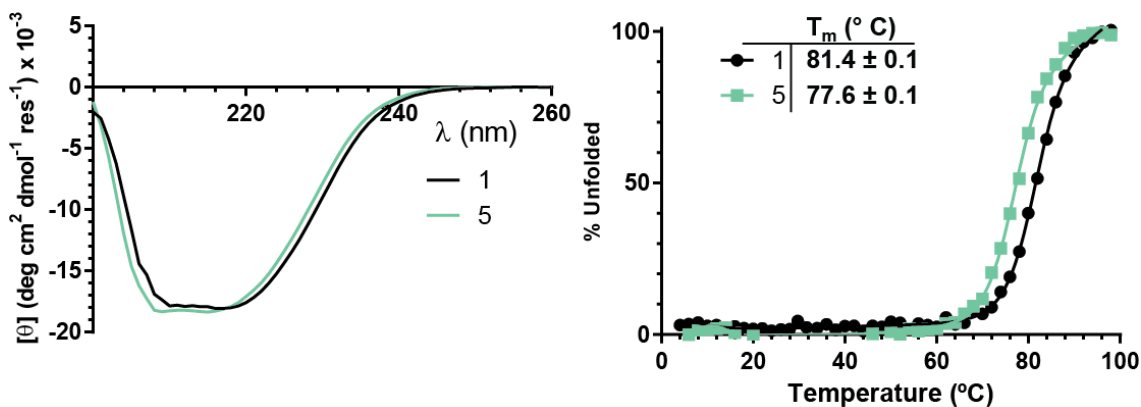
of the 1  $\mu$ s simulation. Dihedrals are plotted either  $0^\circ$  to  $360^\circ$  or  $-180^\circ$  to  $180^\circ$  for clarity. Figure used with permission from reference 91. Copyright 2012 Wiley.

While PEG-derived backbones were tolerated in the loops of GB1, the thermodynamic penalty towards folding was significant ( $\sim 4$  kcal/mol/loop). Given our overarching goal of combining several secondary structure modifications in one sequence, we turned towards more conservative loop replacement strategies. We examined  $\beta^3$ -residues (Figure 22) as an alternative to PEG for loop replacement.  $\beta^3$ -Residues retain the side-chain of the native sequence and have greater backbone conformational freedom relative to  $\alpha$ -residues by virtue of an additional methylene unit. We hypothesized that replacing one residue in each loop with the corresponding  $\beta^3$ -residue would minimally destabilize the overall protein fold.



**Figure 22.** Secondary structure of GB1 and sequence of mutant 5.  $\beta^3$ -residues are highlighted blue and bolded. R is the side-chain group of the corresponding  $\alpha$ -residue.

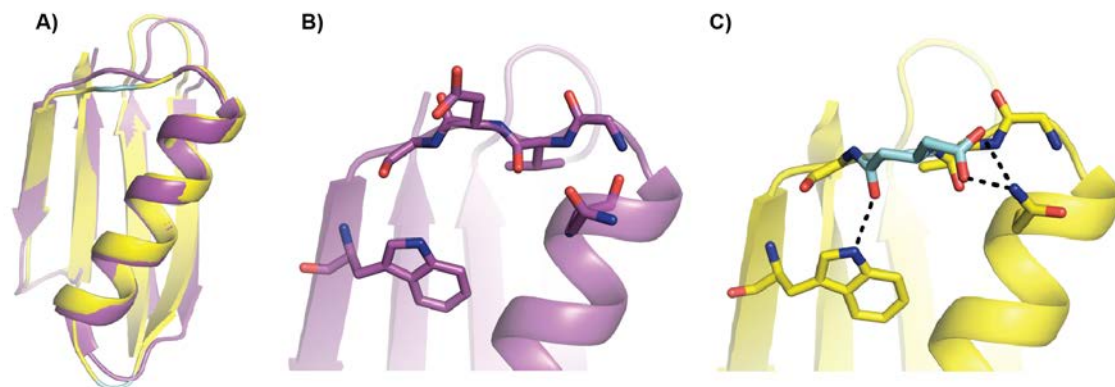
$\beta^3$ -Residues were incorporated into the loops of GB1 (Figure 22). The final residue of each loop was replaced with a corresponding  $\beta^3$ -residue to generate mutant **5**. Mutant **5** showed an almost identical CD scan as that of wild-type **1** (Figure 23). A thermal melt of protein **5** yielded a  $T_m$  of  $77.6 \pm 0.1$  °C and a  $\Delta\Delta G_{\text{fold}}$  of  $0.9 \pm 0.1$  kcal/mol indicating protein **5** is as stable as wild-type **1** within 1 kcal/mol. This small destabilization may arise from the introduction of flexible methylene units in the backbone, increasing the entropy of the unfolded state. This hypothesis is explored more thoroughly in Chapter 3. The enhanced folded stability of **5** over PEG mutants **2-4** confirmed the conservative approach of  $\beta^3$ -residue backbone replacement in GB1 relative to PEG-derived analogs.



**Figure 23.** CD scans at 25 °C and thermal melts of proteins **1** and **5** in 20 mM phosphate buffered water pH 7.

A crystal structure of **5** (Figure 24) was obtained using the hanging drop vapor diffusion method. A small ( $\sim 1$   $\mu\text{L}$ ) volume of a concentrated ( $> 10$  mg/mL) sample of protein is mixed with a small volume of crystallization buffer. The mixture is then suspended over a larger

volume solution of the crystallization buffer in a sealed chamber. Vapor diffusion of water from the protein droplet into the reservoir decreases solubility and drives crystallization of the protein. The crystal was subjected X-ray radiation and diffracted out to 2.00 Å. The diffraction pattern was solved by molecular replacement using the published wild-type GB1 crystal structure (2QMT). The unit cell of the crystal structure of **5** featured two copies of the mutant protein in the asymmetric unit. It is notable that the two chains have slight differences in the N-terminal  $\beta$ -turn with an RMSD value of 0.377 Å. The overlay of the lowest RMSD copy with the wild-type showed a nearly identical tertiary structure, consistent with the CD scan of **5**. The  $\beta^3$ -residues did not alter main chain hydrogen bonds present in the wild-type structure including those of Ala<sub>20</sub> and Glu<sub>19</sub> which participate in the adjacent  $\beta$ -strand. Most of the side-chain polar contacts were retained from the wild-type structure. Interestingly, replacement of Asp<sub>40</sub> with the  $\beta^3$ -homologue caused the carbonyl oxygen of the  $\beta^3$ -residue to participate in a new hydrogen bond with the indole NH of Trp<sub>43</sub>. The side-chain carboxylate also formed a new polar contact with the side-chain amide of Asn<sub>35</sub>. Altogether, the structural data suggested that the additional methylene units in the backbone of the loops are well-accommodated in GB1 and important tertiary contacts in the native structure were maintained in the mutant.



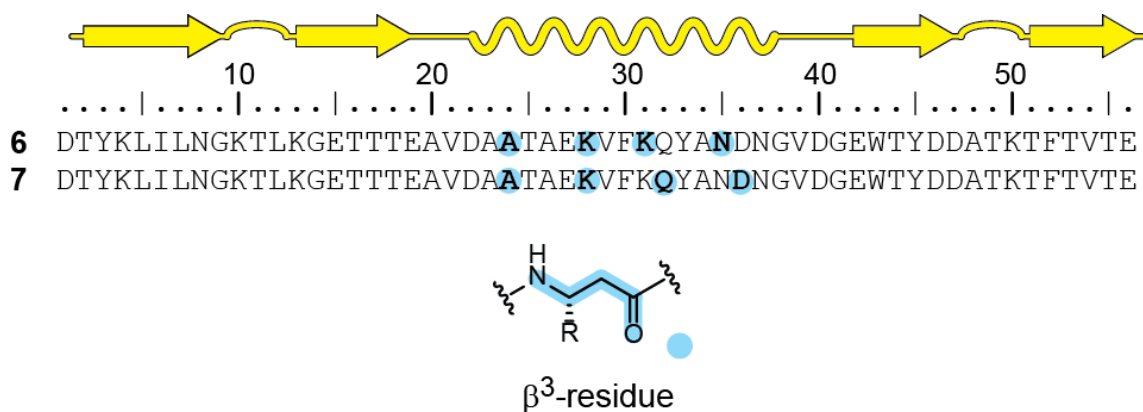
**Figure 24.** A) Protein **5** (yellow) overlaid with wild-type GB1 (purple, PDB: 2QMT). **B, C)** Close ups of C-terminal loop polar contacts (yellow dashed lines) in wild-type GB1 (**B**) and mutant **5** (**C**).  $\beta^3$ -Residues are colored cyan.

Both PEG-based and  $\beta^3$ -residue substitutions have been applied the loops of GB1. The PEG-based residues may be suitable as a tool exploring the role of backbone conformational freedom in protein folding. However, the loss of folded stability from this substitution (~4 kcal/mol for one loop replacement) is too large to be effectively used in concert with strategies for sequence-based modification of other secondary structures.  $\beta^3$ -Residues, like PEG-derived residues, are more flexible than  $\alpha$ -amino acids but their incorporation into the loops of GB1 is less detrimental to protein folding and should be more suitable to use with other unnatural amino acids for wholesale tertiary structure replacement of GB1.

### 2.2.2 Helix

The  $\alpha$ -helix is the most common secondary structure element of proteins and present at many protein-protein interfaces.<sup>103</sup> The vast majority of foldamer research has been aimed at

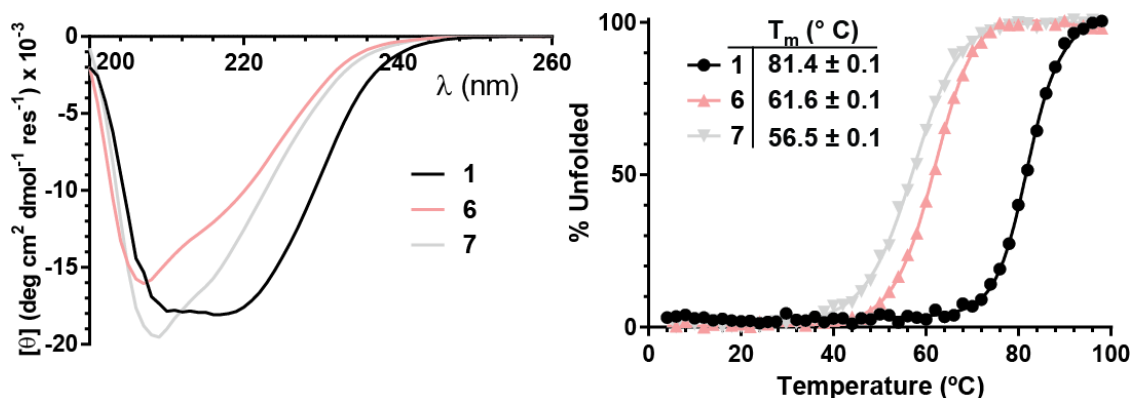
mimicking  $\alpha$ -helical folds with both *de novo*<sup>59, 104-105</sup> and sequence-based design approaches.<sup>70,</sup>  
<sup>106</sup> Several classes of unnatural residues have shown strong preferences for forming helices in solution including  $C_{\alpha}$ -methyl-residues,<sup>35</sup>  $\beta$ -residues,<sup>49</sup>  $\gamma$ -residues,<sup>107</sup> peptoids,<sup>108</sup> and aromatic residues.<sup>59</sup> Prior studies have shown that  $\alpha/\beta^3$ -peptides derived from  $\alpha$ -helix-forming sequences can fold into  $\alpha$ -helices with similar structure to the native  $\alpha$ -peptide in aqueous solution<sup>70</sup> and retain biological function.<sup>71, 106, 109</sup>



**Figure 25.** Sequences of GB1 helix mutants **6** and **7**. R is the side-chain group of the corresponding  $\alpha$ -residue.

We designed two GB1 sequences to examine the impact of two different  $\alpha/\beta^3$ -substitution patterns,  $\alpha\alpha\beta\alpha\alpha\beta$  (**6**) and  $\alpha\alpha\alpha\beta$  (**7**) (Figure 25) in a tertiary structure context; these patterns previously led to effective structural and biological mimics in BH3 domain-derived helical peptides.<sup>106</sup> Both templates were designed to place  $\beta$ -residues distant from the hydrophobic core of GB1 as well as surfaces involved in protein-receptor binding. The  $\alpha\alpha\beta\alpha\alpha\beta$  pattern was expected to display the backbone methylene units as a stripe opposite of the hydrophobic core of

GB1. In contrast, the  $\alpha\alpha\alpha\beta$  pattern should arrange the backbone methylene units as a rotated stripe along the  $\alpha$ -helix.



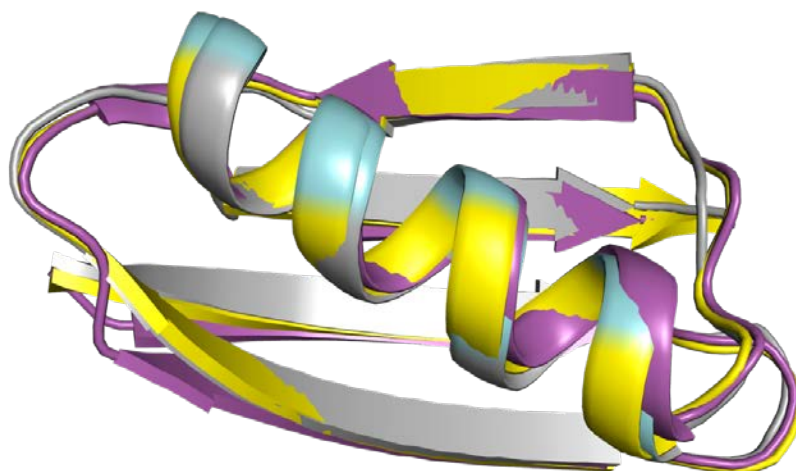
**Figure 26.** CD scans at 25 °C and thermal melts of 40  $\mu$ M wild-type **1** and mutants **6** and **7** in 20 mM phosphate buffered water pH 7.

The CD scans of **6** and **7** are significantly different in shape and magnitude from that of wild-type **1** (Figure 26). The minimum at 209 nm is consistent with minima observed for other  $\alpha/\beta$  helices.<sup>110</sup> The notably larger magnitude of the molar ellipticity of **7** compared to **6** could suggest that the  $\alpha\alpha\alpha\beta$  pattern leads to a greater helical character in GB1 than the  $\alpha\beta\alpha\alpha\beta$  pattern. However, changes to the peptide backbone have resulted in altered CD spectra relative to natural proteins and as such, CD spectra of unnatural backbones may not accurately reflect the extent of folded structure in solution. Interestingly, thermal melts reveal a lower  $T_m$  for **7**, corresponding to a lower folded stability than mutant **6** (Table 2).

**Table 2.** Melting Temperatures and Changes to Free Energy of Folding for Helix Mutants **6-7**.

Sequence	$T_m$ (°C)	$\Delta T_m$ (°C) <sup>a</sup>	$\Delta\Delta G_{\text{fold}}$ (kcal/mol) <sup>a</sup>
<b>6</b>	$61.6 \pm 0.1$	-19.8	$3.5 \pm 0.1$
<b>7</b>	$56.5 \pm 0.1$	-24.9	$4.1 \pm 0.1$

*a.* Relative to **1**.



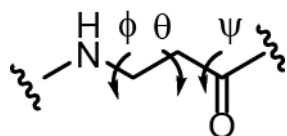
**Figure 27.** Overlay of wild-type GB1 structure (2QMT) with refined crystal structures of **6** (yellow) and **7** (gray).  $\beta^3$ -Residues are colored cyan.

Diffraction quality crystals of **6** and **7** were obtained by hanging-drop vapor diffusion and used to determine high-resolution crystal structures (1.95 Å and 2.20 Å, respectively, Figure 27). Each mutant exhibits high structural homology to the wild-type crystal structure including side-chain display in the helices. Both substitution patterns displayed the  $\beta$ -residues as expected from our design. The  $\beta$ -residues in **6** are aligned on the side of the helix distal from the hydrophobic core, while in **7** they form a rotated stripe about the helix. The average backbone dihedrals of  $\beta^3$ -residues ( $\phi$ ,  $\theta$ , and  $\psi$ ) in both mutants match closely with previously published structures of



helices containing  $\beta^3$ -residues<sup>70</sup> (Table 3). Many native tertiary interactions are preserved in the helix-modified mutants, even when the interaction involves a  $\beta^3$ -residue (Figure 28). RMSD values for **6** and **7** (0.529 Å and 0.927 Å, respectively) relative to wild-type crystal structure suggest that the fold of **6** is closer the native fold than that of **7**. Additional structural parameters of the helices of **1**, **6**, and **7** were analyzed, but no significant differences in measured values were observed (Table 3).

**Table 3.** Comparison of Helical Parameters of  $\alpha/\beta$  Helices and Wild-type GB1 Helix.

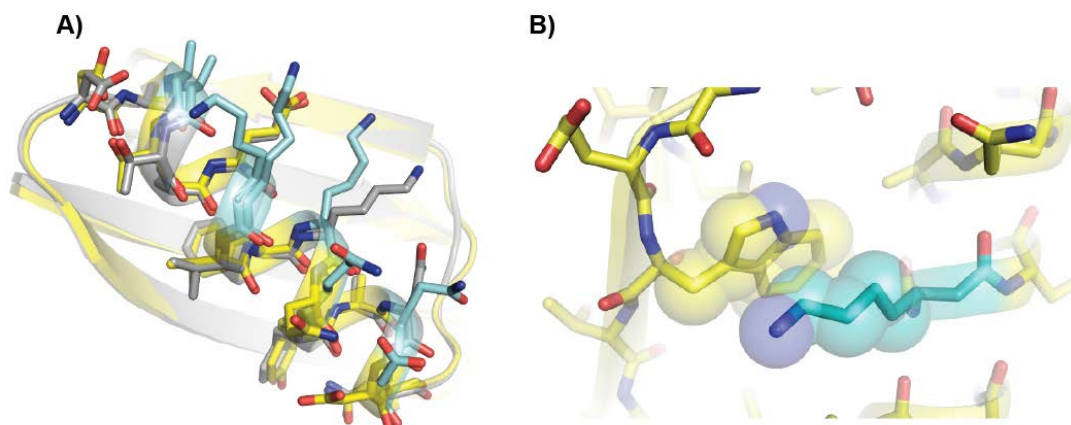


Protein	Phi (°)	Theta (°)	Psi (°)	Radius <sup>a</sup> (Å)	Rise/Residue <sup>a</sup> (Å)	Pitch <sup>b</sup> (Å)
<b>1 (2QMT)</b> <sup>78</sup>	-	-	-	2.3	1.5	6
<b>6</b>	-106	82	-114	2.4	1.5	5
<b>7</b>	-108	85	-111	2.4	1.5	6
<b>2OXJ</b> <sup>70c</sup>	-116	84	-109	2.5	1.5	5

*a.* Calculated using HELANAL software.<sup>111</sup>

*b.* Calculated using PS software.<sup>112</sup>

*c.*  $\alpha/\beta$ -helix from coiled-coil.<sup>70</sup>



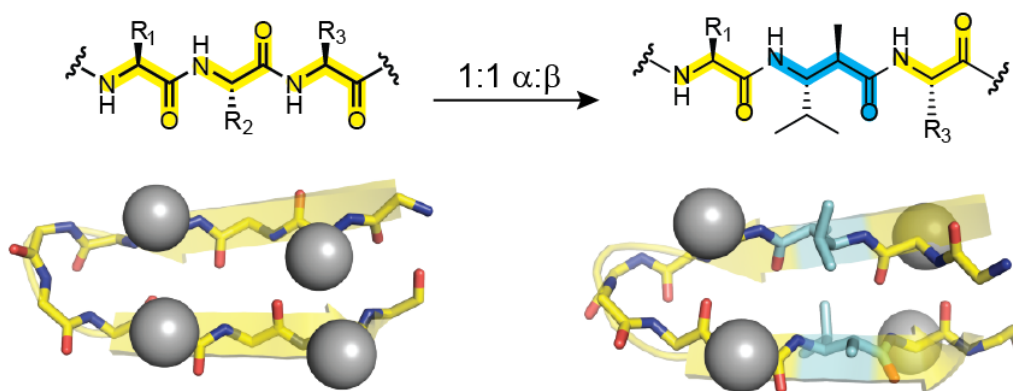
**Figure 28.** **A)** Overlay of helix backbone and side-chains in crystal structures of **6** (yellow) and **7** (gray). **B)** Tertiary interaction in the structure of **6** of side-chain from  $\beta^3$ -Lys<sub>31</sub> with side-chain of Trp<sub>43</sub>.

### 2.2.3 Sheet

$\beta$ -Sheet motifs are ubiquitous in proteins, yet their structural mimicry by foldamers is much less understood relative to  $\alpha$ -helices. One explanation could be the larger entropic barrier to folding in  $\beta$ -sheets; sheet formation requires the precise arrangement of two (or more)  $\beta$ -strands such that sufficient hydrogen bonds are formed and hydrophobic side-chains are buried. Pioneering studies showed that small cyclic  $\beta$ -residues formed sheet-like structures or fibrils in organic solvents.<sup>113-115</sup> Few studies have examined sheet mimicry in the more challenging folding environment of aqueous solvents.<sup>116-118</sup>

Prior work in our lab focused on developing strategies for  $\beta$ -sheet mimicry on unnatural backbones. Early designs using a small model hairpin peptide focused on exploring mixed  $\alpha/\beta$ -backbones to probe the ability of  $\beta$ -residues to influence sheet formation in buffered water.<sup>119</sup>

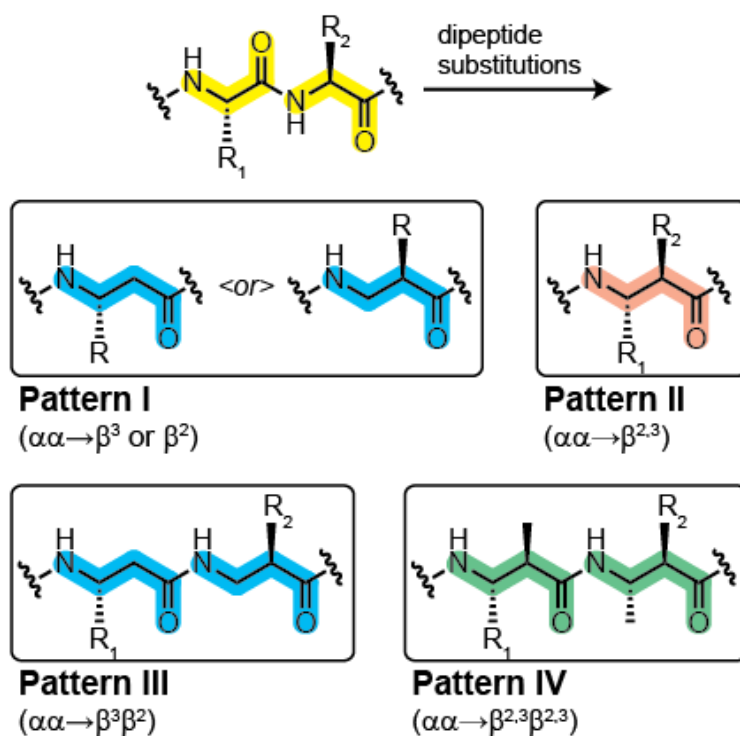
Two central  $\alpha$ -residues in the model hairpin sequence were systematically replaced in a 1:1 fashion by sixteen different  $\beta$ -residues. The selection of  $\beta$ -residues examined how side-chain position ( $\beta^2$  vs  $\beta^3$  vs  $\beta^{2,3}$ ), stereochemistry, and backbone cyclization affected the propensity of the peptide to fold into a hairpin. Of the 16 monomers explored, only three were capable of hairpin formation in mixed  $\alpha/\beta$ -peptides. It was further shown by 2D-NMR analysis that only two peptides, containing either enantiomer of a  $\beta^{2,3}$ -residue, had native hairpin-like folds in aqueous solution. Unfortunately, in all cases, the 1:1  $\alpha/\beta$  substitution led to an inversion of the backbone amide as well as the side-chain (Figure 29), making it unsuitable for sheet modifications in tertiary structures.



**Figure 29.** Change in side-chain display in hairpin peptide as a result of 1:1  $\alpha/\beta$  substitution.  $\beta$ -Residues are colored cyan and side-chain atoms of key hydrophobic core residues are shown as gray spheres. Data from reference 119.

To circumvent the unfavorable inversions introduced by 1:1  $\alpha/\beta$  substitutions, our lab previously explored four alternate strategies in a longer model hairpin peptide derived from the

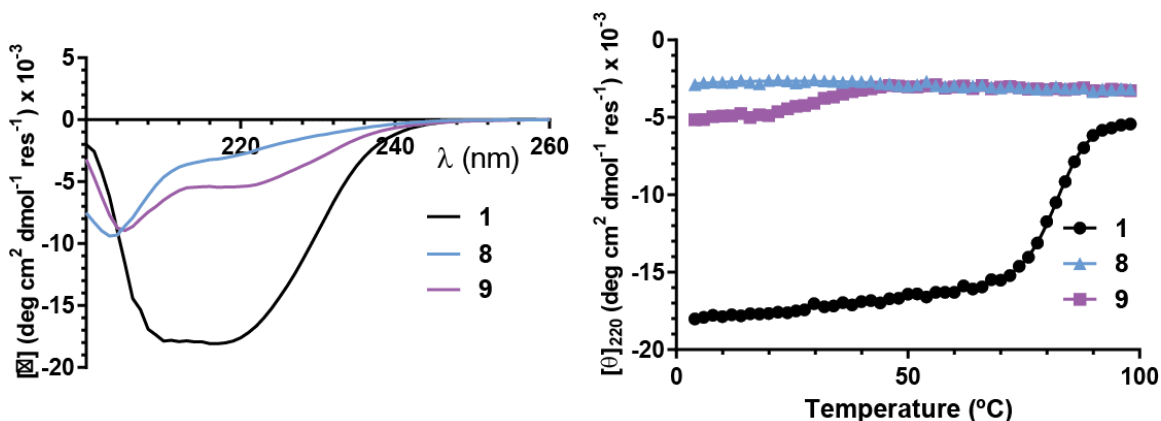
C-terminal hairpin of GB1:  $\alpha\alpha \rightarrow \beta^2$  or  $\beta^3$  (I),  $\alpha\alpha \rightarrow \beta^{2,3}$  (II),  $\alpha\alpha \rightarrow \beta^3\text{-}\beta^2$  (III), and  $\alpha\alpha \rightarrow \beta^{2,3}\text{-}\beta^{2,3}$  (IV). These designs examined the role of two orthogonal variables on sheet backbone modification: 1) backbone length relative to the replaced  $\alpha\alpha$  segment and 2) backbone preorganization of  $\beta$ -residues (Figure 30).<sup>120</sup> Designs I and II are 2:1  $\alpha$ : $\beta$  substitutions whereas III and IV are 2:2  $\alpha$ : $\beta$  replacements. Of the designs investigated, I, II, and IV led to  $\alpha$ / $\beta$ -peptides with hairpin folds by 2D NMR analysis. Peptides based on Design III likely failed to fold as designed because of the significant increase in backbone flexibility introduced by a 2:2  $\alpha$ : $\beta$  substitution. Relative to the all  $\alpha$ -residue model sequence, each design introduced an approximately 0.5 kcal/mol energetic penalty towards folding per substitution. Despite this destabilization, NMR solution structures confirmed the presence of a hairpin-like fold in the  $\alpha$ / $\beta$  peptides.



**Figure 30.** Dipeptide substitution designs for hairpin mimicry in mixed  $\alpha/\beta$ -peptides. R is the side-chain group(s) of the corresponding  $\alpha$ -residue(s). Adapted with permission from reference 120.

Having demonstrated that  $\alpha/\beta$  oligomers based on a model hairpin sequence can fold akin to that of the native peptide, we set our sights on modifying the sheet of GB1. Sheet mutants **8** and **9** were synthesized and examined for their ability to retain the tertiary fold of GB1 (Figure 31). **8** utilizes a 2:1  $\alpha:\beta$  replacement strategy from Design I while a 2:2  $\alpha:\beta$  akin to that of Design III was used in **9**. Synthetic difficulties associated with monomer synthesis precluded the investigations of  $\beta^{2,3}$  residues in GB1.

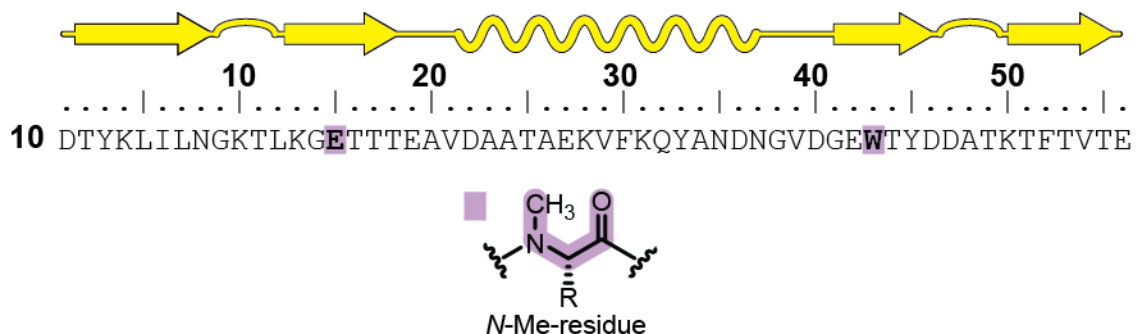




**Figure 32.** CD scans at 25 °C and thermal melts of 40  $\mu$ M GB1 sheet mutants **8** and **9** in 20 mM phosphate buffered water, pH 7.

*N*-methyl- $\alpha$ -residues, methylated variants of  $\alpha$ -amino acids at the backbone nitrogen atom, have been used in small sheet peptides to study aggregation prone sequences especially those from neuronal<sup>121</sup> and pancreatic disorders.<sup>122</sup> *N*-methyl residues are advantageous in these systems because of the replacement of a backbone hydrogen bond donor with a methyl group that disfavors edge to edge association of strands. We hypothesized that these residues would be suitable for sheet backbone modification in GB1 because relative to  $\beta$ -residues, *N*-methyl residue incorporation does not increase the flexibility or length of the backbone, an important consideration for maintaining tertiary interactions in the native fold.

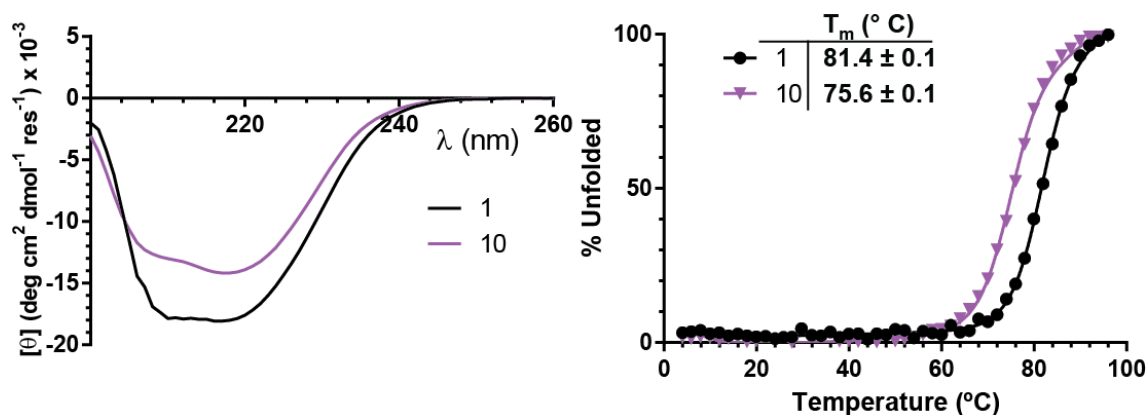
Two *N*-methyl residues were introduced into the central residue of the outer strands of GB1 to generate mutant **10** (Figure 33). These locations were chosen to avoid disruption of interstrand hydrogen bonds necessary for sheet formation. From modelling analysis based on the wild-type crystal structure (2QMT), the unnatural *N*-methyl groups should extend towards solvent and not disrupt the fold of GB1.



**Figure 33.** Sequence of *N*-Methyl GB1 sheet mutant **10**. R is the side-chain group of the corresponding  $\alpha$ -residue.

As expected, a CD scan of **10** shows a native-like spectrum with a somewhat lower magnitude at 220 nm, perhaps indicative of reduced sheet content (Figure 34). Thermal melts of **10** yield a  $T_m$  of 75.6 °C resulting in a  $\Delta\Delta G_{\text{fold}}$  of 0.9 kcal/mol relative to **1**. Given the conservative nature of the modification, the slightly destabilized folded state of **10** is unanticipated but may be a result of tertiary amide *cis/trans* isomerization or changes in solvation of the protein brought about by the exposed and hydrophobic *N*-methyl groups. The difference in free energy for *cis* and *trans* conformations of tertiary amides is much smaller than in secondary amides and many tertiary amides have been shown to display both conformations in proteins.<sup>123-124</sup>



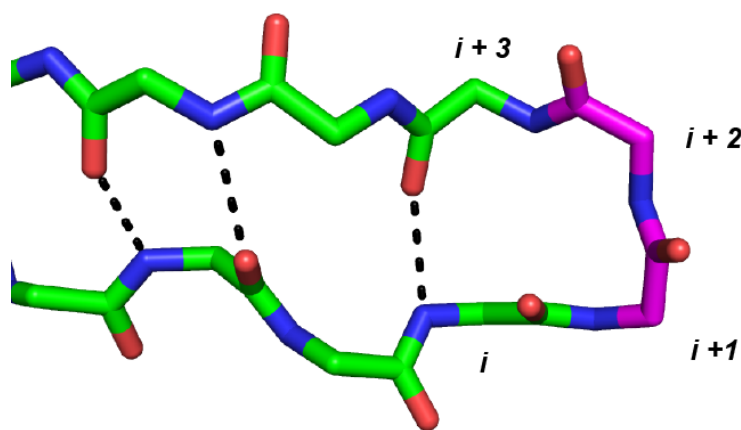


**Figure 34.** CD scans at 25 °C and thermal melts of 40  $\mu$ M wild-type **1** and mutant **10** in 20 mM phosphate buffered water pH 7.

## 2.2.4 Turns

Protein tertiary folding typically requires a reversal of backbone direction at several points in the sequence. This reversal is often realized by the  $\beta$ -turn, and  $\beta$ -turn mimics have been extensively studied. For clarity, a  $\beta$ -loop is defined herein as a group of adjacent residues in a protein or peptide that contain the  $\beta$ -turn residues as well as an N-H to O=C intramolecular hydrogen bond between the first ( $i$ ) and last ( $i + n$ ) residue (Figure 35). The  $\beta$ -turn residues are the residues between the first and last residue of a  $\beta$ -loop ( $i + 1$  to  $i + n - 1$ ). The turn residues reverse the direction of the polypeptide chain and typically do not have any intramolecular backbone hydrogen bonds. The canonical  $\beta$ -loop spans four residues ( $n = 3$ ) and is stabilized by an intramolecular hydrogen bond between residues  $i$  and  $i + 3$  that additionally serves as the first hydrogen bond of the  $\beta$ -sheet.  $\beta$ -turns are classified by the  $\phi$  and  $\psi$  bond dihedrals of residues  $i +$

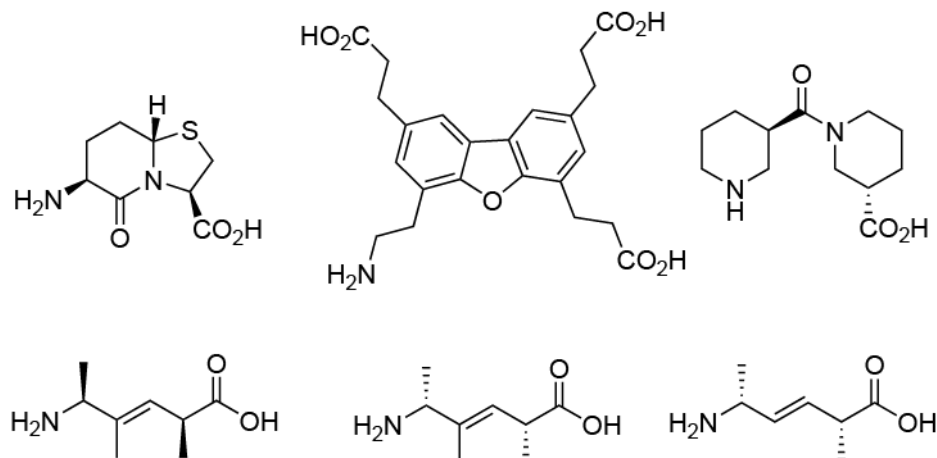
1 and  $i + 2$  and fall into one of nine turn types,<sup>125</sup> the most prevalent of which are I, II, I', and II'. Types I and II differ from I' and II' in the signs of the  $\phi$  and  $\psi$  bond dihedrals.



**Figure 35.** Example of a canonical four-residue  $\beta$ -loop,  $i - i + 3$ .  $\beta$ -Turn residues are colored magenta. Side-chains are omitted for clarity. Dotted black lines indicate interstrand hydrogen bonds.  $\beta$ -Loop adapted from GB1 (PDB: 2QMT).

$\beta$ -Turn mimics have been developed to explore the kinetics of  $\beta$ -sheet nucleation and control the propensity to promote turns in model peptides.<sup>64, 67, 69, 126</sup> Prior work investigated the nucleation capacity of six different  $\beta$ -turn mimics in variants of a Pin1 WW domain model peptide (Figure 36).<sup>69</sup> The study found that the fastest nucleators were those that matched the preferred dihedral angles of the model peptides turn type. In separate work, a dibenzofuran-based turn mimic was incorporated into the N-terminal  $\beta$  loop of GB1;<sup>65</sup> the resulting mutant had nearly identical folded stability as the wild-type protein. 1,2,3-triazoles have also been explored as  $\beta$ -turn mimics.<sup>54, 127</sup> In a larger protein context, 1,4 and 1,5 disubstituted 1,2,3-triazoles were

explored as *cis*-peptide bond mimics in a  $\beta$ -turn of RNase A.<sup>54</sup> 1,5 disubstituted triazoles were shown to be superior than both 1,4 triazoles and the native backbone at recapitulating *cis*-peptide bonds as shown by the higher catalytic activity of the mutant enzyme.

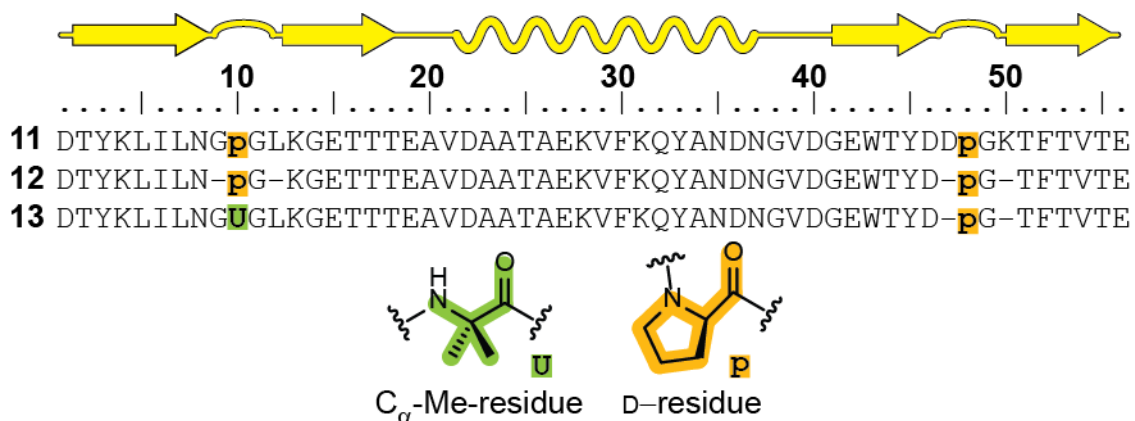


**Figure 36.** Examples of several  $\beta$ -turn mimics studied in the literature.

As closer analogs of  $\alpha$ -amino acids, D-amino acids have been used as turn inducers in short peptides. D-Pro-Gly segments in two-residue  $\beta$ -loops have been shown to strongly promote  $\beta$ -hairpin structures in small peptides.<sup>62, 128-129</sup> In these two-residue loops, types I' and II' are favored for  $\beta$ -hairpin formation over other turn types. In contrast, four-residue turns often fall under Type I or II; surveys of four-residue turns reveal a high frequency for L-proline at position *i* of the turn.<sup>130-131</sup>

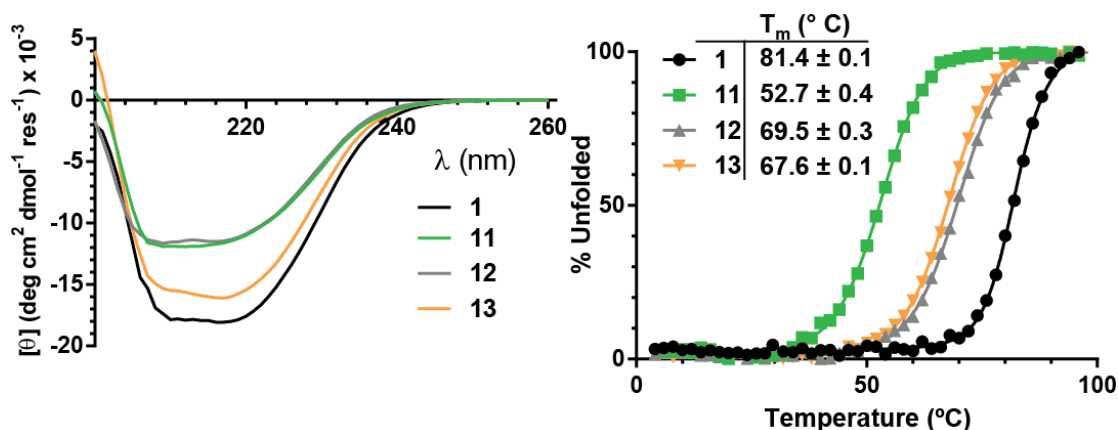
$C_{\alpha}$ -methyl residues such as aminoisobutyric acid (Aib) are known to be strong helix promoters in short peptides and proteins.<sup>35, 132</sup> This is due to the restricted Ramachandran space of  $C_{\alpha}$ -methyl residues resulting from the sterics of the  $\alpha$ -methyl group. The backbone

conformational preferences permits far fewer structures than  $\alpha$ -residues. It has been shown that an Aib-Gly dipeptide segment can induce turns in short hairpin peptides in water with a similar efficacy as D-Pro-Gly.<sup>68</sup>



**Figure 37.** Sequences of turn mutants **11-13**.

We designed three proteins, **11**, **12**, and **13** (Figure 37) to examine the effects of turn replacement in GB1. Protein **11** replaces the central two residues of each  $\beta$ -loop with D-Pro-Gly. In **12**, four-residues of the native turns are entirely replaced with D-Pro-Gly to create two type I'/II' two-residue  $\beta$ -turns. In modifying each  $\beta$ -turn in GB1 to be a two-residue turn, a mirror image turn promoter was hypothesized to be more favorable than other turn type promoters. Mutant **13** retains the C-terminal turn's modification from **12** but uses an alternate strategy for the N-terminal turn replacement; the two turn residues are replaced with an Aib-Gly segment. As Aib-Gly is an achiral segment, it may be better accommodated in GB1 as opposed to the chiral D-Pro-Gly turn by enacting less of a chiral preference on the local folding environment.

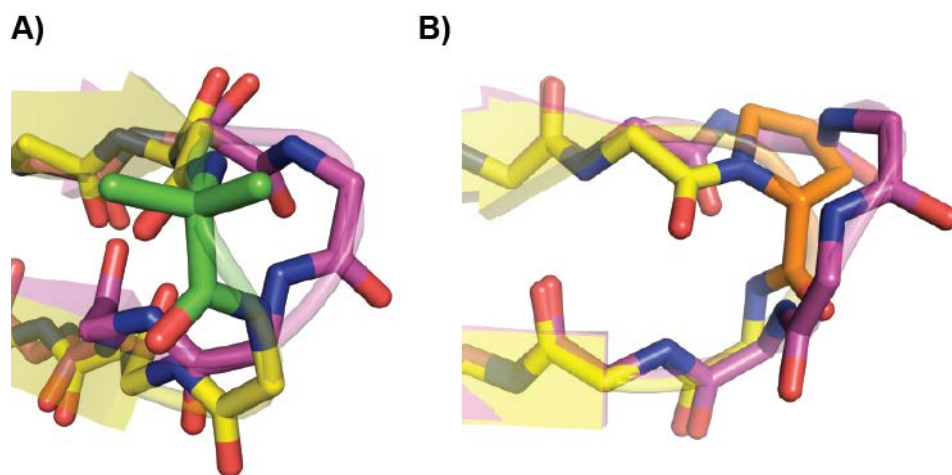


**Figure 38.** CD scans and melts of 40  $\mu$ M turn mutants **11-13** in phosphate buffered water pH 7.

CD spectra of **11**, **12**, and **13** showed similar curve shape yet reduced minima intensity of both proteins relative to the wild-type **1** (Figure 38), possibly suggesting the presence of a less-prominent tertiary structure. Hybrid turn mutant **13** displayed an absorbance with a shape and magnitude closer to that of **1**. Thermal melts of mutants **11**, **12**, and **13** suggested that the D-Pro-Gly segment strongly destabilized the type I turn of the N-terminal hairpin while promoting type I'/II' turns as mutant **12** exhibited higher thermal stability than **11**. Despite having two fewer residues, the shortened C-terminal turns appeared to be well tolerated in GB1 as shown by the higher folded stabilities of **12** and **13** relative to **11**. In contrast, modification of the N-terminal turn with either Aib-Gly or D-Pro-Gly was unfavorable, perhaps due to a native turn type mismatch. It was possible that replacing the four-residue  $\beta$ -loop of the N-terminus as in **12** and **13** resulted in removal of the first hydrogen bond of the  $\beta$ -hairpin, causing a change in the

hydrogen bond register of the  $\beta$ -hairpin. The hydrogen bond register is important to  $\beta$ -sheet formation and changes to the native register may disrupt  $\beta$ -sheet formation.

Only **13** was able to be crystallized and diffracted to give a high resolution structure (2.00 Å). The structure shows a native-like tertiary fold in close agreement to wild-type GB1 with a  $C_{\alpha}$  RMSD of 0.45 Å. As hypothesized, the N-terminal turn shows a turn type mismatch that likely reduced the thermal stability of **13** (Figure 39). In one of the chains of the asymmetric unit, the N-terminal turn appears to more disordered as revealed by locally elevated B-factors. Overlays of the C-terminal turns of **1** and **13** supported the observation from CD melts that the shortened turn is well accommodated and does not significantly alter the position of the  $\beta$ -strands that form part of the central  $\beta$ -sheet.



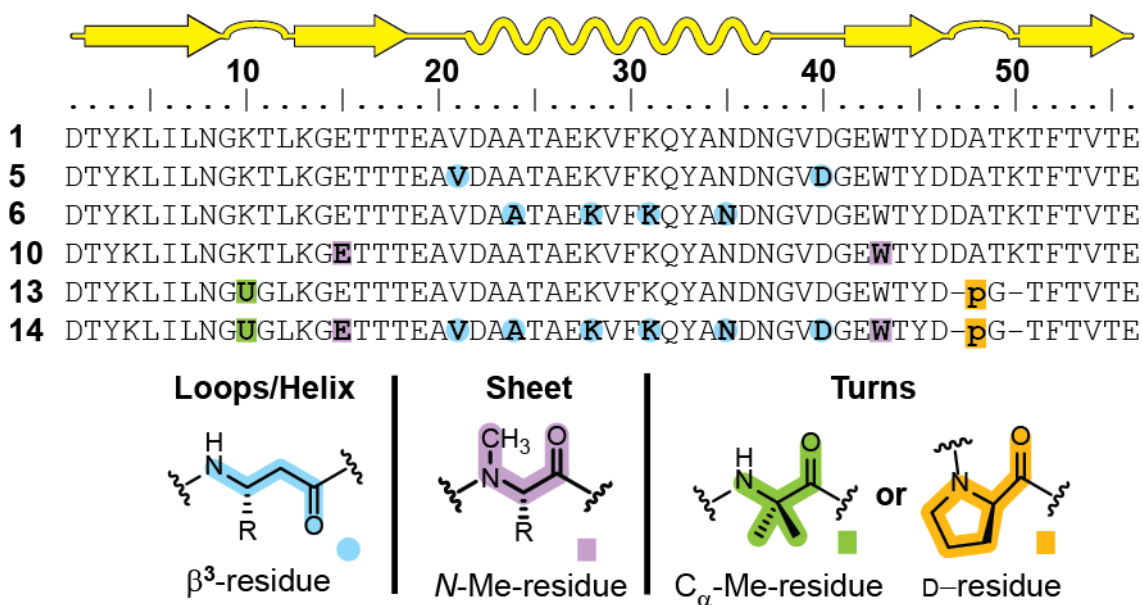
**Figure 39.** Close-up of **A)** N-terminal and **B)** C-terminal turns from **13** (yellow) overlaid with those from wild-type GB1 (magenta) (2QMT).

## 2.3 COMBINED BACKBONE MUTATIONS

### 2.3.1 “All” Mutant

As described throughout the chapter, we systematically explored backbone alterations in the secondary structure elements of GB1 through several unnatural residue types. Although every modification was destabilizing to the tertiary fold, nearly all mutants had evidence of native-like folding behavior through CD spectra and/or X-ray crystallography. We were curious as to how the individual secondary structure mutations would affect the folding thermodynamics of GB1 when combined in one sequence.

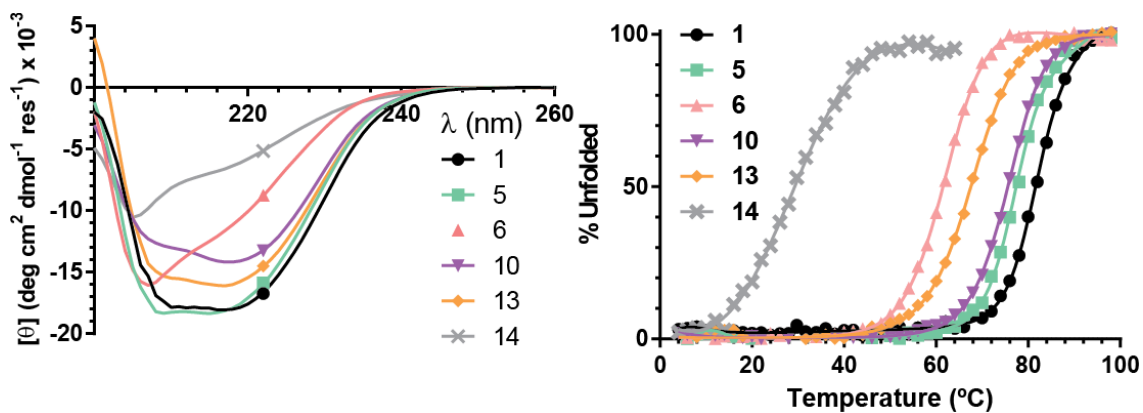
The backbone sequences of the mutants with the highest  $T_m$ , approximately equivalent to the most stable fold, were combined into the design of All mutant **14** (Figure 40). For the turn mutants, the sequence of **13** was selected due to its ability to yield diffraction quality crystals and confirmed native-like three dimensional fold. **14** bears the side-chain sequence of wild-type GB1 on a mixed backbone composed of ~ 20 % unnatural residues.



**Figure 40.** Sequences of wild-type GB1 **1** and mutants used to design All mutant **14**. R is the side-chain group of the corresponding  $\alpha$ -residue.

All mutant **14** had characteristic signatures of its constitutive design elements as observed by CD: an  $\alpha/\beta$ -helix (minima at 209 nm) and wild-type GB1 sheet (broad absorbance centered on 218 nm) (Figure 41). Assessment of folded structure by CD scans alone, however, is not reliable. Stronger evidence for a tertiary fold in **14** came from a thermal melt. The denaturation curve of the All mutant was two-state and cooperative. The low  $T_m$  of 31 °C translated to an approximate  $\Delta\Delta G_{\text{fold}}$  of 8 kcal/mol relative to wild-type GB1 (Table 4). This high destabilization closely matched the sum of the energetic penalties of each individual secondary structure modification, suggesting that the combination of several different strategies for backbone alteration in proteins could lead to foldamers with predictable stabilities.





**Figure 41.** CD scans at 25 °C and thermal melts of 40  $\mu$ M wild-type **1** and mutant **14** in 20 mM phosphate buffered water pH 7.

**Table 4.** Summary of Folding Thermodynamics for Wild-type GB1 and Mutants.

Sequence	2° modified	$T_m$ (°C)	$\Delta T_m$ (°C) <sup>a</sup>	$\Delta\Delta G_{fold}$ (kcal/mol) <sup>a</sup>
<b>1</b>	-	$81.4 \pm 0.1$	-	-
<b>5</b>	loops	$77.6 \pm 0.1$	-3.8	$0.6 \pm 0.1$
<b>6</b>	helix	$61.6 \pm 0.1$	-19.8	$3.2 \pm 0.1$
<b>10</b>	sheet	$75.6 \pm 0.1$	-5.8	$0.9 \pm 0.1$
<b>13</b>	turns	$67.6 \pm 0.1$	-13.8	$2.2 \pm 0.1$
<b>14</b>	all	$\sim 31^b$	-50	$\sim 8$

<sup>a</sup>. Relative to **1**.

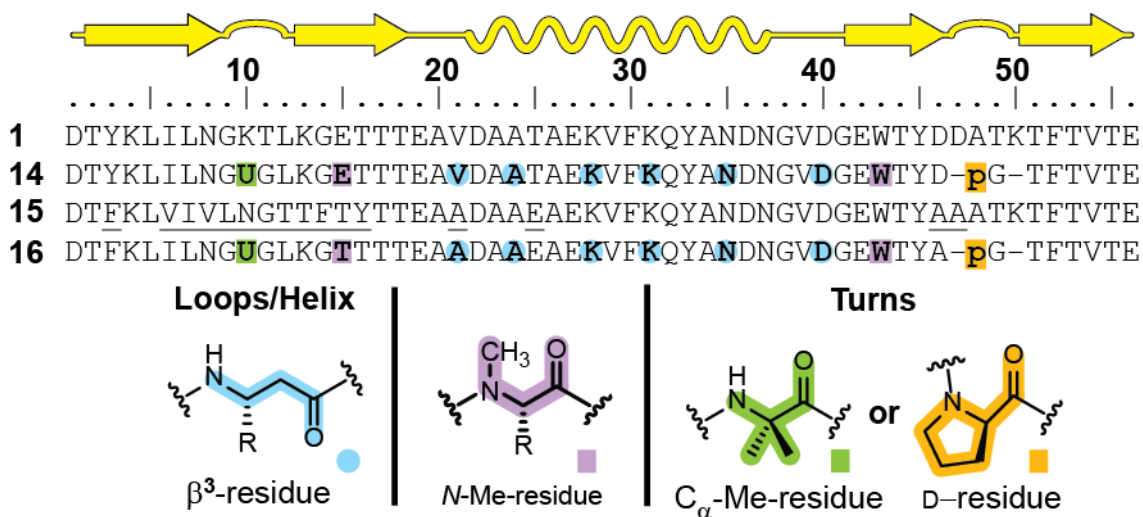
<sup>b</sup>. A complete folded baseline was not observed.

### 2.3.2 “NuG-All” Mutant

As stated in the introduction, the hypothesis guiding our designs is that every protein has two orthogonal sequences: side-chain and backbone. Improvements or deteriorations from changes to one sequence should equivalently affect the orthogonal sequence provided the three dimensional

structure of the oligomer is unchanged. The CD data on **14** suggested the presence of a tertiary structure albeit with low folded stability. Due to the likely similar folded state, wild-type GB1 and All mutant **14** could serve as a model to test our two-sequence hypothesis.

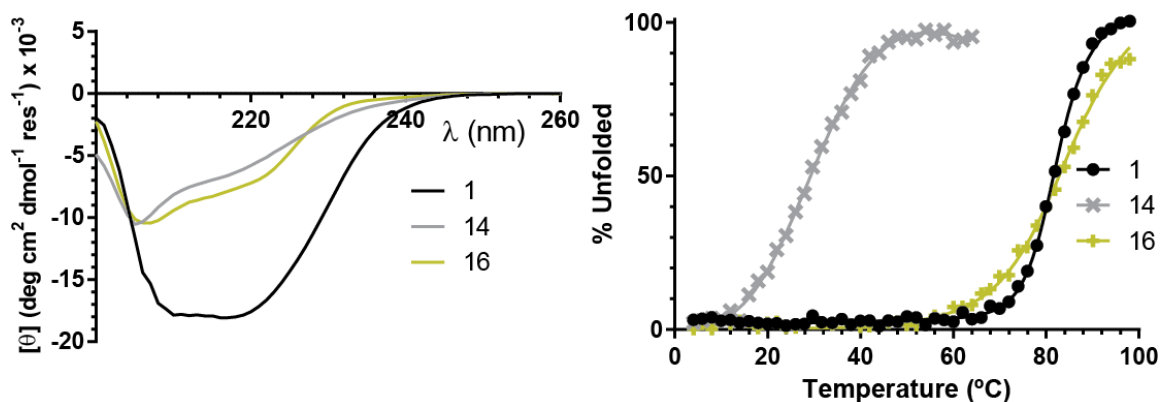
Proteins **1** and **14** share nearly an identical side-chain sequence, yet the backbone sequence differs significantly. If the side-chain sequence were to be altered and found to be stabilizing to the wild-type protein, then that same side-chain modification should stabilize the All mutant as well. Previous computational work on GB1 designed an all  $\alpha$ -residue mutant of GB1, referred to as NuG, that had an identical folded state that was approximately 5 kcal/mol more stable than native GB1.<sup>133</sup> The side-chain sequence of NuG (**15**) was grafted onto the backbone of the All mutant to generate NuG-All mutant **16** (Figure 42). **16** has the identical backbone sequence of **14** and the high stability fold-encoding side chain sequence of NuG.



**Figure 42.** Sequences of wild-type protein **1**, all mutant **14**, NuG **15**, and NuG-All mutant **16**.

Underlined residues indicate side-chain sequence alterations relative to **1**. R is the side-chain group of the corresponding  $\alpha$ -residue.

Gratifyingly, the tertiary fold of **16** was much more stable than **14** as revealed by CD thermal melts (Figure 43, Table 5). The  $T_m$  of NuG mutant **16** was nearly identical to that of wild-type GB1, translating to an approximately 8 kcal/mol stabilization over All mutant **14**. The folding transition of **16** is somewhat less cooperative than either **14** or wild-type **1**. A similar reduction in folding cooperativity was previously observed in NuG.<sup>133-134</sup> Side-chain sequence alterations in GB1 can change the folding mechanism<sup>134</sup> and could explain the differences in folding between **16** and **1**. Despite the large difference in folding free energies between **14** and **16**, CD scans revealed matching signatures underscoring the limitations of CD spectra of proteins with unnatural backbones. Native-like cooperative unfolding transitions were observed for **16**. The observation that side-chain mutations stabilizing to the wild-type protein also stabilize a highly modified backbone variant provided further evidence for the native-like folds of **14** and **16**. Unfortunately, we were unable to obtain crystals of **14** or **16** to confirm the folded structure of either protein.



**Figure 43.** CD scans at 25 °C and thermal melts of 40  $\mu$ M wild-type **1** and mutants **14** and **16** in 20 mM phosphate buffered water pH 7.

**Table 5.** Summary of Folding Thermodynamics for Wild-type GB1 and All Mutants

Sequence	2 ° modified	T <sub>m</sub> (°C)	$\Delta$ T <sub>m</sub> (°C) <sup>a</sup>	$\Delta\Delta$ G <sub>fold</sub> (kcal/mol) <sup>a</sup>
<b>1</b>	-	81.4 ± 0.1	-	-
<b>14</b>	all	~31 <sup>b</sup>	-50	~8
<b>16</b>	NuG-all	~82 <sup>c</sup>	~0	~0

*a.* Relative to **1**.

*b.* A complete folded baseline was not observed.

*c.* A complete unfolded baseline was not observed.

## 2.4 CONCLUSIONS AND OUTLOOK

This work has shown that a native protein with a defined tertiary fold can serve as a blueprint for the design of backbone-modified variants. The folded structure of GB1 possesses every secondary structure common in proteins and each could be modified individually and in tandem.

Although the backbone alterations were destabilizing to the tertiary fold in all cases, nearly all mutants retained the sequence-encoded fold of the wild-type and high resolution crystal structures of several mutants were obtainable. Combining several secondary structure modifications into one sequence led to a near-additive destabilization of the folded state suggesting that the stability of highly modified mutants could be reliably predicted from individual backbone alterations. It should then be possible to use improved substitution strategies to increase the folded stability of foldameric variants of proteins.

The collective dataset also bears favorably on our two sequence design hypothesis: backbone and side-chain sequences are orthogonal design variables inherent to every protein. The folded stabilities of both wild-type protein **1** and All mutant **14** were greatly increased by changing the side-chain sequence to that of NuG. Stated another way, replacing the all  $\alpha$ -residue backbone of either wild-type GB1 or the optimized NuG variant **15** with that of **14** was highly destabilizing to the sequence-encoded folds. This raises the possibility of using a protein of interest to design a proteolytically resistant variant and modulate the folded stability through side-chain modifications. Protease-resistant backbones that display the complex folds of therapeutically relevant proteins could facilitate protein-protein interface modulation, a challenging drug development area.<sup>135</sup>

Our work explored several classes of unnatural residues:  $C_{\alpha}$ -methyl  $\alpha$ -residues,  $N$ -methyl  $\alpha$ -residues,  $D$ - $\alpha$ -residues,  $\beta$ -residues, and PEG-derived residues. It would be worthwhile to explore additional classes as well as different substitution strategies in an attempt to improve the thermodynamic stabilities of the GB1 mutants. Establishing the relationship between proteolytic stability and backbone sequence, determining the physical rationale behind why certain unnatural

residues destabilize protein structures are investigated in subsequent chapters. The applicability of the tertiary design rules established herein to other protein sequences remains to be measured.

## 2.5 EXPERIMENTAL

### 2.5.1 General Information

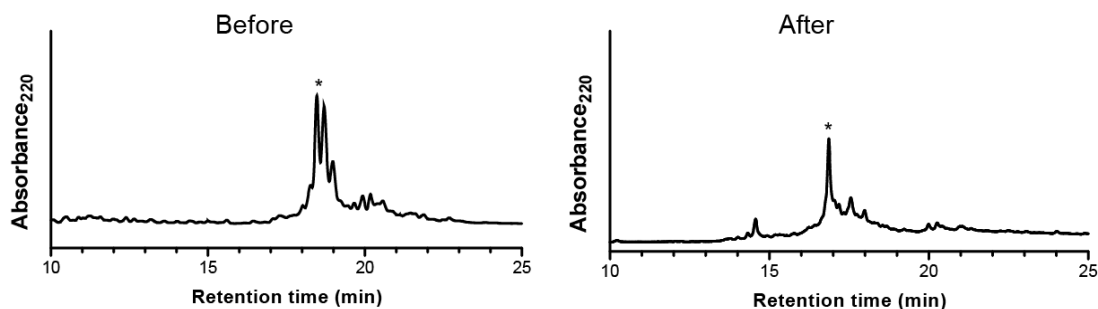
Solvents and all other reagents were purchased from Aldrich, Baker, EMD, or Fisher and used without further purification. HOBt was purchased from Anaspec Inc. HCTU, NovaPEG Rink Amide Resin, and Fmoc-protected  $\alpha$ -amino acids were purchased from Novabiochem. Fmoc-protected  $\beta^3$ -amino acids were purchased from Aapptec. Fmoc-Aib-OH, Fmoc-D-Pro-OH, and the residues preceding adjacent residues were coupled with PyAOP in place of HCTU. Microcleavages were taken after all PyAOP and pseudoproline dipeptide couplings.

### 2.5.2 Protein Synthesis

Optimization of the chemical synthesis of GB1 was developed in order to effectively obtain sufficient amounts of wild-type and synthetic proteins for characterization. Before describing the final procedures used for SPPS of GB1, the development of the optimized method is presented below.

### 2.5.2.1 General Considerations for Fmoc-SPPS of GB1

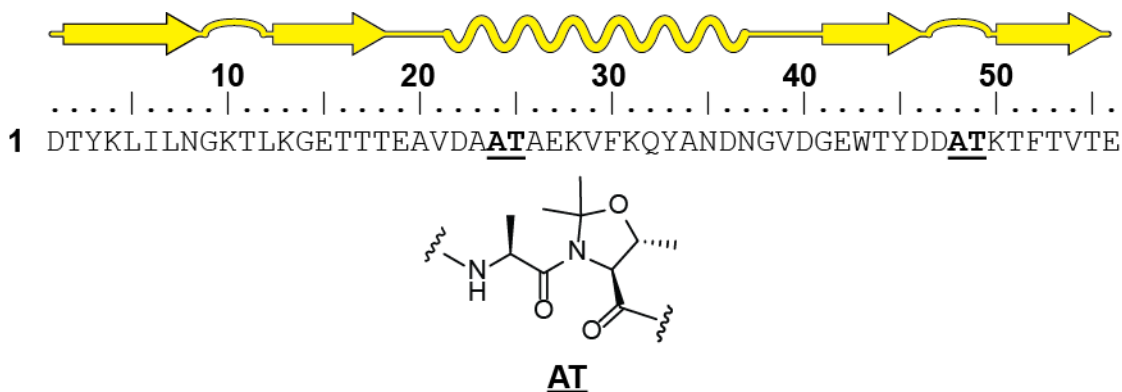
Solid-phase peptide synthesis has been vastly improved since it was invented in 1963.<sup>136</sup> Modern chemical syntheses of proteins utilize microwave-assisted Fmoc-SPPS. A general procedure is as follows. Resin is allowed to swell in DMF before a coupling reaction is carried out. Four equivalents of a pre-activated mixture of a coupling reagent and the last amino acid of the sequence (for example, Glu<sub>56</sub> in GB1) in a polar aprotic solvent such as NMP are added. The reaction vessel is heated with microwave radiation stirred. The resin is washed three times with DMF before a solution of 20% (v/v) 4-methyl piperidine in DMF is added. Deprotection reactions are similarly carried out under microwave heating. After washing three times with DMF, the cycle is repeated starting with the coupling reaction until the peptide sequence is completed. Treatment of the resin with an acidic cleavage cocktail removes side-chain protecting groups and cleaves the peptide from resin. The crude peptide is then subjected to reversed-phase HPLC purification methods. Peptides in excess of 90 residues have been prepared using similar methods.<sup>137</sup>



**Figure 44.** Crude HPLC traces from GB1 microwave-assisted SPPS before and after optimization.

Although GB1 is only 56 residues in length, we found that total chemical synthesis of GB1 and backbone-modified analogues required modification of standard microwave-assisted Fmoc-SPPS methods. To account for the increased sequence length and hydrophobicity of GB1, Fmoc-amino acid and HCTU equivalents were increased from four to six relative to resin. We initially synthesized the C-terminal 28 residues of protein **1** by this method. HPLC traces of crude material obtained showed desired product along with several deletion products (Figure 44). Poor crude purity resulted partially from aspartimide formation, a well-known problem of solid-phase peptide synthesis<sup>138</sup> especially when used with microwave irradiation.<sup>139</sup> To suppress aspartimide formation, the deprotection conditions were changed to 0.1 M HOBt in 5 % piperazine in DMF.<sup>139</sup> Deletion products may be derived from peptide aggregation on resin due to the hydrophobic nature of commonly used amino acid side-chain protecting groups. To mitigate aggregation of the growing peptide chain, pseudoproline dipeptides were incorporated at two positions in the sequence (Figure 45), proximal to the  $\beta$ -strands in the full-length protein. Pseudoproline dipeptides disfavor aggregation of the growing peptide chain through disruption of intermolecular and intramolecular hydrogen bonding.<sup>140</sup> Significant increases in crude purities were observed in subsequent syntheses.

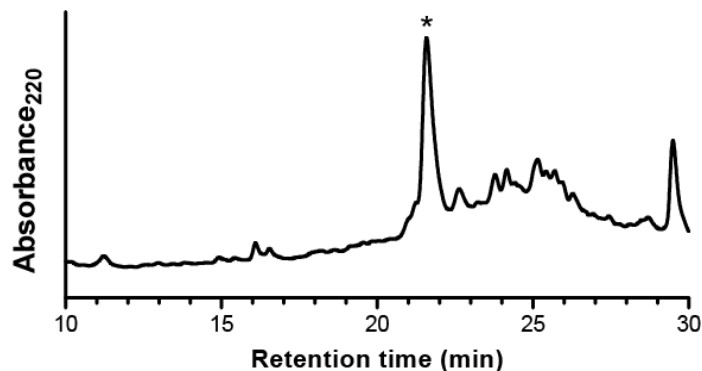




**Figure 45.** Location of pseudoproline dipeptide residues in synthesis of wild-type **1** sequence.

The oxazolidine-based backbone is removed under acidic cleavage conditions to yield two  $\alpha$ -residues.

As an alternative to microwave-assisted SPPS, we employed room temperature Fmoc-SPPS methods using an automated peptide synthesizer for some GB1 analogues. The best crude purities were obtained using a further optimized methodology. Coupling reaction times were increased from 30 minutes to 45 minutes. Two deprotection reactions were performed per amino acid with 4-methyl piperidine in DMF for four minutes. It was found that deletions occurring at Glu<sub>56</sub> and Glu<sub>15</sub> were often the major side products of syntheses. Double coupling of these residues further improved crude purities and yields (Figure 46).



**Figure 46.** Crude HPLC trace from GB1 SPPS by automated room temperature methods.

### 2.5.2.2 Optimized Fmoc-SPPS Methods

Two different SPPS protocols were utilized to synthesize proteins **1-14** and **16**. For all syntheses, after the final deprotection reaction, the resin was washed three times with 3 mL of dichloromethane followed by three 3 mL washes of methanol. After drying in a vacuum desiccator, the resin was cleaved.

**Microwave-assisted synthesis.** Proteins **1-4** were synthesized by this method. Coupling reactions were carried out using six equivalents of protected amino acid, six equivalents HCTU, and eight equivalents of DIEA in 1.2 mL of N-methylpyrrolidinone relative to resin. After two minutes of preactivation, the activated amino acid was added to resin and heated with microwave radiation to 70 °C over two minutes. The reaction was held at 70 °C for four additional minutes. Deprotection reactions were performed with 2.0 mL of 5% piperazine with 0.1 M HOBt in DMF. The resin was then heated to 80 °C over two minutes with microwave radiation. The reaction was

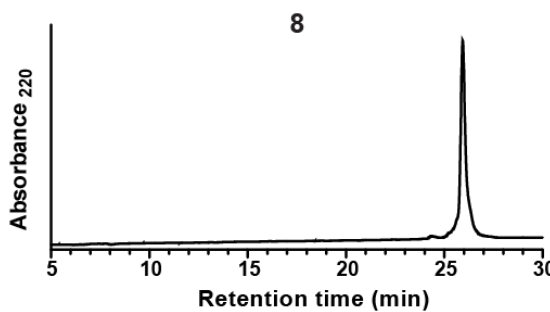
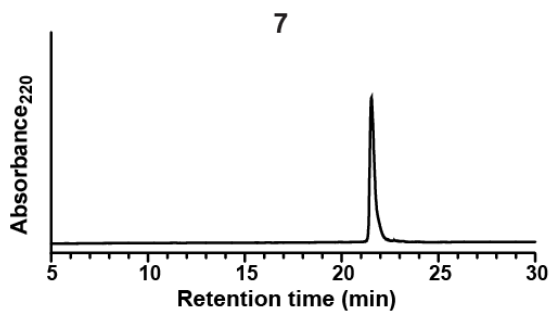
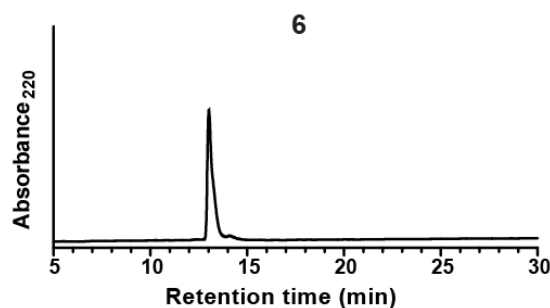
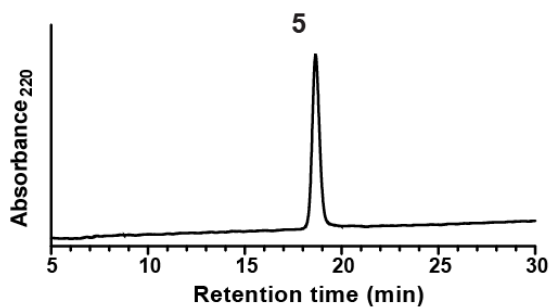
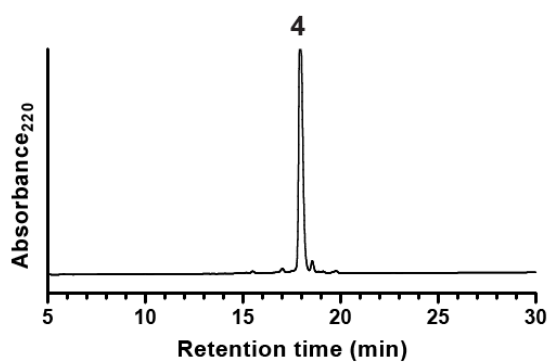
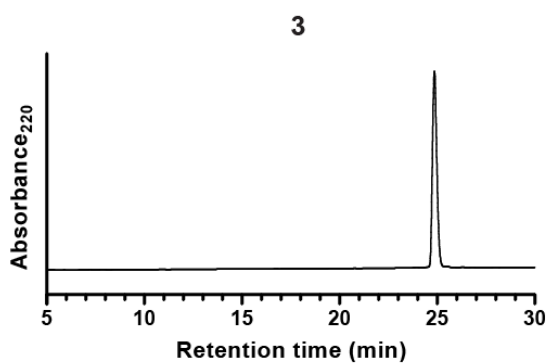
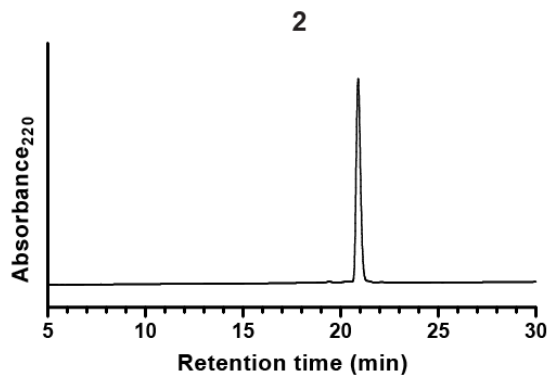
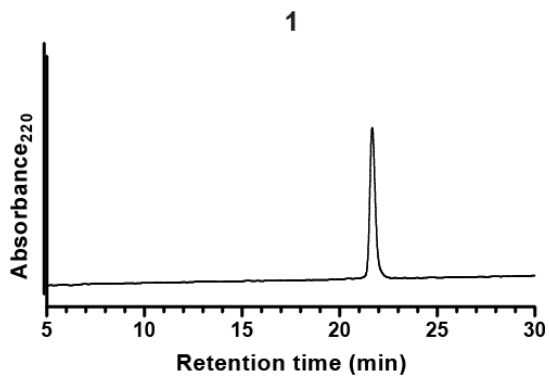
maintained at 80 °C for an additional two minutes. After each coupling and deprotection step, the resin was washed three times with 3 mL of DMF for 40 seconds each. Pseudoproline dipeptides were incorporated into proteins by coupling at room temperature for 45 minutes with identical volumes and reagents as all other Fmoc-protected amino acids.

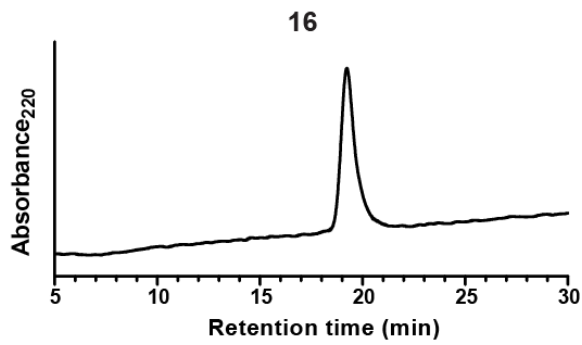
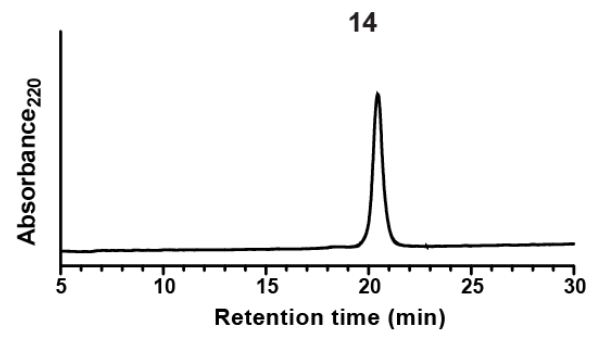
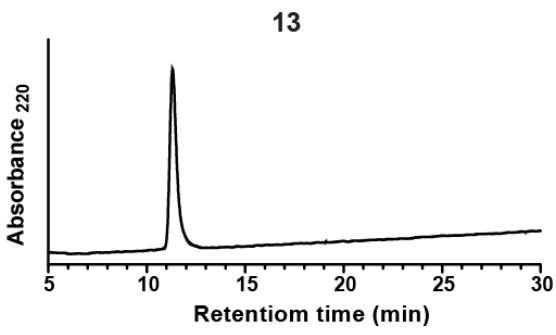
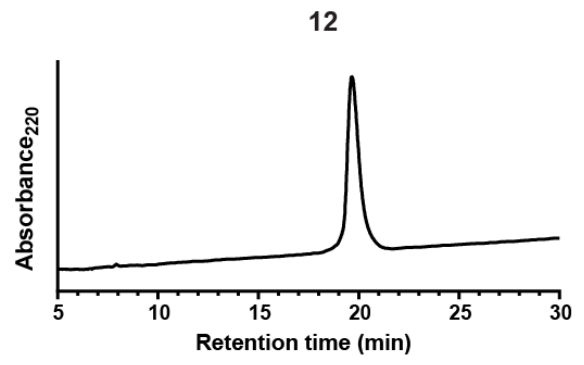
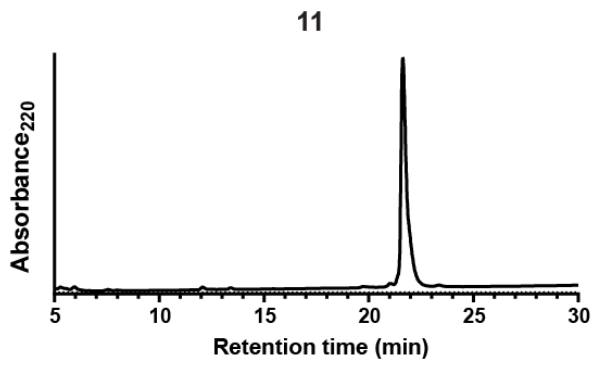
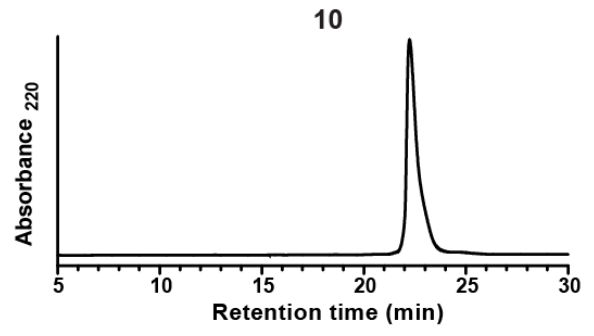
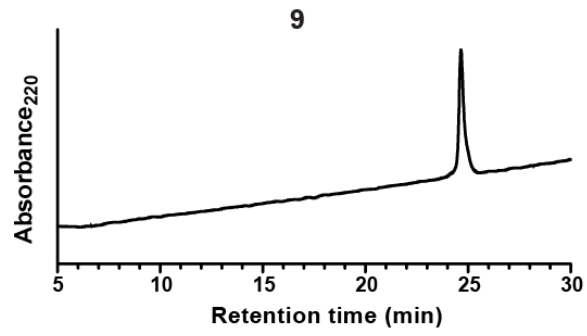
**Automated synthesis.** Proteins **5-14** and **16** were synthesized on a PTI Tribute automated synthesizer on 70 µmol scale of NovaPEG Rink Amide resin. Coupling reactions were performed using 3 mL of a 0.4 M N-methylmorpholine in DMF solution added to 7 equivalents of the Fmoc-amino acid and HCTU. After a two minute preactivation, the activated amino acid is added to the resin and vortexed for 45 minutes. Deprotection reactions were carried out twice with 3 mL of a 20% v/v solution of 4-methylpiperidine in DMF for four minutes. After each coupling or deprotection the resin was washed four times with 3 mL of DMF for 40 seconds each. After the final deprotection reaction, the resin was washed three times with 3 mL of dichloromethane followed by three 3 mL washes of methanol. After drying in a vacuum desiccator, the resin was subjected to cleavage.

**Cleavage.** Peptides were cleaved from resin by treatment with a solution of 94% TFA, 1% TIS, 2.5% water, and 2.5% EDT. Crude peptides were precipitated by addition of cold diethyl ether. The solid was pelleted by centrifugation and dissolved in 6 M guanidinium chloride, 25 mM sodium phosphate, pH 5.7. This solution was then subjected to purification.

### **2.5.3 Protein Purification and Characterization**

Each protein was purified by preparative C18 reverse-phase HPLC using gradients between 0.1% TFA in water and 0.1% TFA in acetonitrile. The identity of each was confirmed by MS analysis on a Voyager DE Pro MALDI-TOF instrument (Table 6). Following HPLC purification, each protein was subjected to anion-exchange chromatography on a monoQ 5/50GL column, (GE Healthcare) using 0.02 M Tris buffer at pH 8 and eluted with increasing concentrations of KCl. Final protein samples were  $\geq 95\%$  pure by analytical reverse-phase HPLC (Figure 47).





**Figure 47.** Analytical HPLC chromatograms of purified proteins **1-14**, and **16**.

**Table 6.** MALDI-TOF MS Data for Proteins **1-14**, and **16**.

#	[M+H] <sup>+</sup> m/z (avg.)	
	Calculated	Observed
<b>1</b>	6179.6	6178.6
<b>2</b>	6024.5	6023.9
<b>3</b>	6052.5	6052.1
<b>4</b>	5898.4	5897.7
<b>5</b>	6207.7	6204.4
<b>6</b>	6235.8	6232.8
<b>7</b>	6235.8	6237.3
<b>8</b>	5789.3	5791.5
<b>9</b>	6271.9	6274.6
<b>10</b>	6207.7	6207.5
<b>11</b>	6085.5	6089.2
<b>12</b>	5672.1	5674.0
<b>13</b>	5831.2	5828.5
<b>14</b>	5943.5	5944.1
<b>16</b>	5995.6	5997.8

#### **2.5.4 Competition ELISA.**

ELISAs were performed using pre-coated Protein G plates (Sigma-Aldrich). Before use, assay plates were blocked for 2 h at ambient temperature in a humid atmosphere with dilution buffer (25 mM sodium phosphate with 0.15 M NaCl, pH 7.4, 0.05% (v/v) Tween-20, 3% BSA). Stock plates were prepared containing serial dilutions of each protein in dilution buffer. 10  $\mu$ L of this protein stock was then added to 80  $\mu$ L of dilution buffer, followed by addition of 10  $\mu$ L of 0.25  $\mu$ g/mL rabbit anti-goat IgG-HRP (Biorad) conjugate antibody to wells. Plates were mixed, incubated for 1 h at ambient temperature, and washed with wash buffer (25 mM sodium phosphate with 0.15 M NaCl, pH 7.4, 0.05% (v/v) Tween-20) four times for 10 seconds each. 100  $\mu$ L of TMB Substrate System (Sigma) was added to wells, mixed, and incubated for 20

minutes at ambient temperature. 100  $\mu$ L of 1 M HCl was then added to wells and plates were read at 450 nm on a Modulus II multimode plate reader (Turner Biosystems). Reported values are averages of six replicate measurements. Each assay plate included control wells containing only TMB Substrate System that were developed as described above and used to determine the A450 value for 0 % binding of antibody to well. Data were fit to a sigmoidal dose-response curve with variable slope using GraphPad Prism. The Hill slope was found to be -0.65; values less than 1 have previously been observed for monovalent inhibitors and multivalent receptors in a competition ELISA.<sup>141</sup>

### **2.5.5 Circular Dichroism Spectroscopy.**

CD measurements were made on an Olis DSM17 Circular Dichroism Spectrometer. Measurements were carried out on samples consisting of 40  $\mu$ M protein in 20 mM sodium phosphate buffer, pH 7.0 with 2 mm quartz cells. Scans were taken at 25  $^{\circ}$ C over the range of 200-260 nm with 1 nm increments and a 1 nm bandwidth. Scan data were smoothed by the Savitsky-Golay method as implemented in GraphPad Prism. Melts were monitored at 220 nm over the range of 4  $^{\circ}$ C to 98  $^{\circ}$ C with 2  $^{\circ}$ C increments, a dead band of 0.5  $^{\circ}$ C, and a 2 minute equilibration time for each temperature point. All measurements were baseline corrected for blank buffer. Temperature-dependent CD data were fit to simple two-state unfolding model using GraphPad Prism to obtain melting temperature ( $T_m$ ) and enthalpy of folding at  $T_m$  ( $\Delta H_{fold}$ ). The change in free energy of folding for each mutant relative to wild-type ( $\Delta\Delta G_{fold}$ ) was calculated according to equation **2.1**:



### Equation 2.1

$$\Delta\Delta G_{fold} = \Delta H_{fold} \frac{\Delta T_m}{T_m}$$

where  $\Delta H_{fold}$  is the enthalpy of folding at  $T_m$  for the wild-type,  $T_m$  is the melting temperature for the wild-type, and  $\Delta T_m$  is the change in melting temperature for mutant relative to wild-type.<sup>96-97</sup> This relationship assumes that the heat capacity change of folding is identical for the mutant and the wild-type.

### 2.5.6 Gel Permeation Chromatography.

GPC was performed on a Superdex 75 10/300 GL column (10 x 300 mm, 24 mL bed volume, 13  $\mu\text{m}$  average particle size, GE Healthcare). The column was equilibrated with 150 mM NaCl in 20 mM sodium phosphate, pH 7.0. Protein samples (40  $\mu\text{M}$  protein in equilibration buffer) were loaded onto the column and eluted at a flow rate of 1.0 mL/min.

### 2.5.7 Molecular Dynamics.

MD simulations were performed by collaborators (Eli Musselman and Adrian Elcock, University of Iowa). Constructs for proteins **1-4** were modeled starting from a published 1.05  $\text{\AA}$  crystal structure of GB1 from *Staphylococcus aureus* (PDB 2QMT). In all cases, the first two residues were mutated from Met-Gln- to Asp-Thr- and the C-terminal Glu residue was amidated so as to be consistent with the corresponding experimental sample. The initial structures for the three

PEG-chimeras were completed by deleting the sidechains of the appropriate amino acids and altering their backbone N-C $\alpha$ -C atoms to the C-C-O atoms of the PEG repeating unit.

MD simulations were performed in explicit solvent using GROMACS 4 software.<sup>142-145</sup> Amino acid residues were modeled with the OPLS-AA forcefield<sup>146</sup> and PEG residues were modeled using the TraPPE-UA parameters developed by Fischer *et al.*<sup>147</sup> In separate work (Musselman & Elcock; manuscript in preparation), we have verified that our implementation of the PEG parameters produces conformational behavior consistent with that seen in previous simulation<sup>147-148</sup> and experimental<sup>149</sup> studies. In order to maintain consistency of the force fields as far as possible, all pairwise interactions involving PEG atoms were treated using Lorentz-Berthelot mixing rules,<sup>150</sup> otherwise standard OPLS-mixing rules<sup>151</sup> were used. Each protein construct was immersed in a cubic (periodic) simulation box of side 60 Å, with approximately 7000 water molecules added; in order to electrically neutralize each system, 2 or 3 Na<sup>+</sup> ions were added depending on the net charge of the construct. In all cases water was described by the TIP3P model<sup>152</sup> and Na<sup>+</sup> ions were modeled using Aqvist's parameters.<sup>153</sup>

Prior to MD simulation all systems were first energy minimized using the steepest descent algorithm for 1000 steps and then incrementally heated to 298 K in 50K intervals over the course of 500 ps. Each system was then equilibrated for a further 10.5 ns before the production phase of the simulation was begun. Each construct was simulated for 1  $\mu$ s of production time; system coordinates were saved at intervals of 0.1 ps for subsequent analysis. All simulations were performed in the NPT ensemble: pressure (1 atm) was maintained with the Parrinello-Rahman barostat<sup>154</sup> and temperature (298 K) was maintained using the Nosé-Hoover thermostat.<sup>155-156</sup>

Cutoffs for both the short-range electrostatic and van der Waals interactions were set to 10 Å; all long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method.<sup>157</sup> In all simulations covalent bond lengths were constrained using the LINCS algorithm,<sup>158</sup> enabling a 2.5 fs time step to be used. Each 1 μs production simulation required ~3 weeks of dedicated time on a 48-core Dell R815 server. MD trajectories for 1-4 were analyzed using various tools in the GROMACS package to extract time-dependent RMSD plots, dihedral trajectories and distributions, and dihedral autocorrelation functions. The dihedral autocorrelation functions were fit to two-phase exponential decays using GraphPad Prism.

## 2.5.8 Crystallization of Proteins

Diffraction-quality crystals were obtained for proteins **5-7** and **13** by hanging drop vapor diffusion. ~17 mg/mL solution of protein in water was mixed (1 μL + 1 μL) with well buffer (Table 7). The drop was allowed to equilibrate at room temperature (**5, 7, 13**) or at 4 °C after two weeks at room temperature (**6**).

**Table 7.** Crystallization Conditions for Proteins **5-7** and **13**.

<b>Protein</b>	<b>Buffer</b>	<b>Precipitant(s)</b>
<b>5</b>	100 mM NaOAc pH 4.6	16% PEG 3350
<b>6</b>	150 mM NaOAc pH 4.6	20% PEG 4000
<b>7</b>	100 mM NaOAc pH 4.6	30% isopropanol, 0.2 M CaCl <sub>2</sub>
<b>13</b>	100 mM NaOAc pH 4.6	8% PEG 4000

X-ray diffraction data collection and structure determination was performed by W. Seth Horne. A single crystal of each protein was flash frozen in liquid nitrogen after cryoprotection in well buffer supplemented with 30% v/v glycerol. Diffraction data were collected using CuK $\alpha$  radiation on a Rigaku/MSM diffractometer (FR-E generator, VariMax optics, AFC-Kappa goniometer, Saturn 944 CCD detector) equipped with an X-Stream 2000 low temperature system operated at 100 K. Raw diffraction data were indexed, integrated, and scaled with d\*TREK.

Structure solution and refinement were carried out using the CCP4<sup>159</sup> and Phenix<sup>160</sup> software suites. The structures were solved by molecular replacement using the program Phaser<sup>161</sup> with a model derived from a published structure of wild-type GB1 (PDB: 2QMT<sup>78</sup>). Refinement was performed by a combination of Refmac<sup>162</sup> and Phenix for automated refinement, Coot<sup>163</sup> for manual model building, and Phenix<sup>164</sup> for automated model building. Geometric restraints for the  $\beta$ -residues were assembled and used in a custom cif library during refinement.

In the case of protein **6**, there was some ambiguity in space group assignment. The diffraction pattern readily indexed as tetragonal but analysis of intensities during integration and scaling suggested the crystal lattice was actually C2221 with  $a \approx b$ . Molecular replacement in this space group was successful, but stalled R/Rfree during refinement indicated the possible presence of twinning. Analysis of the diffraction data with Xtriage in the Phenix software suite suggested the true space group to be P21 with near perfect pseudomerohedral twinning (twin operator: L,  $-K$ , H). The raw diffraction data were reprocessed in P21 and structure refinement carried out using the twin refinement options implemented in Refmac and Phenix.

**Table 8.** X-ray Diffraction Data and Refinement Statistics for Proteins **5-7**, and **13**.

#	<b>5 (4KGS)</b>	<b>6 (4KGR)</b>	<b>7 (4OZA)</b>	<b>13 (4KGT)</b>
<b>Data Collection</b>				
<b>Unit cell dimensions</b> (Å, °)	$a = 52.2, b = 81.2$ $c = 52.1$ $\alpha = 90, \beta = 90,$ $\gamma = 90$	$a = 80.7, b = 35.7,$ $c = 46.5$ $\alpha = 90, \beta = 120.4,$ $\gamma = 90$	$a = b = 65.9, c =$ $21.9$ $\alpha = \beta = \gamma = 90$	$a = b = 83.8, c = 97.5$ $\alpha = \beta = \gamma = 90$
<b>Space group</b>	P21	C2	I41	I4122
<b>Resolution (Å)</b>	32.03–2.00 (2.07–2.00)	28.27–1.95 (2.02–1.95)	23.31-2.20 (2.28- 2.20)	23.28–2.00 (2.07–2.00)
<b>Total observations</b>	97,753	51,377	13,203	166,708
<b>Unique observations</b>	27,508	8,453	2,418	12,054
<b>Redundancy</b>	3.55 (2.68)	6.08 (3.64)	5.46 (5.42)	13.83 (13.81)
<b>Completeness (%)</b>	93.7 (82.7)	99.6 (96.3)	96.6 (98.8)	100 (100)
<b>I/σ</b>	23.6 (5.1)	16.2 (4.1)	11.0 (3.2)	16.7 (4.5)
<b>Rmerge (%)</b>	4.4 (20.3)	7.8 (22.5)	10.2 (26.4)	7.6 (39.6)
<b>Refinement</b>				
<b>Resolution (Å)</b>	32.03–2.00	28.27–1.95	23.31-2.20	23.28–2.00
<b>R (%)</b>	16.2	17.6	23.3	19.8
<b>Rfree (%)</b>	19.2	21.7	26.0	21.0
<b>Avg. B factor (Å<sup>2</sup>)</b>	23.6	24.9	38.1	39.4
<b>RMSD</b>				
<b>Bonds (Å)</b>	0.011	0.008	0.007	0.007
<b>Angles (°)</b>	1.54	1.12	1.14	1.05

### 3.0 FOLDING THERMODYNAMICS OF BACKBONE-MODIFIED PROTEINS

Some of the results in this chapter have been published in:

Z.E. Reinert, W.S. Horne. "Folding thermodynamics of protein-like oligomers with heterogeneous backbones." *Chemical Science*. **2014**, 5, 3325-3330.

Crystallography data collection and structure refinement were performed by W. Seth Horne and Nathan A. Tavenor. Chemical synthesis, purification, and characterization of C<sub>α</sub>-methyl containing GB1 mutants were performed by Nathan A. Tavenor.

Incorporation of backbone mutations in GB1 were destabilizing to the tertiary fold as observed in Chapter 2. One open question was the physical basis for the energetic penalties. We sought to answer this by exploring the fundamental folding thermodynamics of proteins with unnatural backbones.

## 3.1 HELIX SUBSTITUTION STRATEGIES

### 3.1.1 $\alpha \rightarrow \beta^3$ Substitution

Our previous work established that backbone modifications to the helix, loops, sheet, and turns of GB1 can be combined without sacrificing the sequence-encoded fold, however, at the cost of thermodynamic stability (see Chapter 2). In an orthogonal manner, it was shown that side-chain sequence changes that improved the folded stability of the wild-type protein also stabilized a highly unnatural backbone variant. This suggests that stabilizing alterations to the backbone sequence of a protein would enhance the folded stability as well. One route towards improving backbone modification of proteins is to understand the physical principles behind the thermodynamic destabilization of current strategies.

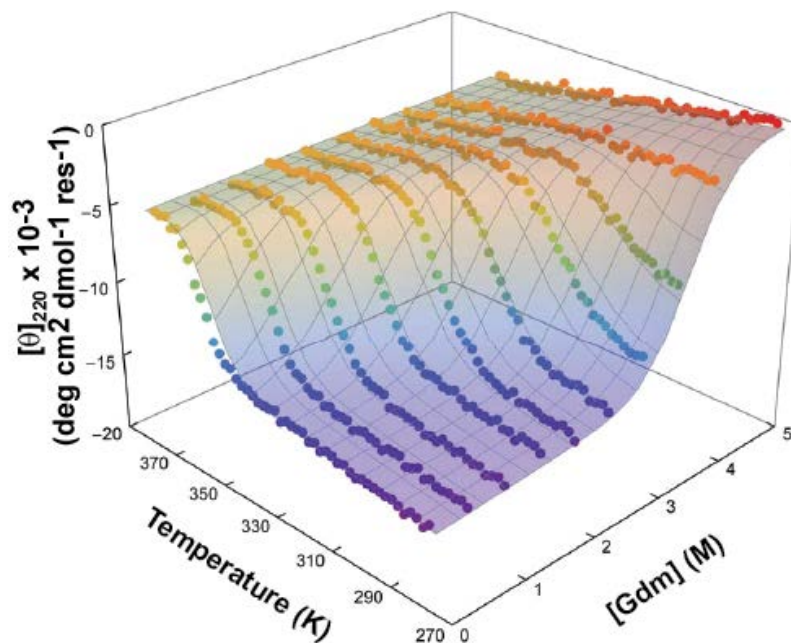
The thermodynamics of protein folding have been a longstanding area of intense research.<sup>165</sup> For most proteins, folding in water is enthalpically driven and entropically opposed.<sup>166</sup> In the denatured state, water hydrogen bonds to polar groups on the polypeptide chain and forms ordered “icebergs” around exposed hydrophobic groups. The increase in solvent entropy from the release of ordered water as a protein folds greatly compensates for the steep chain entropic penalty of ordering a peptide into the native conformation. The newly formed intrachain hydrogen bonds and burial of exposed hydrophobic surface enthalpically drive folding of the protein. In general, the folded state of proteins are marginally (5-15 kcal/mol) more stable than the denatured state.<sup>166</sup>

Compared to native sequences, the folding thermodynamics of proteins with unnatural backbones are much less understood. We set out to examine how unnatural residues can alter classical protein folding dynamics. Given their prevalence in our work as well as the literature,

we chose to focus initially on  $\alpha \rightarrow \beta^3$  substitutions in the helix and loops of GB1 (mutants **5-7**). From a cursory examination,  $\beta^3$ -residues would be expected to increase the chain entropic penalty toward folding relative to an all  $\alpha$ -peptide due to the additional rotatable bond present in the backbone. To better understand why  $\beta^3$ -residues were unfavorable in the helix, the thermodynamic origin of the destabilization, i.e. the enthalpic and entropic changes that result in the decreased change in folding free energy, would need to be determined.

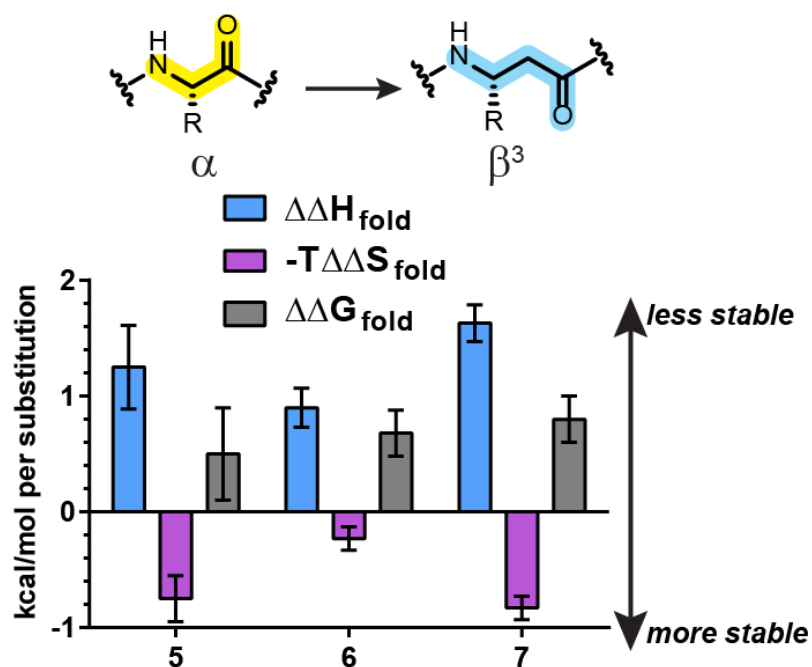
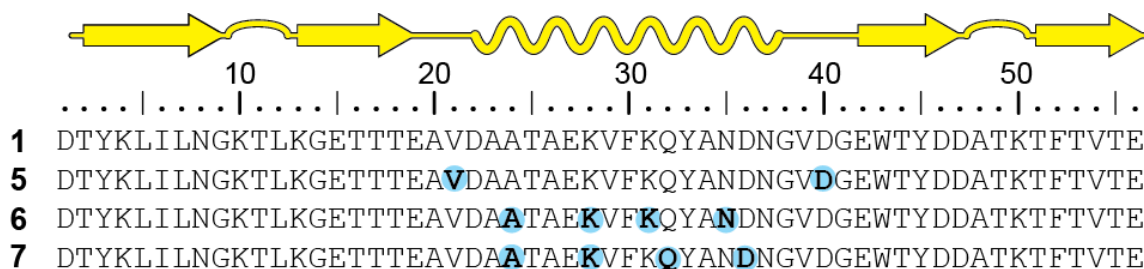
We utilized combined thermal and chemical denaturation experiments monitored by CD to obtain detailed thermodynamic data on wild-type **1** (Figure 48) and mutants **5-7**. This known method<sup>167-168</sup> monitors the 220 nm CD signal of a protein as a function of temperature over several parallel samples containing incrementally varying amounts of a chemical denaturant such as guanidinium chloride. The resulting three-dimensional plot is then fit to a surface using a modified Gibbs-Helmholtz equation (see experimental) to obtain changes to enthalpy ( $\Delta H^\circ$ ), entropy ( $\Delta S^\circ$ ), heat capacity ( $\Delta C_p^\circ$ ), and free energy ( $\Delta G^\circ$ ) upon folding as well as the sensitivity of the protein fold to denaturant concentration ( $m$ ). A fundamental challenge in protein thermodynamics is discerning contributions to folding entropy. While it is impossible to quantitatively determine changes to  $\Delta S_{\text{fold}}$  from chain and solvent entropies experimentally,  $\Delta C_p$  and  $m$  provide an indirect measure of changes to solvation of the oligomer. These thermodynamic parameters and structural data on the folded state of the mutants can inform hypotheses on changes to chain and solvent entropies.





**Figure 48.** Global fit of thermal and guanidinium chloride denaturation of 40 $\mu$ M **1** in 20 mM phosphate buffer, pH 7.

Thermodynamic data for protein **1** and heterogeneous backbone analogues **5-7** are summarized in Table 9. Our results agreed with the general observation that protein folding in water is enthalpically favored and entropically opposed<sup>169</sup> and previously reported thermodynamic values for GB1 using differential scanning calorimetry.<sup>95</sup> Importantly, the dataset provided detailed information on the relationship between backbone alterations and changes to protein folding thermodynamics.



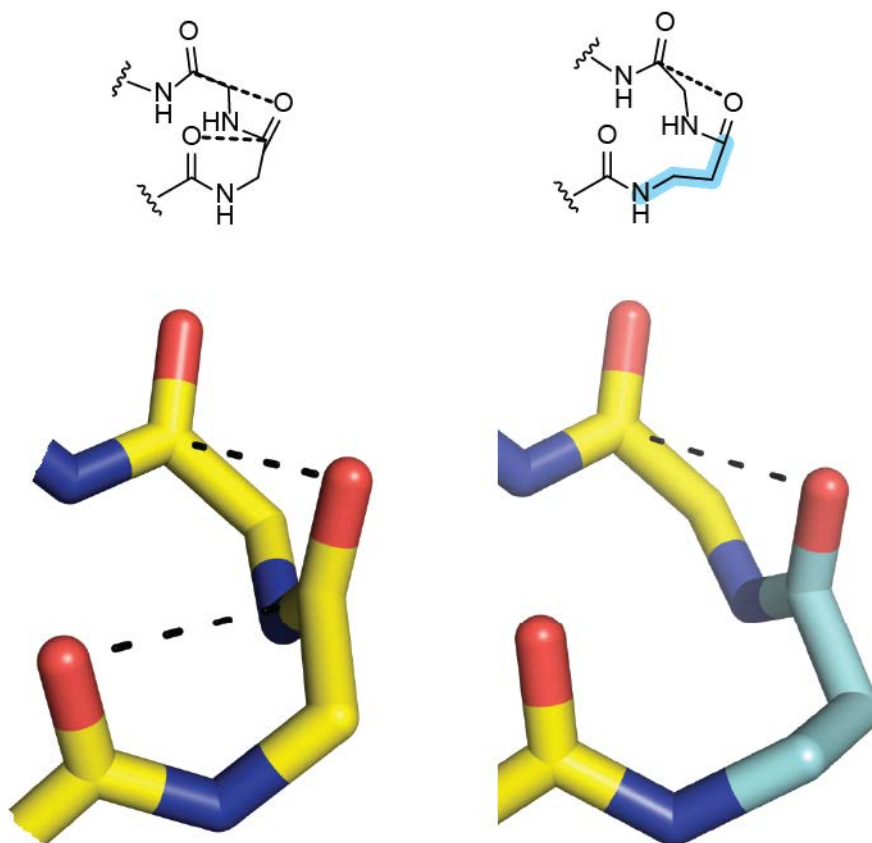
**Figure 49.** Changes in thermodynamic parameters of folding for mutants **5-7** relative to wild-type GB1 (**1**).

**Table 9.** Thermodynamic Parameters for the Folding Transitions of **1** and **5-7** at 298 K.

#	$\Delta H^\circ$ (kcal mol <sup>-1</sup> )	$T\Delta S^\circ$ (kcal mol <sup>-1</sup> )	$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$\Delta C_p^\circ$ (kcal mol <sup>-1</sup> K <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta\Delta G^\circ$ (kcal mol <sup>-1</sup> ) <sup>a</sup>
<b>1</b>	-21.9 ± 0.6	-16.1 ± 0.3	-5.8 ± 0.6	-0.60 ± 0.02	-1.80 ± 0.04	-
<b>5</b>	-19.4 ± 0.4	-14.6 ± 0.3	-4.8 ± 0.5	-0.59 ± 0.02	-1.82 ± 0.04	+1.0 ± 0.8
<b>6</b>	-18.3 ± 0.4	-15.2 ± 0.3	-3.1 ± 0.5	-0.53 ± 0.02	-2.48 ± 0.05	+2.7 ± 0.8
<b>7</b>	-15.4 ± 0.4	-12.8 ± 0.3	-2.6 ± 0.5	-0.63 ± 0.02	-2.58 ± 0.05	+3.2 ± 0.8

*a.* Relative to **1**.

In each mutant,  $\alpha \rightarrow \beta^3$  replacement led to an unfavorable change in folding enthalpy. The magnitude of the change in  $\Delta H^\circ$  relative to **1** depended on location (**5** vs **6** and **7**) and pattern of  $\beta^3$ -residue incorporation (**6** vs **7**, Figure 49). The change in folding enthalpies of loop mutant **5** and helix mutant **6** were within 1 kcal/mol despite **6** having twice the number of  $\beta^3$ -residues incorporated into a more demanding helical backbone environment. The  $\alpha\alpha\alpha\beta$  substitution pattern of helix analogue **7** was far less enthalpically favorable toward folding than the  $\alpha\alpha\beta\alpha\alpha\beta$  design of **6**. In trying to rationalize the physical basis for the consistently unfavorable enthalpy change accompanying  $\alpha \rightarrow \beta^3$  substitution in proteins **5-7**, we were drawn to recent work exploring the importance of orbital interactions in protein folding.<sup>170</sup> Partial donation of a backbone carbonyl oxygen  $n_p$  electrons into the empty  $\pi^*$  of the subsequent backbone carbonyl is thought to stabilize helical conformations. As backbone-elongated analogues of  $\alpha$ -residues,  $\beta$ -residues can only act as donors and not acceptors for this possible orbital interaction. Thus  $\alpha \rightarrow \beta^3$  replacement would lead to a loss of one interaction per substitution resulting in an enthalpic penalty (Figure 50).



**Figure 50.** Putative  $n \rightarrow \pi^*$  interactions in helices of wild-type GB1 and **6**. Dotted lines indicate orbital interactions between  $n_p$  electrons of carbonyl oxygens and  $\pi^*$  orbital of neighboring carbonyl carbons.

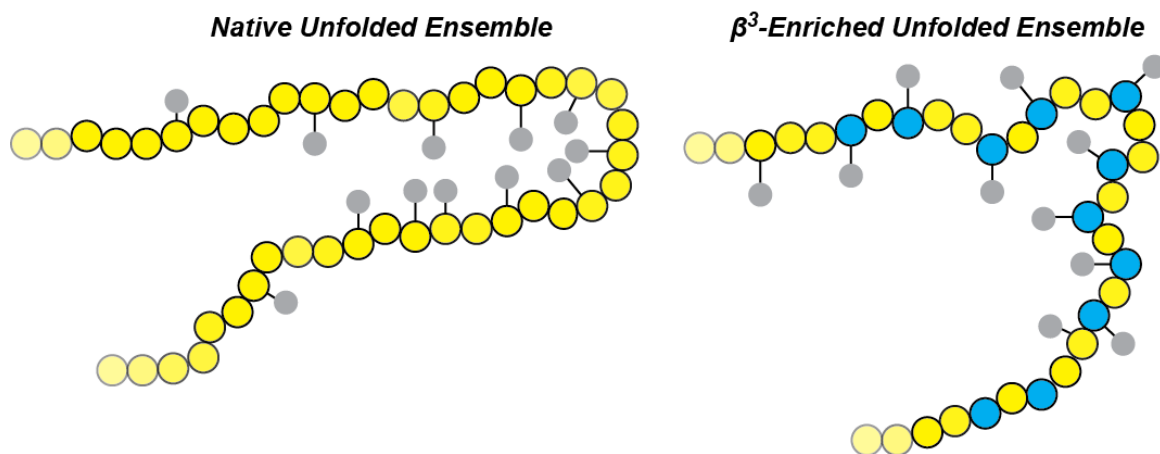
Replacement of  $\alpha$ -residues with  $\beta^3$ -analogues was expected to be entropically destabilizing towards folding due to the additional degree of backbone conformational freedom that would raise the entropy of the unfolded state. Instead, fits of the global denaturation curves of **5-7** revealed that incorporation of  $\beta^3$ -residues into the loops or helix of GB1 was entropically favorable towards folding. As observed with the change in folding enthalpies, the magnitude of the changes in  $\Delta S^\circ$  depended on location and pattern of backbone substitution. The origin of the

entropic change could be from changes in the folded state of the protein; however the crystal structures of proteins **5-7** closely matched that of wild-type GB1 and did not show elevated B factors in the areas around the  $\beta^3$ -residues. An alternate explanation is that the  $\beta^3$ -residues altered the solvation of the folded and unfolded protein states relative to **1**.

The denatured state of proteins is generally not an extended peptide chain but better described as a dynamic mixture of structures.<sup>171</sup> The complexity of the denatured ensemble is challenging to probe experimentally, yet it is the reference state against which all folding thermodynamics of proteins are measured. We hypothesized that like most proteins,<sup>169</sup> the unfolded state of GB1, minimizes solvent exposure of hydrophobic groups. In contrast, the additional conformational flexibility imparted by the  $\beta^3$ -residues may raise the entropy of the protein backbone to a sufficient degree such that hydrophobic groups are exposed to solvent. This exposure of nonpolar surface area causes unfavorable ordering of water molecules into iceberg-like structures<sup>166, 169</sup> around the hydrophobic groups. As the heterogeneous backbone folds, the nonpolar groups are buried and the ordered water is released, yielding a large favorable increase in solvent entropy that drives folding. Although the more flexible backbone of  $\beta^3$ -residues likely contributes unfavorably to the change in folding chain entropy in the protein, the large increase in solvent entropy upon folding significantly compensated any entropic penalty introduced into the protein backbone (Figure 51).

The  $\Delta C_p^\circ$  for protein folding is dependent on several factors including change in solvent accessible surface area ( $\Delta ASA$ ), hydrogen bonding, conformational entropy, and residual structure in the denatured state.<sup>172</sup> For many proteins, a negative change in heat capacity accompanies the folding transition.<sup>172</sup> Wild-type **1** and each of the  $\beta^3$ -containing mutants displayed a negative  $\Delta C_p^\circ$  of approximately the same value suggesting that  $\beta^3$ -residues do not

significantly alter the heat capacity of GB1 folding. This is somewhat surprising given the enthalpic and entropic contributions to  $\Delta C_p^\circ$  and likely reflects the complex interplay of additional factors that contribute to heat capacity in proteins.



**Figure 51.** Hypothesized differences between denatured ensembles of native and  $\beta^3$ -containing proteins. Gray spheres represent hydrophobic side-chains.

Like heat capacity,  $m$ , the sensitivity of the protein backbone to chemical denaturation, is dependent on the change in solvent accessible surface area. Changes in  $m$  can offer insight into differences to the complex denatured ensemble. In contrast to the observed changes to  $\Delta C_p^\circ$ ,  $m$  depended strongly on the secondary structure modified. While loop mutant **5** had an identical value of  $m$  as wild-type,  $\alpha \rightarrow \beta^3$  substitution in the helix of GB1 in either **6** or **7** led to a significant increase in denaturant sensitivity. One explanation could be that the  $\beta^3$ -residues altered the change in ASA through increased backbone flexibility (Figure 51). The helix of wild-type GB1

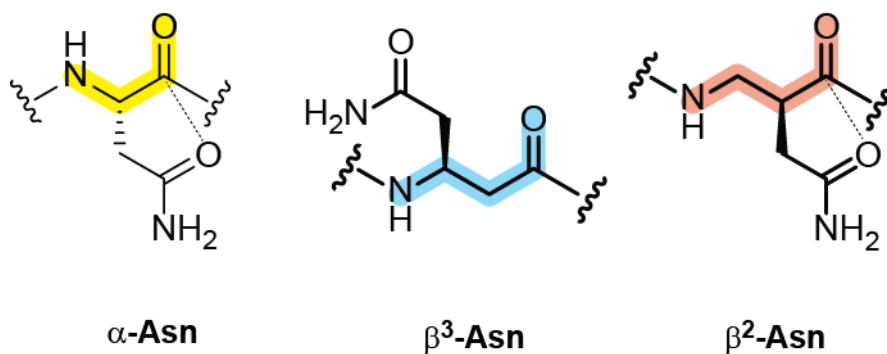
may retain residual structure in the unfolded state as opposed to the more disordered loops which could explain the difference in  $m$  values observed for **5** and **6/7**.

### 3.1.2 $\beta^3 \rightarrow \beta^2$ Substitution

In our work on sheet-based mimicry, we explored the use of both  $\beta^3$  and  $\beta^2$ -residues as replacements for  $\alpha$ -residues in model hairpin peptides.<sup>119-120</sup> As regioisomers of  $\beta^3$ -residues,  $\beta^2$ -residues have been studied primarily in  $\beta$ -peptides and found to promote helical structures in organic solvents, especially in combination with other classes of  $\beta$ -residues.<sup>48</sup> Prior computational studies compared the conformational preferences of model  $\beta^2$  and  $\beta^3$ -amino acids in explicit water.<sup>173</sup> Comparison of the lowest energy structures for each residue type with dihedrals from the crystal structure of **6** suggested that  $\beta^2$ -residues would be better accommodated in an  $\alpha/\beta$ -helix. Using  $\beta^3$ -helix mutant **6** as a starting point, we synthesized proteins **17-19** to examine the thermodynamic effects of  $\beta^3 \rightarrow \beta^2$  substitution (Figure 52). **17** and **18** explored the energetics of a single substitution, either an alanine (**17**) or asparagine (**18**) whereas the design of **19** replaces all four  $\beta^3$ -residues with the  $\beta^2$ -analogue.

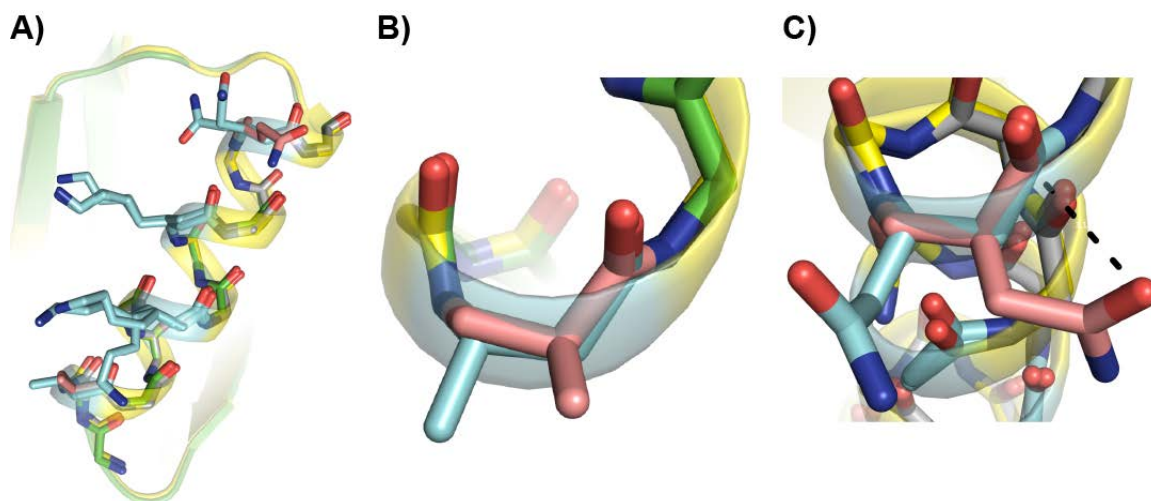






**Figure 53.** Proposed intraresidue  $n \rightarrow \pi^*$  interaction possible in  $\beta^2$ -Asn.

Proteins **17** and **18** have a fold identical to that of **6** as shown in high-resolution crystal structures (Figure 54). In both **17** and **18**,  $\beta^2$ -residues did not elevate local B-factors in the helix region. The relative positions of the  $C_\alpha$  and  $C_\beta$  atoms of the  $\beta^2$ -residues in the helices of **17** and **18** matched those of the corresponding  $\beta^3$ -analogues in **6**. In two out of four copies of mutant **18** from the crystal asymmetric unit, the side-chain carbonyl oxygen of  $\beta^2$ -Asn appeared to engage in an intraresidue  $n \rightarrow \pi^*$  interaction, supporting our design hypothesis. The distance between the side-chain oxygen and the main chain carbonyl, however, was at the upper limit for a  $n \rightarrow \pi^*$  interaction ( $> 3 \text{ \AA}$ ).<sup>174</sup>



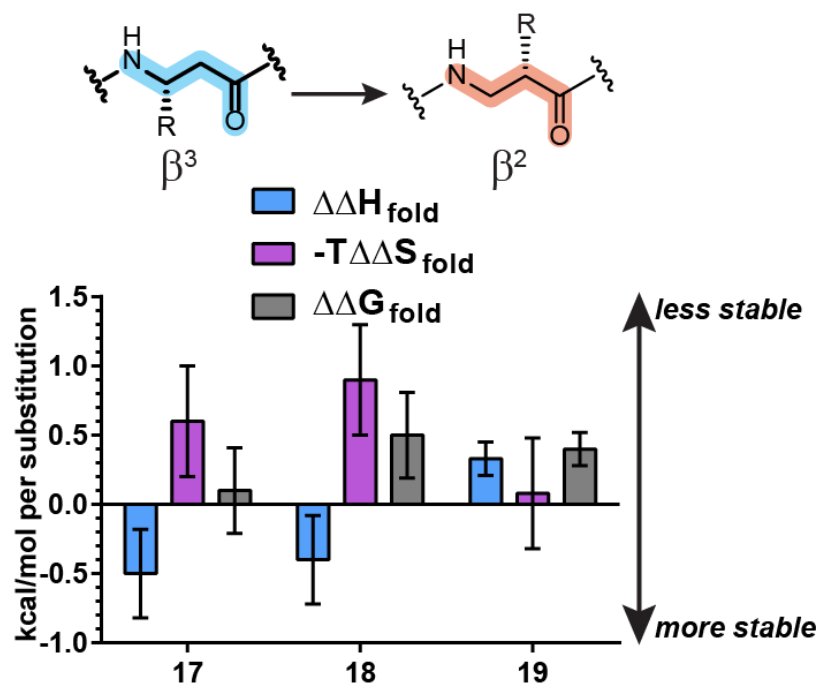
**Figure 54.** A) Overlay of helices from crystal structures of **6** (yellow), **17** (green), and **18** (gray). B) Close-up of  $\beta^3/\beta^2$ Ala<sub>24</sub> in **6** and **17**. C) Close-up of  $\beta^3/\beta^2$ -Asn<sub>35</sub> in **6** and **18**. Dotted line highlights a putative intraresidue orbital interaction in  $\beta^2$ -Asn<sub>35</sub> that may contribute favorably to folded stability of protein.

Overall,  $\beta^2$ -residues did not significantly alter the free energy of folding relative to  $\beta^3$ -residues as revealed by fits of combined thermal and chemical denaturation of proteins **17-19** (Figure 55). The enthalpy and entropy of folding for **17** and **18** were indistinguishable from that of the blueprint sequence **6**. Changes in  $\Delta G^\circ_{\text{fold}}$  were minimal except in the case of **19** where a 1.6 kcal/mol destabilization was observed. We attributed this to modification of Lys<sub>31</sub> in the center of the helix. The side-chain of Lys<sub>31</sub> packs against the aromatic face of the indole ring from Trp<sub>43</sub> from the sheet in the crystal structures of both wild-type GB1 and helix-modified analogue **6**. (see Chapter 2). Moving the lysine side-chain to the  $\beta^2$ -position would abolish this favorable interaction resulting in an enthalpic penalty. To discern the thermodynamic impact of losing this packing interaction of Lys<sub>31</sub> with Trp<sub>43</sub>, we bacterially expressed a GB1 Lys<sub>31</sub>→Ala mutant (sequence **20**). Point mutant **20** had a  $\Delta\Delta G^\circ_{\text{fold}}$  of +1.0 kcal/mol relative to an expressed

wild-type GB1 protein, approximately the same energetic difference in folded stabilities between **19** and **18**. It is notable that replacement of Lys<sub>31</sub> with Ala reduced the folded stability by approximately 20 % relative to the recombinant wild-type sequence (Table 10).

Contrary to our hypothesis regarding a potential role of  $n \rightarrow \pi^*$  interactions, replacement of  $\beta^3$ -Asn with a  $\beta^2$ -analogue did not significantly alter the  $\Delta H^\circ$  of folding relative to **6**. Relocation of the side-chain of  $\beta^3$ -Asn<sub>35</sub> to the C <sub>$\alpha$</sub>  atom as in **18** resulted in a small ( $\sim 0.5$  kcal/mol) enthalpic stabilization towards folding. As observed in a subset of chains in the crystal structure of **18** (Figure 54), the side-chain of  $\beta^2$ -Asn occupied a conformation suggestive of an intraresidue  $n \rightarrow \pi^*$  interaction, however, the distance between the participating atoms strongly diminishes the energetic stabilization towards folding. Any contribution of this transient interaction to the overall folded stability of the protein is at best marginal.

Changes to  $\Delta C_p^\circ$  and  $m$  for  $\beta^2$ -mutants **17-19** were small relative to **6** (Figure 55, Table 10), though some trends were observed.  $\beta^3 \rightarrow \beta^2$  substitution slightly decreased both changes in heat capacity and  $m$ , suggesting possibly small alterations to the denatured ensemble. Collectively, the data suggested that  $\beta^2$ -residues are isoenergetic relative to  $\beta^3$ -residues and key side-chain interactions should guide the choice of one over the other for  $\alpha$ -residue replacement in proteins.



**Figure 55.** Changes in thermodynamic parameters of folding for mutants **17-19** relative to **6**.

**Table 10.** Thermodynamic Parameters for the Folding Transitions of **6** and **17-19** at 298 K.

#	$\Delta H^\circ$ (kcal mol <sup>-1</sup> )	$T\Delta S^\circ$ (kcal mol <sup>-1</sup> )	$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$\Delta C_p^\circ$ (kcal mol <sup>-1</sup> K <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta\Delta G^\circ$ (kcal mol <sup>-1</sup> ) <sup>a</sup>
<b>6</b>	-18.3 ± 0.4	-15.2 ± 0.3	-3.1 ± 0.5	-0.53 ± 0.02	-2.48 ± 0.05	-
<b>17</b>	-18.8 ± 0.5	-15.8 ± 0.3	-3.0 ± 0.5	-0.45 ± 0.02	-2.25 ± 0.05	+0.1 ± 0.8
<b>18</b>	-18.7 ± 0.5	-16.1 ± 0.3	-2.6 ± 0.5	-0.48 ± 0.02	-2.36 ± 0.04	+0.5 ± 0.8
<b>19</b>	-17.0 ± 0.4	-15.5 ± 0.3	-1.5 ± 0.5	-0.48 ± 0.02	-2.45 ± 0.04	+1.6 ± 0.8
<b>20</b>	-22.7 ± 0.4	-18.5 ± 0.3	-4.2 ± 0.5	-0.67 ± 0.02	-1.77 ± 0.03	+1.0 ± 0.8 <sup>b</sup>
<b>GB1<sup>c</sup></b>	-23.0 ± 0.5	-17.9 ± 0.3	-5.1 ± 0.6	-0.57 ± 0.02	-1.71 ± 0.04	-

*a.* Relative to **6**.

*b.* Relative to biologically expressed wild-type GB1.

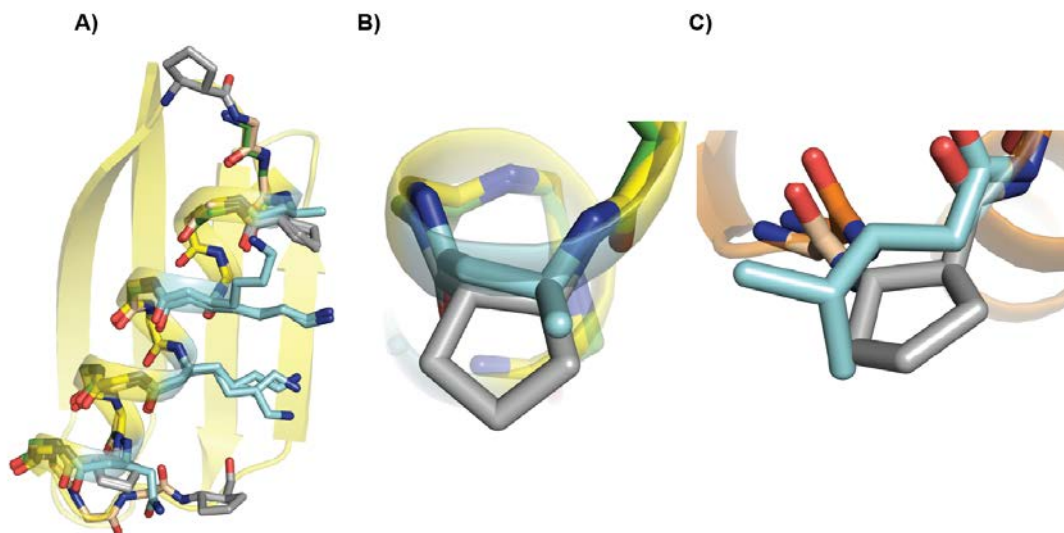
*c.* From recombinant expression in *E. coli*.

### 3.1.3 $\beta^3 \rightarrow \beta^{\text{cyc}}$ Substitution

The increased backbone flexibility of  $\beta^3/\beta^2$ -residues inspired prior experimental<sup>176-178</sup> work to design  $\beta$ -residues that “prepay” the chain entropic cost towards folding. Backbone cyclization through the side-chain atoms led to  $\beta^{\text{cyc}}$ -residues that were shown to promote helical folds in  $\beta$ -peptides.<sup>176</sup> In particular, (*S,S*)-ACPC was found to support  $\alpha$ -helix-like folds in mixed  $\alpha/\beta$  peptides.<sup>71, 110, 179</sup>

$\beta^{\text{cyc}}$ -Residue containing GB1 analogues **21-24** were synthesized to quantify the thermodynamic impact of backbone cyclization of  $\beta$ -residues in the helix of GB1 (Figure 56). Sequences **21** and **22** replace either the outer or inner pair of  $\beta^3$ -residues while **23** substitutes all four  $\beta^3$ -residues with the  $\beta^{\text{cyc}}$ -residue. To determine if there would be synergistic effects of multiple ACPC residues (**X**) as well as explore backbone cyclization in the loops of GB1, we replaced the final residue of each loop with residue **X** and grafted the most thermodynamically stable  $\beta^{\text{cyc}}$  backbone sequence (*vide infra*) to generate combined loop/helix mutant **24**.



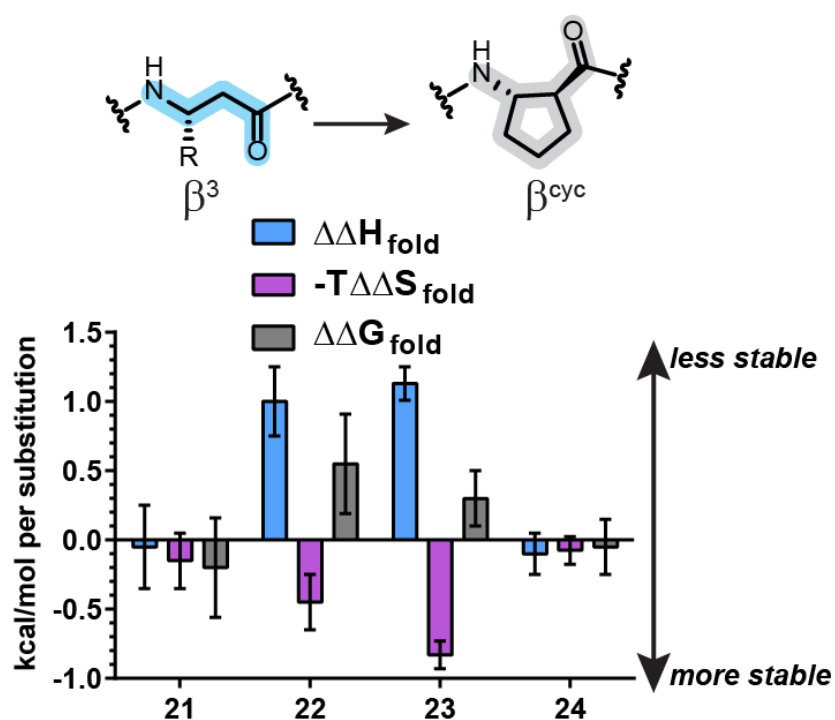


**Figure 57.** A) Overlay of helices from crystal structures of **6** (yellow), **21** (green), and **24** (ivory).  
 B) Close-up of helices in **6** and **21**. C) Close-up of loop in **5** (orange) and **24**.

Relative to  $\beta^3$ -residues,  $\beta^{\text{cyc}}$ -residues did not appreciably alter the folding enthalpy except in cases where residues with charged side-chains (lysine) were replaced, akin to the observations of  $\beta^3 \rightarrow \beta^2$  substitution (Figure 58, Table 11). As expected, fits of combined thermal and chemical melts of backbone cyclized mutants **21-24** showed a consistently lower  $\Delta S^\circ$  than all  $\beta^3$ -mutant **6** in accordance with design hypotheses. In cases where polar residues were replaced, the entropic stabilization afforded by backbone preorganization was surpassed by the enthalpic penalty of losing the polar contacts of Lys<sub>31</sub>. The combined loop/helix mutant **24** had a fold identical in stability to both **6** and **21**. Accounting for the energetic penalty of introducing  $\beta^3$ -residues into the loop of GB1 ( $\sim 0.3$  kcal/mol per replacement), the data suggested that  $\beta^3 \rightarrow \beta^{\text{cyc}}$  substitution in the loops was overall favorable towards folding.

The ability of the  $\beta^{\text{cyc}}$ -residues to preorganize the backbone was apparent from changes to  $\Delta C_p^\circ$  and  $m$ .  $\beta^3 \rightarrow \beta^{\text{cyc}}$  replacement consistently decreased the change in heat capacity and  $m$ .

Cyclic residues were anticipated to promote structure even in the denatured state based on restricted backbone conformational freedom. Due to the dependence of both  $\Delta C_p^\circ$  and  $m$  on  $\Delta ASA$ , this increase of residual structure would decrease the change in accessible surface area. This effect scaled with the number of  $\beta^{cyc}$ -residues in the helix with mutant **23** having the smallest  $\Delta C_p^\circ$  and  $m$  values.



**Figure 58.** Changes in thermodynamic parameters of folding for mutants **21-24** relative to helix mutant **6**.



**Table 11.** Thermodynamic Parameters for the Folding Transitions of **6** and **20-23** at 298 K.

#	$\Delta H^\circ$ (kcal mol <sup>-1</sup> )	$T\Delta S^\circ$ (kcal mol <sup>-1</sup> )	$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$\Delta C_p$ (kcal mol <sup>-1</sup> K <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta\Delta G^\circ$ (kcal mol <sup>-1</sup> ) <sup>a</sup>
<b>6</b>	-18.3 ± 0.4	-15.2 ± 0.3	-3.1 ± 0.5	-0.53 ± 0.02	-2.48 ± 0.05	-
<b>21</b>	-18.4 ± 0.5	-14.9 ± 0.3	-3.5 ± 0.6	-0.48 ± 0.02	-2.00 ± 0.05	-0.4 ± 0.8
<b>22</b>	-16.3 ± 0.3	-14.3 ± 0.3	-2.0 ± 0.4	-0.49 ± 0.02	-1.76 ± 0.03	+1.1 ± 0.8
<b>23</b>	-13.8 ± 0.3	-11.9 ± 0.3	-1.9 ± 0.4	-0.39 ± 0.02	-1.49 ± 0.03	+1.2 ± 0.8
<b>24</b>	-18.2 ± 0.5	-14.9 ± 0.3	-3.3 ± 0.6	-0.52 ± 0.02	-1.71 ± 0.04	-0.2 ± 0.8

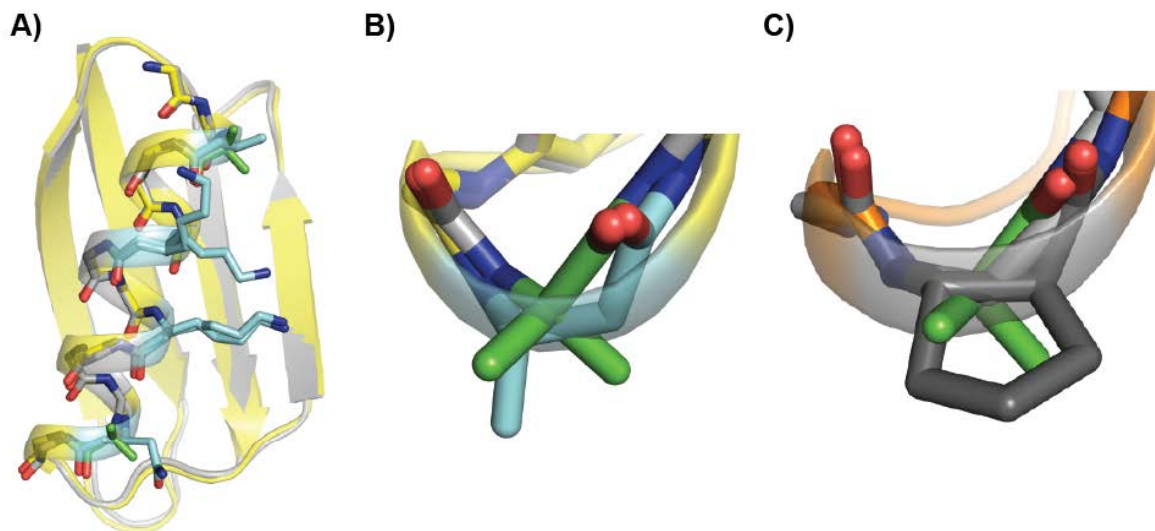
*a.* Relative to **6**.

### 3.1.4 $\beta^3 \rightarrow C_\alpha$ Methyl Substitution

$\beta$ -Residues can be thought of as backbone methylated analogues of  $\alpha$ -residues where the additional methylene unit is inserted between the carbonyl carbon and  $C_\alpha$ .  $C_\alpha$ -methyl residues are similarly backbone methylated analogues of  $\alpha$ -amino acids but instead place the methylene group between  $C_\alpha$  and  $H_\alpha$ . The prototypical  $C_\alpha$ -methyl residue Aib induces helical folding in peptides due to a restricted Ramachandran space where only helical conformations are significantly populated.<sup>180</sup> Studies on replacement of alanine with Aib in a helical subdomain of thermolysin found that Aib enhanced the thermostability of the folded protein by ~3 kcal/mol on average.<sup>36</sup> Oligomers of Aib are achiral and can equally adopt either left or right-handed  $3_{10}$  helices, a naturally occurring and narrower type of helix.<sup>35</sup> Incorporation of Aib into L- $\alpha$ -peptides stabilizes a right-handed  $\alpha$ -helical fold.<sup>35, 132</sup> Efforts to engineer a right-handed preference into Aib have resulted in chiral  $C_\alpha$ -methyl residues that show a strong preference



As predicted, Aib was well tolerated in the helix of GB1. **25** had a fold identical to that of **6** (RMSD = 0.4 Å) as shown in the crystal structure (Figure 60). Side-chain positions of **25** matched those of  $\beta^3$ -mutant **6** and  $\beta^{\text{cyc}}$ -mutant **21** even at locations substituted with Aib.

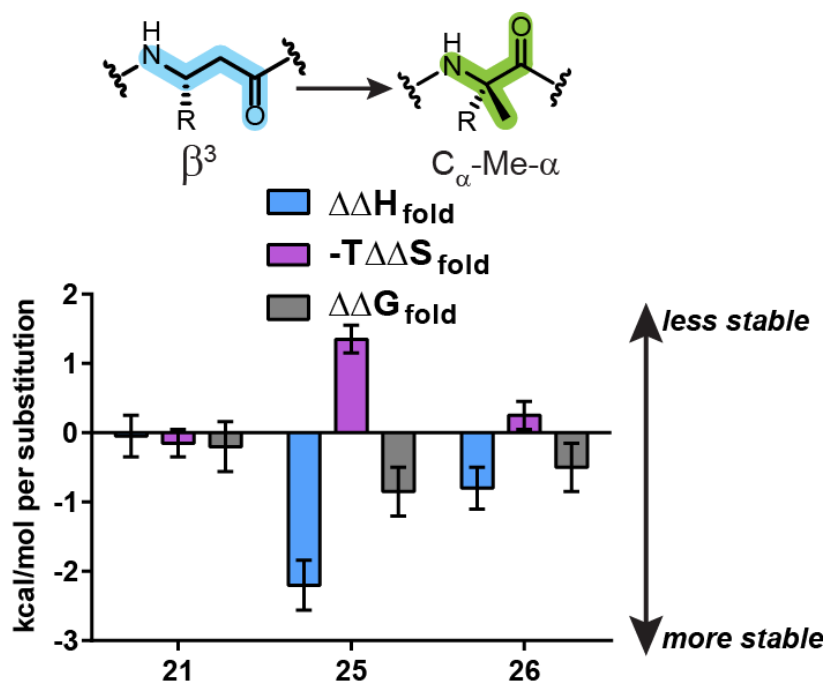


**Figure 60.** **A)** Overlay of helices from **6** (yellow) and **25** (gray). **B)** Close-up of Ala/Aib<sub>24</sub> in **6/25**. **C)** Close-up of ACPC/Aib<sub>35</sub> in **21** (orange)/**26**. Aib residues are colored green in all structures.

Combined thermal and chemical denaturation CD experiments of **25** and **26** supported the hypothesis that C<sub>α</sub>-methyl residues promote helical folds.  $\beta^3 \rightarrow \text{C}_\alpha\text{-Me}$  substitution consistently stabilized protein folding (Figure 61, Table 12). Relative to mutant **6**, C<sub>α</sub>-methylation was enthalpically favorable and entropically unfavorable towards folding. The magnitude of the changes to  $\Delta H^\circ$  and  $\Delta S^\circ$  was much greater (>2x) for replacement of a  $\beta^3$ -residue with Aib than the corresponding substitution with C<sub>α</sub>-Me-Val. Both Aib and C<sub>α</sub>-Me-Val do not have an extended backbone relative to  $\alpha$ -residues as opposed to  $\beta^{\text{cyc}}$ -residues. This feature of **25** and **26**

should retain any native enthalpically favorable towards folding orbital interactions and may explain the observed differences in  $\Delta H^\circ$ . It may be that the bulky isopropyl side-chain of C $_{\alpha}$ -methylated valine introduced steric clashes and/or required significantly more solvation in the folded state. Although  $\beta^{\text{cyc}}$ -residues were slightly more entropically favorable towards folding, the enthalpic folding stabilization afforded by C $_{\alpha}$ -methylation overcompensated for their reduced entropic benefit towards folding. The entropic penalty toward folding of Aib incorporation was larger than that of C $_{\alpha}$ -methyl valine. The lack of chiral preference in Aib may have increased the entropy of the unfolded state.

Introduction of C $_{\alpha}$ -methyl residues did not significantly alter  $\Delta C_p^\circ$  but did reduce  $m$  relative to **6**. This trend was observed for  $\beta^3 \rightarrow \beta^{\text{cyc}}$  replacement and was attributed to a reduction in  $\Delta\text{ASA}$  due to backbone preorganization. In this context, restricted backbone dihedral angles likely promoted folded structure, even in the unfolded state. The extent of reduction in  $m$  values compared to **6** was greater for C $_{\alpha}$ -methyl residues than  $\beta^{\text{cyc}}$ -residues suggesting a higher degree of residual structure and/or a less exposed nonpolar surface in the denatured ensemble. **26** had a larger value of  $m$  likely due to the exposure of bulky isopropyl groups in the folded state which increased the  $\Delta\text{ASA}$ .



**Figure 61.** Changes in thermodynamic parameters of folding for mutants **21**, **25**, and **26** relative to helix mutant **6**.

**Table 12.** Thermodynamics of Folding for Proteins **6**, **21**, **25**, and **26** at 298 K.

#	$\Delta H^\circ$ (kcal mol <sup>-1</sup> )	$T\Delta S^\circ$ (kcal mol <sup>-1</sup> )	$\Delta G^\circ$ (kcal mol <sup>-1</sup> )	$\Delta C_p$ (kcal mol <sup>-1</sup> K <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta\Delta G^\circ$ (kcal mol <sup>-1</sup> ) <sup>a</sup>
<b>6</b>	-18.3 ± 0.4	-15.2 ± 0.3	-3.1 ± 0.5	-0.53 ± 0.02	-2.48 ± 0.05	-
<b>21</b>	-18.4 ± 0.5	-14.9 ± 0.3	-3.5 ± 0.6	-0.48 ± 0.02	-2.00 ± 0.05	-0.4 ± 0.8
<b>25</b>	-22.7 ± 0.6	-17.9 ± 0.3	-4.8 ± 0.5	-0.48 ± 0.02	-1.57 ± 0.03	-1.7 ± 0.8
<b>26</b>	-19.9 ± 0.4	-15.8 ± 0.3	-4.1 ± 0.5	-0.46 ± 0.02	-1.72 ± 0.04	-1.0 ± 0.7

*a.* Relative to **6**.

## 3.2 CONCLUSIONS

Our dataset on the thermodynamics of protein-like oligomers with unnatural backbones revealed several interesting insights. Relative to the wild-type sequence,  $\beta^3$ -residues were enthalpically destabilizing towards folding when inserted into either the loops or helix of GB1. Based on the increase in backbone conformational freedom,  $\alpha \rightarrow \beta^3$  substitution was expected to increase the entropic penalty towards folding in a protein. Instead, incorporation of  $\beta^3$ -residues was entropically favorable towards folding. Based on observed changes to  $\Delta C_p^\circ$  and  $m$ , we attributed this effect to altered solvation of the unnatural backbone in the denatured ensemble. The entropic stabilization of  $\alpha \rightarrow \beta^3$  replacement was always outweighed by a larger enthalpic penalty towards folding.

We used a mutant of GB1 bearing  $\beta^3$ -residues in the helix as a reference for further backbone modifications. Introduction of  $\beta^2$ -residues into the helix of GB1 was overall isoenergetic relative to a  $\beta^3$ -analogue. Replacement of all four  $\beta^3$ -residues led to a mutant with  $\sim 1.5$  kcal/mol less folded stability. The loss of a favorable tertiary interaction due to the translocation of a side-chain compensated for any differences in folding free energies between  $\beta^3$  and  $\beta^2$ -residues. In designing unnatural oligomers, both classes of  $\beta$ -residues may be used, however, key side-chain interactions should guide the choice of one regioisomer over another. In this aspect, GB1 may not have been the best system to examine the thermodynamic effect of  $\beta^3 \rightarrow \beta^2$  substitutions. One other important consideration is the ease of monomer synthesis. Many  $\beta^3$ -residues bearing natural side-chains are commercially available. In contrast, the preparation of protected  $\beta^2$ -analogues of polar or charged amino acids is nontrivial and requires multi-step syntheses.<sup>188</sup>

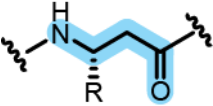
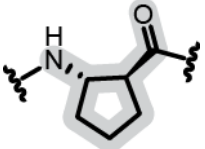
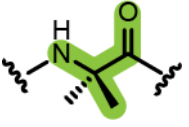
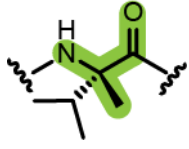
Cyclization of the  $\beta$ -residue is thought to prepay the entropic cost towards arranging the backbone into a folded conformation. Replacement of  $\beta^3$  residues with  $\beta^{\text{cyc}}$ -residues was at best energetically neutral.  $\beta^3 \rightarrow \beta^{\text{cyc}}$  substitution was entropically favorable to folding beyond that observed for  $\beta^3$ -residues. However, replacement of side-chains involved in tertiary interactions resulted in steep enthalpic folding penalties that yielded a net loss of folded stability. Our data suggested that  $\beta^{\text{cyc}}$  residues promote structure in the denatured ensemble as the physical basis for the favorable  $\Delta S^\circ$ . While  $\beta^{\text{cyc}}$ -residues may alter solvation of the backbone like  $\beta^3$ -residues, this is likely a minor contribution to the observed entropic stabilization toward folding.

As an alternate to  $\beta$ -residues,  $C_\alpha$ -methyl residues were expected to strongly promote helical folds based on restricted accessible dihedral angles. Structural and biophysical characterization of two distinct  $C_\alpha$ -methylated GB1 mutants revealed highly stable tertiary folds. In both mutants,  $\beta^3 \rightarrow C_\alpha\text{-Me}$  substitution was enthalpically favorable towards folding. Compared to  $\beta$ -residues,  $C_\alpha$ -methyl residues have one fewer backbone atom. The lengthened backbone of  $\beta$ -residues precludes their participation as acceptors for  $n \rightarrow \pi^*$  interactions, which may explain the more favorable folding enthalpies of the  $C_\alpha$ -methyl GB1 mutants. The  $\Delta H$  of helices containing Aib was found to be significantly greater than those with  $C_\alpha$ -methyl valine. The larger isopropyl chain of  $C_\alpha$ -methyl valine relative to a methyl group of Aib could introduce steric clashes with backbone and/or side-chain atoms in the folded state and result in enthalpic penalties towards folding.

Several trends were observed for the  $\Delta\Delta S_{\text{fold}}$  relative to  $\beta^3$ -residues. Both  $\beta^{\text{cyc}}$  and  $C_\alpha$ -methyl residues are expected to reduce the conformational freedom of the denatured ensemble and provide an entropic driving force towards folding.  $\beta^{\text{cyc}}$ -residues introduced a favorable entropic stabilization towards folding in GB1, yet analogous mutants containing  $C_\alpha$ -methyl

residues were entropically unfavorable towards folding. The restricted backbone dihedrals of C $_{\alpha}$ -methyl residues should promote helical structure in a similar manner to backbone cyclization of  $\beta$ -residues, however, there are clearly other contributing factors to the folding entropy of proteins. The smaller  $m$  values for C $_{\alpha}$ -methyl-residues compared to  $\beta^{\text{cyc}}$ -residues suggest that the former are stronger promoters of  $\alpha$ -helical structure in both unfolded and folded states. It has been previously observed that C $_{\alpha}$ -methylation reduces the configurational entropy of folded state when incorporated at flexible regions in a native protein structure.<sup>36</sup> A similar reduction in the entropy of the folded of GB1 may have resulted from C $_{\alpha}$ -methylation near the termini of the helix.

**Table 13.** Proposed Physical Bases for Observed Thermodynamic Changes to Folding Enthalpy and Entropy in Backbone-Modified GB1 Mutants Relative to  $\alpha$ -Residues

	<b>Acyclic <math>\beta</math>-Residues</b>	<b><math>\beta^{\text{cyc}}</math>-Residues</b>	<b>C<math>_{\alpha}</math>-Me Alanine (Aib)</b>	<b>C<math>_{\alpha}</math>-Me Valine</b>
<b>Observed Changes to</b>				
	+ retains side-chain	+ promotes helical fold	+ promotes helical fold	+ promotes helical fold
<b><math>\Delta H_{\text{fold}}</math></b>	- lengthens backbone (loss of orbital interactions)	- loss of side-chain - lengthens backbone		- may introduce steric clashes
<b><math>\Delta S_{\text{fold}}</math></b>	+ altered solvation  - flexible backbone	+ altered solvation + preorganized backbone + chiral	+ preorganized backbone  - achiral	+ preorganized backbone + chiral



In every thermodynamic dataset obtained, changes to folding enthalpy were at least partially offset by changes to folding entropy. This trend has been observed in several other protein systems and has been referred to as enthalpy-entropy compensation (EEC).<sup>189-191</sup> Although EEC in protein folding is a highly controversial subject, it is noteworthy that heterogeneous backbone oligomers that have tertiary folds also share this feature of natural proteins.

Future endeavors will include the use of unnatural residues to stabilize protein folds comparable to or better than that of  $\alpha$ -amino acids. The design of such residues will utilize the wealth of information on the thermodynamics of folding of proteins with unnatural backbones shown in this chapter. In particular, investigations of the effects of C $_{\alpha}$ -methylation of polar or charged  $\alpha$ -amino acids on protein folding will add to our thermodynamic dataset. Combining second generation helix, loop, sheet, and turn substitution strategies and measuring the changes to folding thermodynamics as well as proteolytic resistance will enhance the current field of protein mimetics. Application of the strategies developed here to other, more complex targets will enable a test of the robustness of the backbone modification methods.

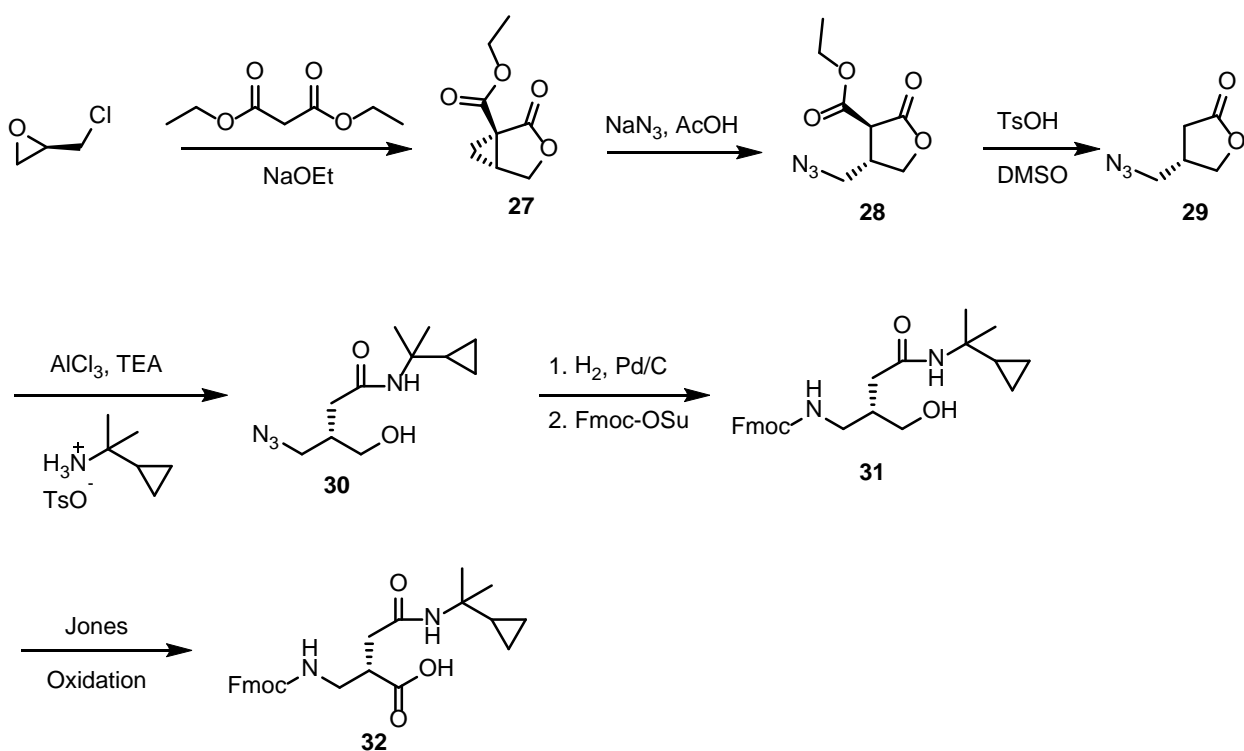
### **3.3 EXPERIMENTAL**

#### **3.3.1 General Information**

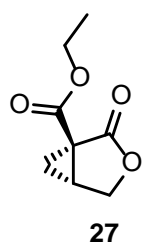
Solvents and all other reagents were purchased from Aldrich, Baker, EMD, or Fisher and used without further purification. HOBt was purchased from Anaspec Inc. HCTU, NovaPEG Rink Amide Resin, and Fmoc-protected  $\alpha$ -amino acids were purchased from Novabiochem. Fmoc-

protected  $\beta$ -amino acids were purchased from Aapptec. Flash column chromatography was performed using Silicycle SiliaFlash P60 (230-400 mesh) silica gel. Optical rotations were measured on a Perkin-Elmer 241 digital polarimeter with a sodium lamp at ambient temperature. NMR spectra were recorded on a Bruker Advance 400 spectrometer.

### 3.3.2 Synthesis of Fmoc- $\beta^2$ -Asn(Dmcp)-OH

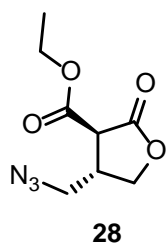


**Scheme 3.1** Synthesis of Fmoc- $\beta^2$ -Asn(Dmcp)-OH



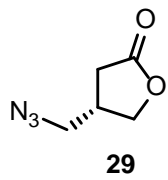
**Ethyl (1S, 5R)-2-oxo-3-oxabicyclo [3.1.0] hexane-1-carboxylate (27):** Compound **27** was synthesized according to a published protocol.<sup>192</sup> Sodium (1.15 g, 50 mmol)

was dissolved in ethanol (115 mL) under a N<sub>2</sub> atmosphere. The solution was cooled in an ice bath and diethyl malonate (9.2 mL, 60 mmol) was slowly added. After 10 minutes, *R*-epichlorohydrin (4.0 mL, 50 mmol) was slowly added over 20 minutes. The solution was refluxed overnight and concentrated. The residue was dissolved in water (100 mL) and the aqueous solution was extracted three times with 100 mL DCM. The organic layers were combined, dried with magnesium sulfate and concentrated. The concentrate was purified using FCC (20 % ethyl acetate in hexanes) to afford the product as a colorless oil (4.4 g, 26 mmol, 52 % yield). NMR data agreed with previously reported results.<sup>192</sup>



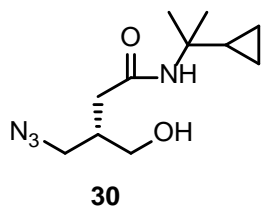
**Ethyl (3*S*,4*S*)-4-(azidomethyl)-2-oxotetrahydrofuran-3-carboxylate (28):**

Compound **28** was synthesized according to a published protocol.<sup>192</sup> A solution of **27** (4.4 g, 26 mmol), sodium azide (6.7 g, 104 mmol), glacial acetic acid (5.9 mL, 104 mmol), and TEA (72  $\mu$ L, 0.5 mmol) in anhydrous DMF (100 mL) was heated at 70 °C for 4 h under a N<sub>2</sub> atmosphere. The DMF was removed under reduced pressure and 200 mL saturated ammonium chloride solution was added. The aqueous solution was extracted three times with 200 mL DCM. The organic layers were washed three times with 100 mL saturated ammonium chloride, dried with magnesium sulfate, and concentrated to give the product as a colorless oil (3.9 g, 18.4 mmol, 71 % yield). NMR data agreed with previously published results.<sup>192</sup>



**(S)-4-(azidomethyl) dihydrofuran-2(3H)-one (29):** Compound **29** was synthesized according to a published protocol.<sup>193</sup> A solution of **28** (3.9 g, 18.4 mmol) and *p*-toluenesulfonic acid monohydrate (14.0 g, 74 mmol) in DMSO was

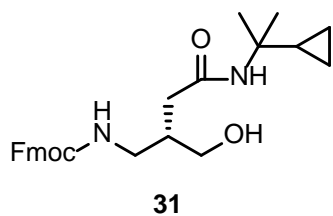
heated at 120 °C overnight. The reaction was cooled and 200 mL water was added. The solution was extracted four times with 200 mL ethyl acetate. The organic layers were washed three times with 100 mL saturated ammonium chloride, dried with magnesium sulfate, and concentrated to give the product as a colorless oil (1.5 g, 10.6 mmol, 57 % yield). NMR data agreed with previously published results.<sup>193</sup>



**(S)-4-azido-N-(2-cyclopropylpropan-2-yl)-3-(hydroxymethyl)**

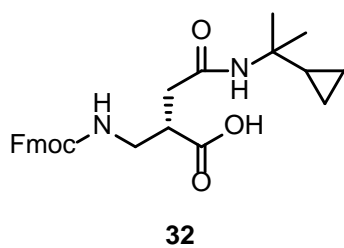
**butanamide (30):** To a solution of aluminum chloride (3.1 g, 23.3 mmol) suspended in anhydrous DCM (100 mL) cooled on ice under a N<sub>2</sub>

atmosphere was added TEA (6.5 mL, 47 mmol). The solution was stirred for 15 minutes before **29** (1.5 g, 10.6 mmol) and  $\alpha,\alpha$ -dimethyl-cyclopropylmethanamine *p*-toluenesulfonic acid (3.2 g, 11.7 mmol) were added. The solution was stirred overnight at ambient temperature and concentrated. The residue was dissolved in 200 mL ethyl acetate and washed twice with 100 mL 5% sodium bicarbonate solution, once with 100 mL saturated ammonium chloride, and once with 100 mL brine. The organic layers were dried with magnesium sulfate and concentrated to afford the product as a colorless oil (1.6 g, 6.7 mmol, 63 % yield).  $[\alpha]_D^{20} = -7.1$  ( $c = 1.1$  in CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.67 (s, 1 H), 3.71 (m, 2 H), 3.47 (m, 2 H), 2.25 (m, 3 H), 1.28 (m, 8 H), 0.43 (m, 4 H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$  170.0, 62.6, 52.9, 51.5, 37.1, 36.7, 23.0, 20.4, 0.0. HRMS  $m/z$  calculated for C<sub>11</sub>H<sub>21</sub>N<sub>4</sub>O<sub>2</sub> (M+H)<sup>+</sup> 241.1659; found 241.1660.



**(9H-fluoren-9-yl) methyl-(S)-(4-((2-cyclopropylpropan-2-yl)amino)-2-(hydroxymethyl)-4-oxobutyl) carbamate (31):** A solution of **30** (1.6 g, 6.7 mmol), palladium hydroxide on carbon (0.3 g), and palladium on carbon (0.3 g) in MeOH (75 mL) was stirred for

two days under a H<sub>2</sub> atmosphere. The reaction was filtered through Celite with MeOH and concentrated to afford the amino alcohol which was carried forward without further purification. To a solution of the amino alcohol in DCM (50 mL) was added FmocOSu (2.5 g, 7.4 mmol) and DIEA (4.8 mL, 27 mmol). The reaction was stirred for 4 h and concentrated. The residue was dissolved in 100 mL ethyl acetate, washed with 30 mL 5% sodium bisulfate, dried with magnesium sulfate, and concentrated. FCC (50 % ethyl acetate in hexanes) afforded the product as a white solid (1.5 g, 3.4 mmol, 50% yield).  $[\alpha]^{20}_D = +1.25$  (c = 0.8 in CHCl<sub>3</sub>). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 7.90 (d, *J* = 8 Hz, 2 H), 7.70 (d, *J* = 8 Hz, 2 H), 7.43 (t, *J* = 4 Hz, 2 H), 7.34 (t, *J* = 4 Hz, 2 H), 7.28 (s, 1 H), 7.23 (t, *J* = 4 Hz, 1 H), 4.46 (t, *J* = 4 Hz, 1 H), 4.30, (m, 3 H), 3.30 (m, 2 H), 3.00 (m, 2 H), 2.01 (m, 3 H), 1.27 (m, 1 H), 1.15 (s, 6 H), 0.28 (m, 4 H). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>) δ 169.9, 155.3, 142.8, 139.6, 126.5, 125.9, 124.1, 119.0, 64.0, 60.5, 51.2, 45.6, 40.1, 35.2, 23.6, 19.4, 0.0. HRMS *m/z* calculated for C<sub>26</sub>H<sub>33</sub>N<sub>2</sub>O<sub>4</sub> (M+H)<sup>+</sup> 437.2440; found 437.2427.

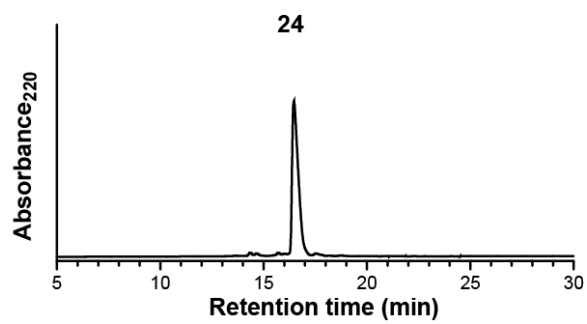
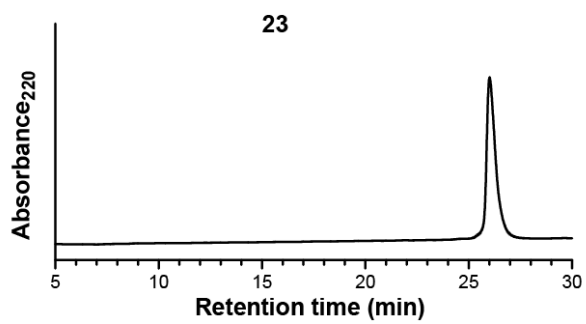
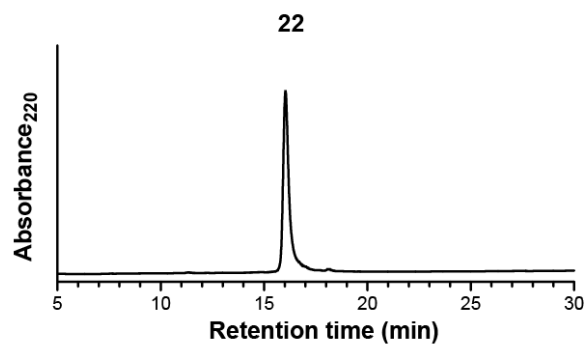
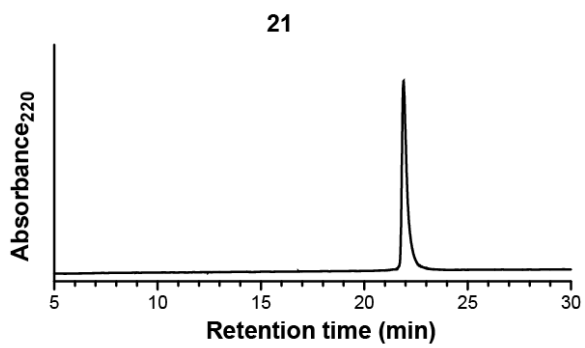
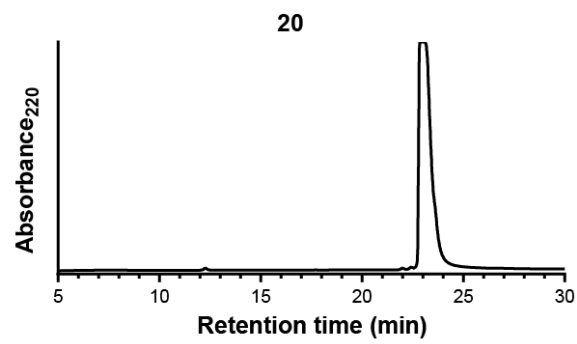
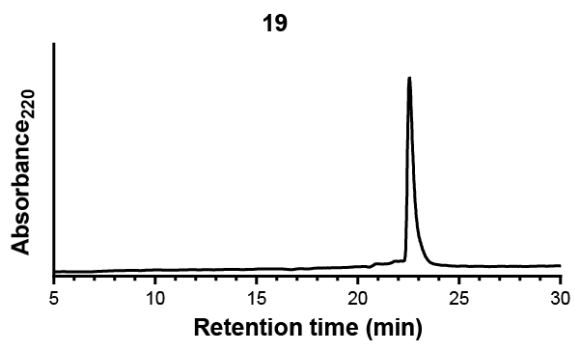
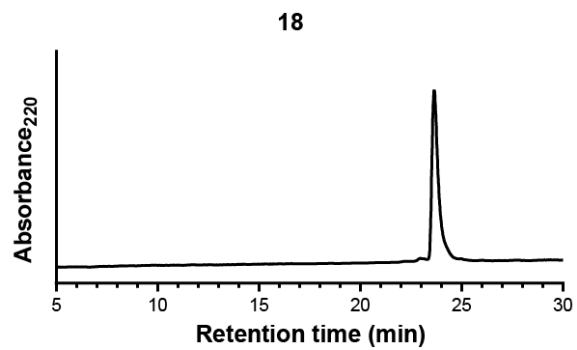
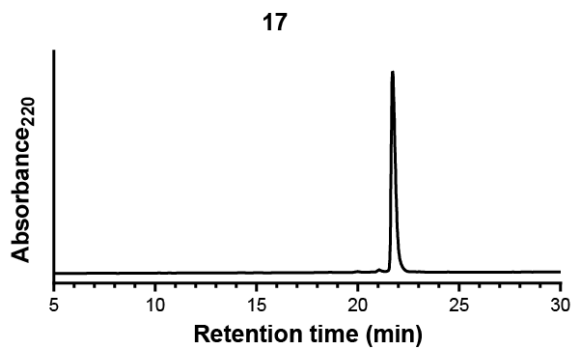


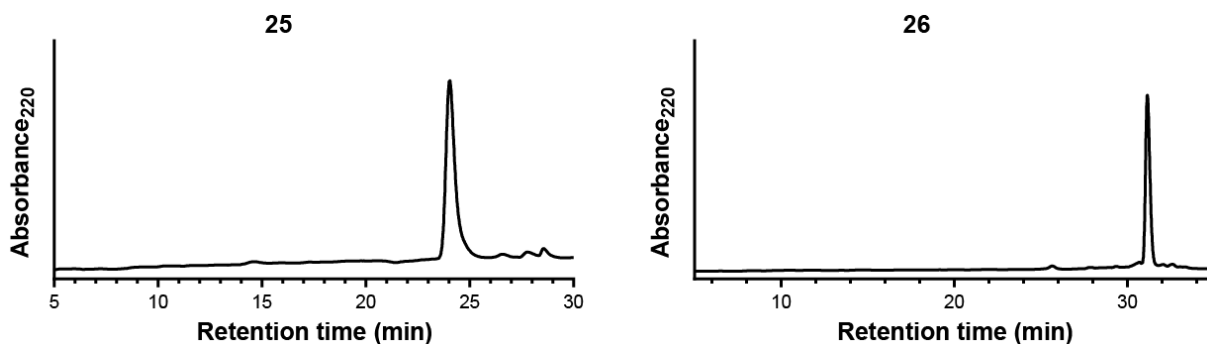
**Fmoc-(S)-β<sup>2</sup>-Asn(Dmcp)-OH (32):** To a stirred solution of sodium dichromate in water (1 M) was added concentrated sulfuric acid to a

final concentration of 4 M. The solution was then diluted with water to a final concentration of 0.5 M sodium dichromate and 2 M sulfuric acid. To a solution of **31** (2.2 g, 5.0 mmol) in acetone (78 mL) was added Jones reagent (12.5 mL). After 2.5 h, 20 mL isopropanol was added and the solution was stirred for 30 minutes. The reaction was diluted with 400 mL ethyl acetate, washed once with 200 mL 5 % sodium bisulfate, twice with 200 mL brine, dried with magnesium sulfate, and concentrated. The residue was purified by FCC (dry-loaded in 1 % TEA 50 % ethyl acetate in hexanes, eluted with 1 % acetic acid in ethyl acetate) to afford the product as a white solid (0.58 g, 1.3 mmol, 26 % yield).  $[\alpha]^{20}_D = -28.8$  ( $c = 1.0$  in  $\text{CHCl}_3$ ).  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-d}_6$ )  $\delta$  7.87 (d,  $J = 8$  Hz, 2 H), 7.68 (d,  $J = 8$  Hz, 2 H), 7.41 (t,  $J = 4$  Hz, 2 H), 7.34 (t,  $J = 4$  Hz, 2 H), 4.23 (m, 3 H), 3.19 (m, 2 H), 2.80 (m, 1 H), 2.22 (m, 2 H), 1.15 (m, 7 H), 0.26 (m, 4 H).  $^{13}\text{C NMR}$  (100 MHz,  $\text{DMSO-d}_6$ )  $\delta$  175.2, 170.3, 156.6, 144.4, 144.3, 141.2, 128.1, 127.5, 125.7, 120.6, 65.9, 52.8, 47.2, 42.4, 42.3, 36.1, 25.3, 25.2, 21.1, 1.64, 1.56. HRMS  $m/z$  calculated for  $\text{C}_{26}\text{H}_{29}\text{N}_2\text{O}_5$  (M-H) $^-$  449.2071; found 449.2081.

### 3.3.3 Protein Synthesis and Purification

Proteins **17-19**, **21-26** were synthesized using the automated synthesis protocol described in Section 2.5.2. Mutant **20** and a wild-type GB1 analogue were biologically expressed following published protocols<sup>194</sup> using a GB1 plasmid graciously provided by Timothy F. Cunningham and Sunil Saxena (University of Pittsburgh). Proteins were cleaved as described in Section 2.5.2. All proteins were purified by reversed-phase HPLC as described in Section 2.5.3.





**Figure 62.** Analytical HPLC chromatograms of purified proteins 17-26.

**Table 14.** MALDI-TOF MS Data for Proteins 17-24.

#	[M+H] <sup>+</sup> <i>m/z</i> (average)	
	Calculated	Observed
17	6235.8	6232.6
18	6235.8	6233.4
19	6235.8	6234.8
20	6164.7	6164.7
21	6244.8	6241.0
22	6173.6	6169.2
23	6182.7	6179.6
24	6252.9	6247.7
25	6191.7	6191.9
26	6247.8	6247.5

### 3.3.4 Circular Dichroism

CD measurements were performed on an Olis DSM17 Circular Dichroism Spectrometer in 1 mm quartz cells. Samples consisted of 40  $\mu$ M protein in 20 mM sodium phosphate buffer, pH 7 with varying concentrations of guanidinium hydrochloride. Thermal melts were monitored at 220 nm



over the range of 4 °C to 98 °C with 2 °C increments, a dead band of 0.5 °C, and a 2 min equilibration time at each temperature. All measurements were baseline corrected for buffer. Raw CD data were fit using Mathematica 8 (Wolfram) and equations reported previously,<sup>167</sup> summarized briefly below. The protein folding free energy ( $\Delta G$ ) at a given temperature ( $T$ ) and concentration of guanidinium ( $[Gdm]$ ) is given by Equation 3.1:

**Equation 3.1**

$$\Delta G = \Delta H^\circ - T\Delta S^\circ + \Delta C_p * (T - T_0 + T * \ln(\frac{T_0}{T})) - m * [Gdm]$$

where  $\Delta H^\circ$  and  $\Delta S^\circ$  are the folding enthalpy and entropy at a reference temperature  $T^\circ$ ,  $\Delta C_p$  the change in heat capacity, and  $m$  the dependence of the folding free energy on  $[Gdm]$ . The observed ellipticity ( $\theta_{obs}$ ) at a particular  $T$  and  $[Gdm]$  is given by Equation 3.2:

**Equation 3.2**

$$\theta_{obs} = \frac{(\theta_n + \theta_u * \exp(\frac{-\Delta G^\circ}{RT}))}{(1 + \exp(\frac{-\Delta G^\circ}{RT}))}$$

where  $\theta_n$  and  $\theta_u$  are the ellipticity of the folded and unfolded states. Based on literature precedent<sup>167-168</sup> and inspection of the raw data,  $\theta_n$  and  $\theta_u$  were assumed to vary linearly with  $T$  and  $[Gdm]$  according to Equations 3.3 and 3.4:

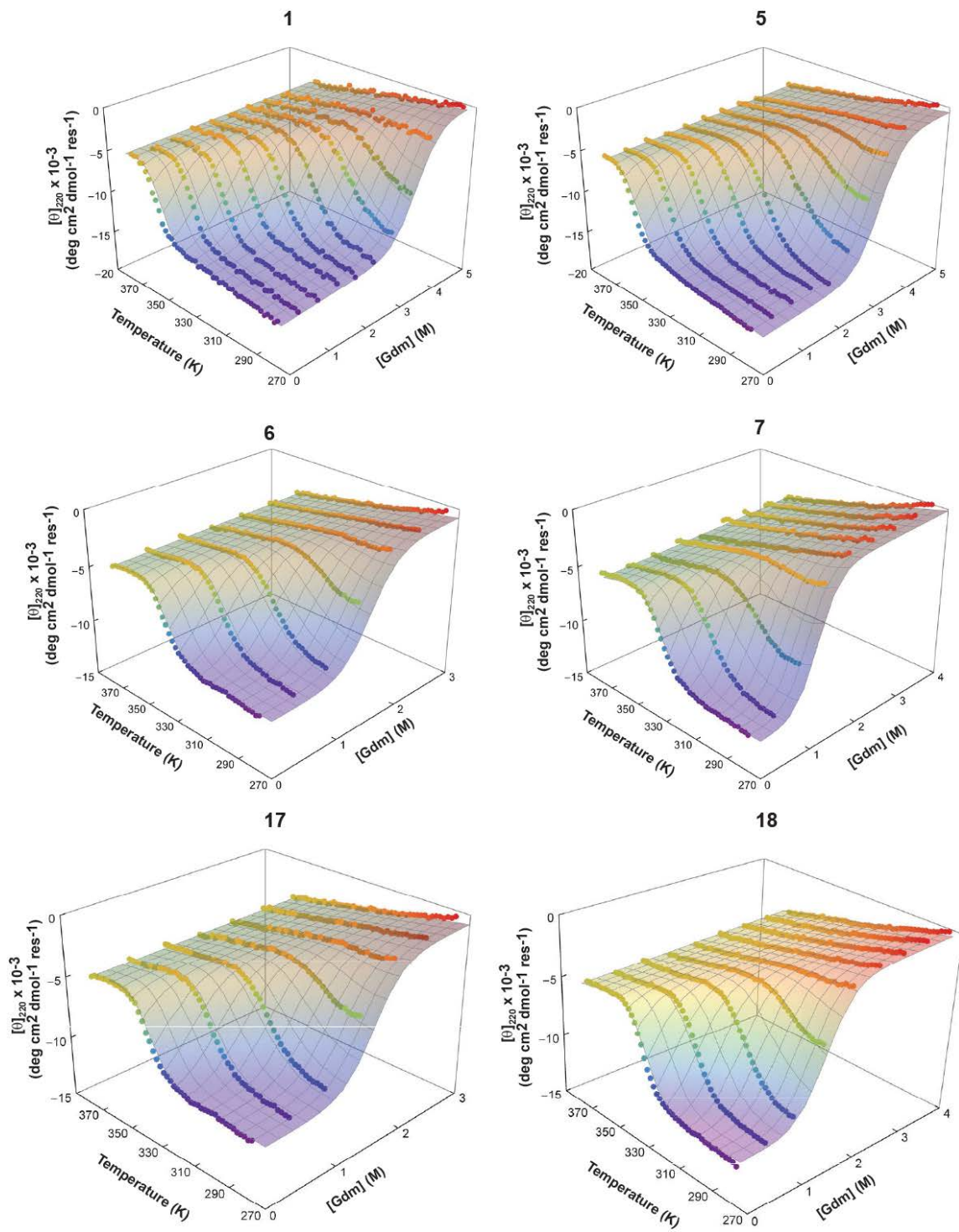
**Equation 3.3**

$$\theta_u = a + bT + c * [Gdm]$$

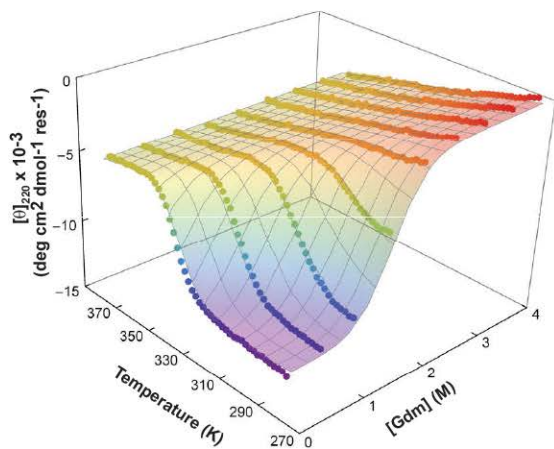
**Equation 3.4**

$$\theta_n = d + eT + f * [Gdm]$$

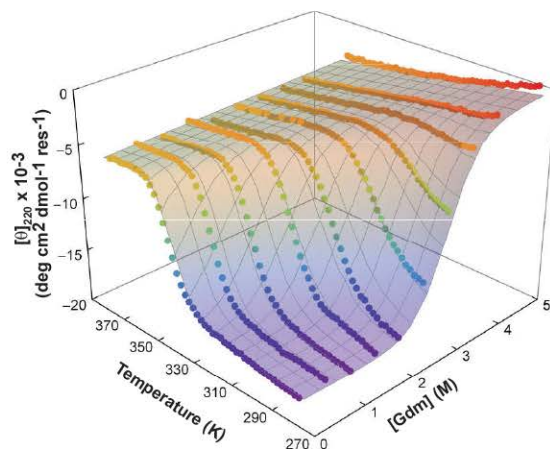
Some proteins (**5-7, 19, and 22-23**) lacked well-defined fully folded baselines as a function of [Gdm]. The parameter  $f$  was constrained to zero for these fits; this approximation did not change the observed thermodynamic values by more than 10% when applied to proteins with better defined baselines.<sup>168</sup> The folding/unfolding transitions of GB1 and its analogues were assumed to follow a two-state model,<sup>95</sup> and it was assumed that  $\Delta C_p$  and  $m$  do not vary over the range of experimental conditions.



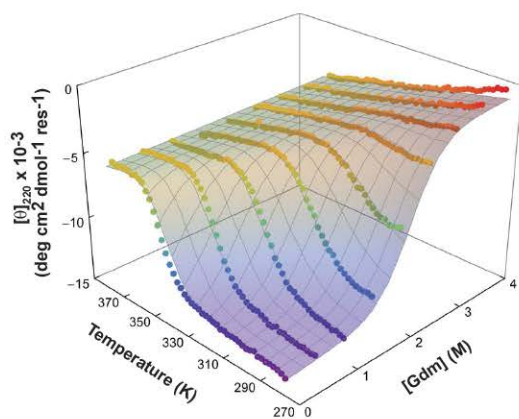
19



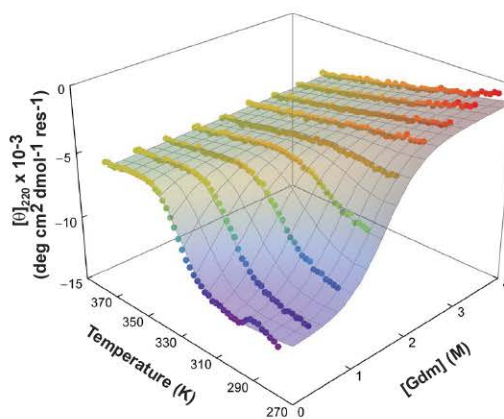
20



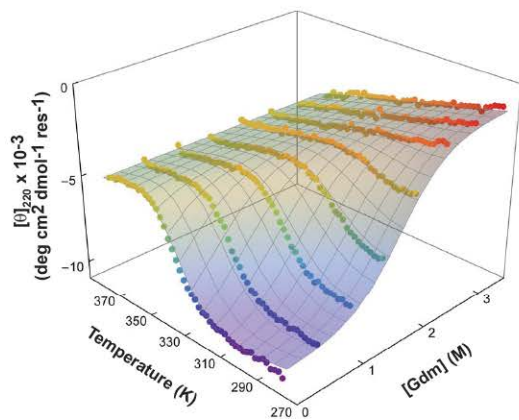
21



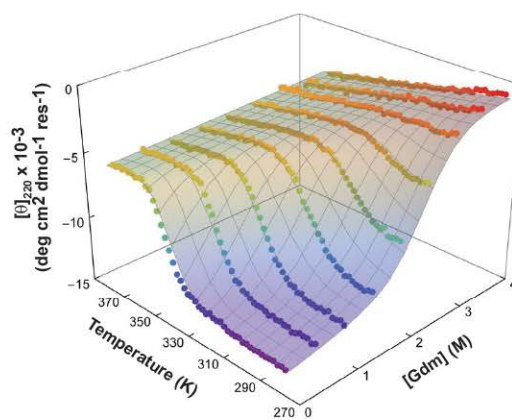
22

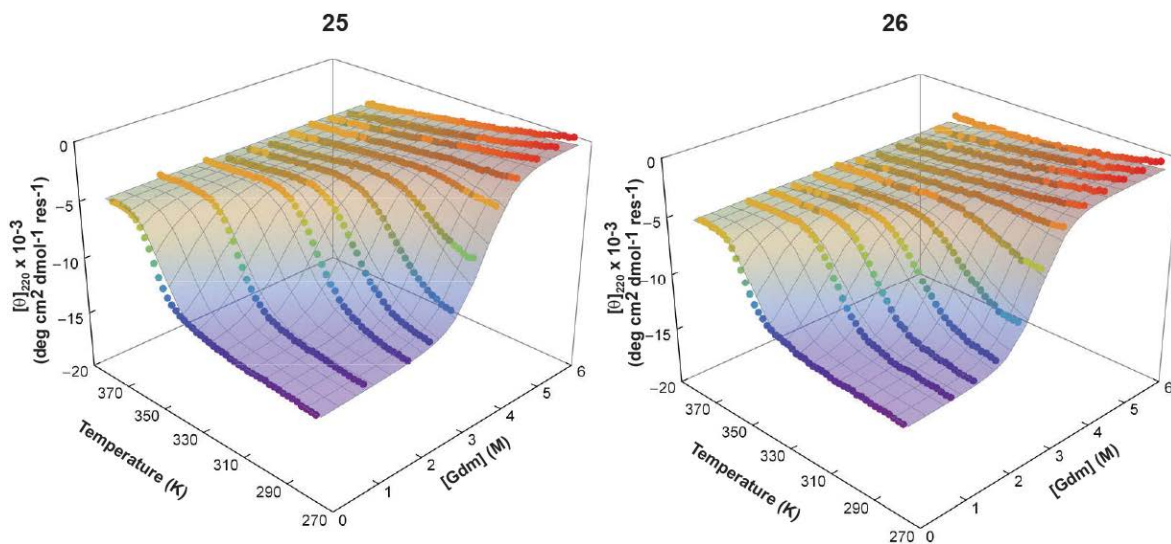


23



24





**Figure 63.** Fits of combined thermal and chemical melts of proteins **1, 5-7, 17-26.**

### 3.3.5 Crystallization of Proteins

Diffraction quality crystals of proteins **17, 18, 21, 24,** and **25** were obtained by the hanging drop vapor diffusion method as described in Section 2.5.8 using ~15 mg/mL solutions of protein in water. Table 15 summarizes the buffers for each protein.

**Table 15.** Crystallization Conditions for Proteins **17, 18, 21, 24,** and **25.**

<b>Protein</b>	<b>Buffer</b>	<b>Precipitant(s)</b>
<b>17</b>	200 mM NaOAc pH 4.6	20 % PEG 4000
<b>18</b>	100 mM Na Cacodylate pH 6.5	0.15 M (MgOAc) <sub>2</sub> , 20% PEG 4000
<b>21</b>	100 mM Na Citrate pH 5.6	20% isopropanol, 20% PEG 4000
<b>24</b>	100 mM HEPES pH 7.5	0.2 M MgCl <sub>2</sub> , 30% isopropanol
<b>25</b>	100 mM NaOAc pH 4.5	0.2 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 20 % PEG 4000

Data collection and structure refinement were performed by W. Seth Horne and Nathan Tavenor. Data collection methods used were identical to those described in Section 2.5.8 except for protein **18** which used a RAXIS HTC image plate. Analysis of the diffraction data with Xtrriage in the Phenix software suite suggested that the structure of **18** showed evidence of twinning. The final structure was solved with twin law operators (H, -K, -L) with a twin fraction of 0.27.

**Table 16.** X-ray Diffraction Data and Refinement Statistics for Proteins **17, 18, 21, 24, and 25.**

#	17	18	21 (4OZB)	24 (4OZC)	25
<b>Data Collection</b>					
<b>Unit cell dimensions</b> (Å, °)	$a = 92.8, b = 22.4$ $c = 65.3$ $\alpha = \gamma = 90, \beta = 134.1,$	$a = b = 51.9,$ $c = 96.4$ $\alpha = \beta = \gamma = 90$	$a = b = 65.9,$ $c = 21.9$ $\alpha = \beta = \gamma = 90$	$a = b = 83.8, c = 97.5$ $\alpha = \beta = \gamma = 90$	$a = 74.4, b = 73.4,$ $c = 79.4$ $\alpha = \gamma = 90, \beta = 99.4$
<b>Space group</b>	C2	P41	I41	I4122	C2
<b>Resolution (Å)</b>	23.44-1.95 (2.02-1.95)	51.95-1.80 (1.86-1.80)	23.31-2.20 (2.28-2.20)	23.28-2.00 (2.07-2.00)	41.5-2.15 (2.23-2.15)
<b>Total observations</b>	34,719	297,770	13,203	166,708	260,925
<b>Unique observations</b>	7,188	22,477	2,418	12,054	22,880
<b>Redundancy</b>	4.8 (3.2)	13.25 (13.18)	5.46 (5.42)	13.83 (13.81)	11.40 (3.04)
<b>Completeness (%)</b>	97.8 (90.4)	95.0	96.6 (98.8)	100 (100)	99.2 (93.2)
<b>I/σ</b>	18.5 (3.6)	25.3 (4.9)	11.0 (3.2)	16.7 (4.5)	15.9 (4.2)
<b>Rmerge (%)</b>	6.9 (15.2)	6.6 (40.5)	10.2 (26.4)	7.6 (39.6)	13.7 (23.9)
<b>Refinement</b>					
<b>Resolution (Å)</b>	23.44-1.95	51.95-1.80	23.31-2.20	23.28-2.00	41.15-2.15
<b>R (%)</b>	19.9	19.9	23.3	19.8	21.7
<b>Rfree (%)</b>	23.0	21.7	26.0	21.0	25.2
<b>Avg. B factor (Å<sup>2</sup>)</b>	22.1	27.3	38.1	39.4	25.5
<b>RMSD</b>					
<b>Bonds (Å)</b>	0.005	0.006	0.007	0.007	0.005
<b>Angles (°)</b>	1.04	1.13	1.14	1.05	1.02

## 4.0 IMPROVEMENTS TO SHEET MIMICRY STRATEGIES

The results in this chapter have been published in:

G.A. Lengyel, Z.E. Reinert, B.D. Griffith, W.S. Horne. "Comparison of backbone modification in protein  $\beta$ -sheets by  $\alpha \rightarrow \gamma$  residue replacement and  $\alpha$ -residue methylation." *Organic & Biomolecular Chemistry*. **2014**, *12*, 5375-5381.

All Fmoc-protected  $\gamma$ -residues were synthesized by George Lengyel. Synthesis of *N*-methyl GB1m2A mutants was performed by Brian Griffith. Analysis of 2D-NMR data was performed with Brian Griffith and George Lengyel. Synthesis and purification of GB1 mutants with  $\gamma$ -residues was performed by George Lengyel.

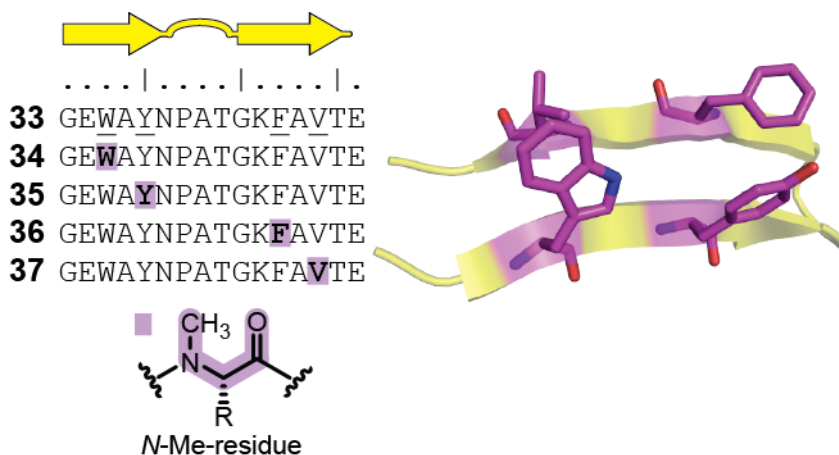
In Chapter 2, it was observed that *N*-methyl residues are tolerated in the sheets of GB1 with a small penalty towards folding. The following work details experiments probing the source of this destabilization in a model hairpin sequence as well as employing second generation sheet mimicry strategies in the full-length protein.



## 4.1 N-METHYL MODEL HAIRPIN INVESTIGATIONS

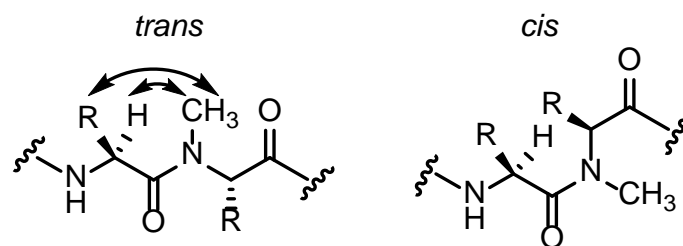
Our work in sheet mimicry with unnatural backbones yielded several insights into our first generation design strategies. Mixed  $\alpha/\beta$  backbones can form hairpins in aqueous solution in small oligomers but are prohibitively destabilizing towards folding to be used in a larger sheet system such as the tertiary fold of GB1. As a more conservative approach, we replaced one residue in each of the outer strands of GB1 with an *N*-methyl analogue. The resulting mutant had a folded stability near that of wild-type ( $\Delta\Delta G^\circ_{\text{fold}} = 1.0$  kcal/mol). Given the subtle backbone modification relative to the parent sequence, we were surprised by the observed energetic penalty.

To probe the small destabilization to folding resulting from incorporation of *N*-methyl residues, we used a model hairpin sequence derived from the C-terminal strands of GB1 referred to as GB1m2A<sup>120</sup> (**33**) to design several sequences exploring *N*-methyl substitution (**34-37**, Figure 64). In each modified sequence, one of four residues bearing a nonpolar side-chain was replaced with an *N*-methyl analogue. These four side-chains form the hydrophobic core that helps to stabilize the hairpin fold of **33**. We reasoned that any small changes introduced by *N*-methylation would be most apparent at these positions. Importantly, these four positions also do not participate in any interstrand backbone N-H hydrogen bonds in the folded state that would be abolished by the modifications.



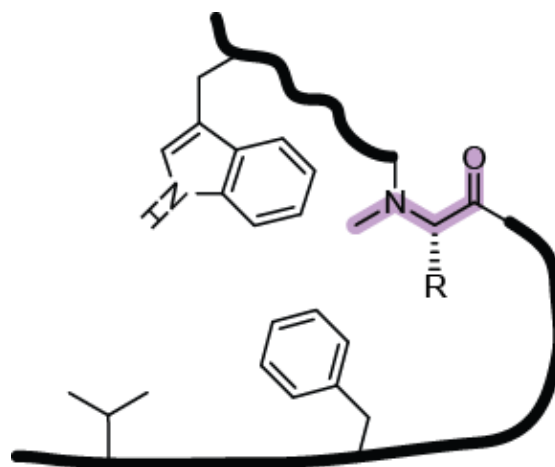
**Figure 64.** Sequences of model hairpin **33** and *N*-methyl mutants **34-37**. An NMR solution structure of GB1m2A<sup>120</sup> with the key hydrophobic core residues colored purple is shown on the right.

Compared to GB1, model hairpin **33** possessed two main advantages that made it amenable to the current study. First, the sequence of GB1m2A is much shorter than that of the parent protein GB1 and chemical synthesis requires less time and material investment. Second, changes to the stability of the hairpin fold of **33** are readily quantified by 2D NMR analysis.<sup>120</sup> Finally, *N*-methyl residues are known to exhibit *cis/trans* isomerization about tertiary amides in solution.<sup>195</sup> The *cis* and *trans* amide conformations of *N*-methylated residues in peptides **33-37** should be distinguishable by 2D NMR methods.



**Figure 65.** Rationale of *trans* amide identification based on NOE contacts observed in 2D NMR experiments. Black arrows indicate key observed NOE contacts.

We performed <sup>1</sup>H homonuclear COSY, TOCSY, and NOESY experiments on peptides **33-37** to examine the effect of *N*-methylation on the backbone of **33** (Table 17). We gauged the changes to folded populations through differences in Gly H<sub>α</sub>/H<sub>α'</sub> proton chemical shifts, a well-established method in hairpin-forming sequences.<sup>120</sup> As observed in the larger sheet system of GB1, *N*-methyl substitution repeatedly destabilized the hairpin fold of **33**. The ratio of *cis:trans* amide conformers was assigned through integration of a well resolved Asn<sub>6</sub> TOCSY peak. The *cis/trans* isomerization of the tertiary amides complicated the interpretation of the NMR spectra of each backbone-modified mutant. We assigned the *trans* isomer based on analysis of the NOESY spectra, which revealed contacts between the backbone *N*-methyl group and side-chain and H<sub>α</sub> protons of the preceding residue (Figure 65).



**Figure 66.** Cartoon of how a *cis* amide could severely destabilize a hairpin fold by displacing a critical hydrophobic side-chain from the core.

The position of backbone *N*-methylation affected the folded populations of the *trans* isomer (Table 17). The presence of an *N*-methyl *trans* amide at either of the outer positions (**34** and **37**) did not appreciably alter the folded population relative to **33**. In contrast, *N*-methyl substitution closer to the  $\beta$ -turn led to considerably destabilized ( $\sim 0.8$  kcal/mol) hairpins. The *cis* amide at any of the four positions consistently showed very small differences in Gly  $H_{\alpha}/H_{\alpha'}$  chemical shifts corresponding to a highly destabilized fold. This observation is reasonable given the probable outcome with respect to folding of a *cis* amide bearing a critical core hydrophobic side-chain (Figure 66). What is perhaps less obvious is why the *trans* amides near the center of the hairpin in peptides **35** and **36** had significantly smaller folded populations relative to the model hairpin.

**Table 17.** Folding Thermodynamics of Hairpin Peptides **33-37** from NMR Measurements.<sup>a</sup>

#	$\Delta\delta$ Gly <sub>10</sub> H $\alpha$ /H $\alpha'$ (ppm)	Fraction Folded (%)	$\Delta G_{\text{fold}}$ (kcal/mol)	$\Delta\Delta G_{\text{fold}}$ vs <b>33</b> (kcal/mol)
<b>33</b>	0.20	65	-0.3	-
<b>34<sup>b</sup></b>		37	+0.3	+0.6
34 <sub>trans</sub> (60 %)	0.19	61	-0.3	+0.0
34 <sub>cis</sub> (40 %)	0.09			
<b>35<sup>b</sup></b>		19	+0.8	+1.1
35 <sub>trans</sub> (76 %)	0.09	30	+0.5	+0.8
35 <sub>cis</sub> (24 %)	0.00			
<b>36<sup>b</sup></b>		29	+0.5	+0.8
36 <sub>trans</sub> (65 %)	0.12	38	+0.3	+0.6
36 <sub>cis</sub> (35 %)	0.00			
<b>37<sup>b</sup></b>		55	-0.1	+0.2
37 <sub>trans</sub> (87 %)	0.20	63	-0.3	+0.0
37 <sub>cis</sub> (13 %)	0.12			

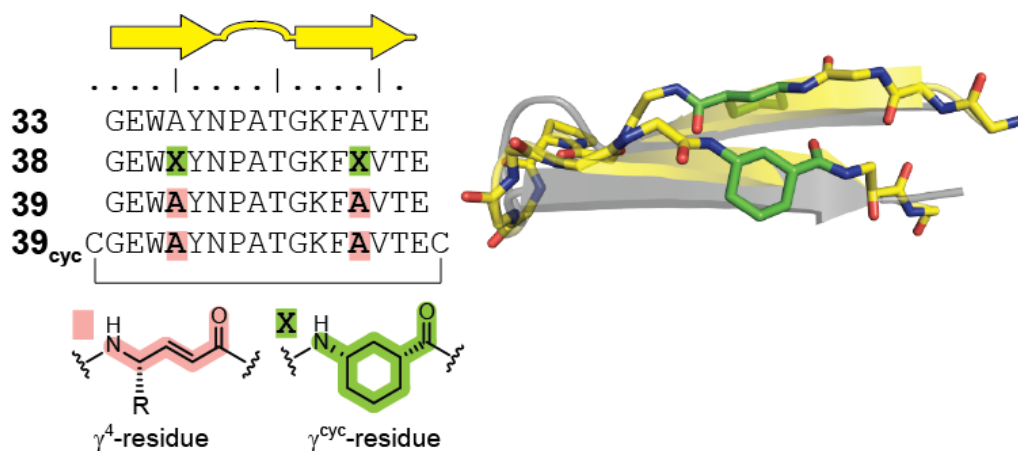
a. NMR carried out at 5 °C in pH 6.3 phosphate buffer. Assuming a 0.01 ppm uncertainty in measured Gly H $\alpha$ /H $\alpha'$  separation, error propagation estimates uncertainties of 5 % for fraction folded and 0.2 kcal/mol for  $\Delta G_{\text{fold}}$  and  $\Delta\Delta G_{\text{fold}}$ .

b. Overall folded population calculated as product of the fraction of peptide in the *trans* amide configuration and fraction folded for *trans* isomer.

Prior computational work has shown that *N*-methyl residues have less backbone conformational freedom than  $\alpha$ -residues.<sup>196</sup> Specifically, one region of the Ramachandran plot that becomes less energetically favorable corresponds to conformations typical of strands in an antiparallel  $\beta$ -sheet.<sup>197</sup> This observation likely explains the somewhat destabilized *trans* amide hairpins of **35**, **36**, and GB1 sheet mutant **10**. Further supporting this rationale is the marked unfavorable incorporation of *N*-methylation of residues near the turn of **33**. Due to the restricted backbone space of *N*-methyl residues, the increased configurational freedom required of turn segments resulted in a larger energetic penalty towards backbone modification than the outer positions.

## 4.2 SECOND GENERATION GB1 SHEET MUTANTS

As demonstrated earlier in Chapter 2, model hairpin sequence **33** has served as the basis of comparison for several backbone modification strategies including  $\alpha \rightarrow \beta$ <sup>119-120</sup> and  $\alpha \rightarrow N\text{-Me-}\alpha$  substitutions. Prior work in our lab examined an alternate design using a  $\gamma^{\text{cyc}}$ -residue derived from *cis*-3-aminocyclohexane carboxylic acid (ACC).<sup>198</sup> When incorporated into the sequence of **33**, the resulting mutant retained the hairpin fold of the parent sequence (**38**, Figure 67). Despite the increased backbone length, the  $\gamma^{\text{cyc}}$ -modified hairpin had a larger folded population (83 %) than the all  $\alpha$ -peptide **33** (67 %).<sup>198</sup>



**Figure 67.** Sequences of GB1m2A hairpin **33** and mutants **38** and **39** with NMR solution structures of **33** (gray) and **38** (yellow) overlaid. Data from reference 198.

Inspired by the success of  $\gamma^{\text{cyc}}$ -residues for sheet mimicry in hairpin, we looked to apply this strategy in the tertiary fold of GB1. However, the side-chains in larger proteins are often

involved in critical interactions necessary to stabilize the folded state. In prior work, a model peptide containing two  $\alpha$ ,  $\beta$ -unsaturated  $\gamma$ -residues bearing a side-chain at  $C_\gamma$  (referred to as a  $\gamma^4$ -residue) was shown to adopt a hairpin fold in organic solvent.<sup>199-200</sup> We designed and synthesized a mutant of GB1m2A with two  $\alpha \rightarrow \gamma^4$  substitutions to explore the ability of this residue type to promote hairpin folds in aqueous solutions (**39**, Figure 67).

**Table 18.** Folding Thermodynamics of Hairpin Peptides **33**, **38**, and **39** from NMR Measurements.<sup>a</sup>

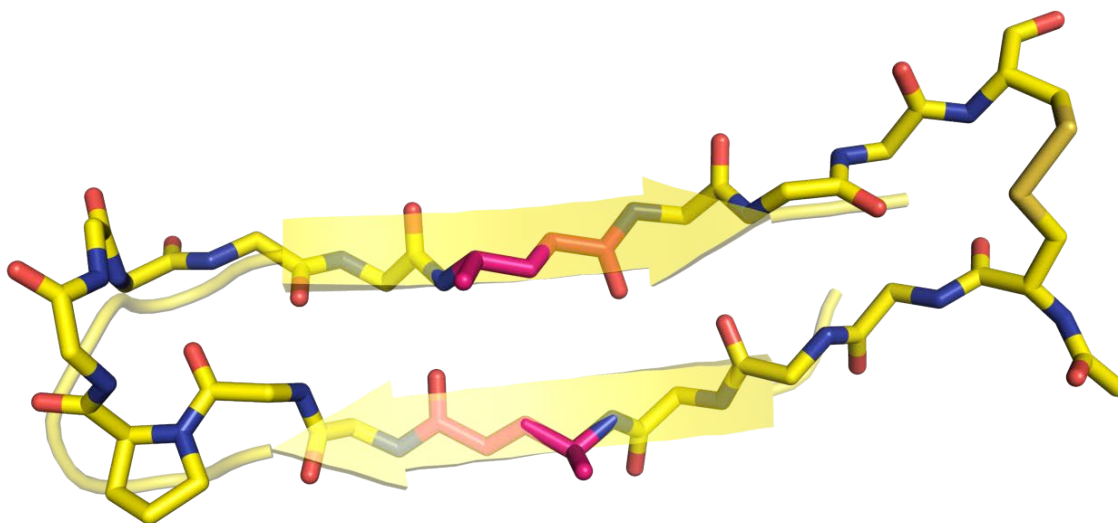
#	$\Delta\delta$ Gly <sub>10</sub> H $\alpha$ /H $\alpha'$ (ppm)	Fraction Folded (%)	$\Delta G_{\text{fold}}$ (kcal/mol)	$\Delta\Delta G_{\text{fold}}$ vs <b>33</b> (kcal/mol)
<b>33</b>	0.20	65	-0.3	-
<b>38<sup>b</sup></b>	0.26	83	-0.9	-0.6
<b>39</b>	0.12	39	+0.2	+0.5

*a.* NMR carried out at 5 °C in pH 6.3 phosphate buffer. Assuming a 0.01 ppm uncertainty in measured Gly H $\alpha$ /H $\alpha'$  separation, error propagation estimates uncertainties of 5 % for fraction folded and 0.2 kcal/mol for  $\Delta G_{\text{fold}}$  and  $\Delta\Delta G_{\text{fold}}$ .

*b.* Data from reference 198.

The folding propensity of **39** was characterized by 2D NMR experiments analogous to those used for *N*-methyl mutants **34-37** (Table 18). Whereas the  $\gamma^{\text{cyc}}$ -residues stabilized the hairpin fold,  $\gamma^4$ -mutant **39** destabilized the sequence-encoded fold of GB1m2A by an equal amount (0.6 kcal/mol). Due to the low folded population of **39**, we prepared a mutant bearing a Cys residue at each terminus (**39<sub>cyc</sub>**) to use for NMR solution structure determination. When the two terminal thiols are oxidized, the resulting disulfide stabilizes a hairpin fold. **39<sub>cyc</sub>** formed a hairpin very similar to that of **33** (Figure 68). A comparison of the structures of  $\gamma^{\text{cyc}}$  and  $\gamma^4$ -

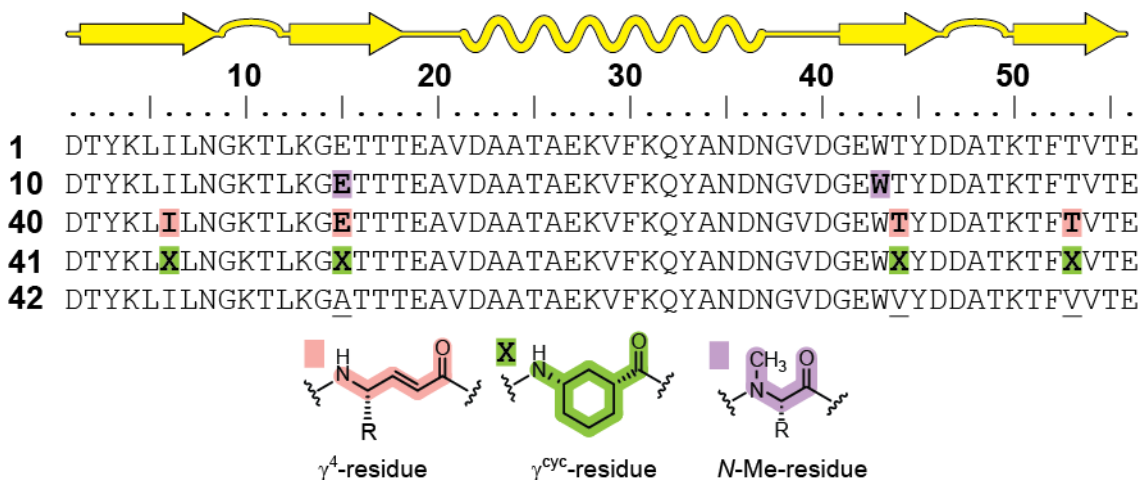
residues suggested that the additional rotatable bond present in the acyclic  $\gamma$ -residue likely accounts for the difference in observed folded populations of peptides **38** and **39**.



**Figure 68.** Solution structure of **39<sub>cyc</sub>** from NMR analyses.  $\gamma^4$ -Residues are colored pink.

Having shown that  $\gamma$ -residues are tolerated in model hairpin systems, we applied our  $\gamma^{\text{cyc}}$  and  $\gamma^4$ -based design strategies to the sheet of GB1 (**40** and **41**, Figure 69). Four  $\alpha$ -residues in the sheet of GB1 were replaced with the corresponding  $\gamma^4$ -analogue in **40** or an ACC-derived  $\gamma^{\text{cyc}}$ -residue (**X**) in **41**. To discern the thermodynamic cost toward folding of replacing polar residues in the sheet of GB1 with a hydrophobic  $\gamma^{\text{cyc}}$ -residue from any changes to the backbone of the protein we synthesized a point mutant of GB1 (**42**) that has three polar residues in the sheet replaced with nonpolar  $\alpha$ -residues.

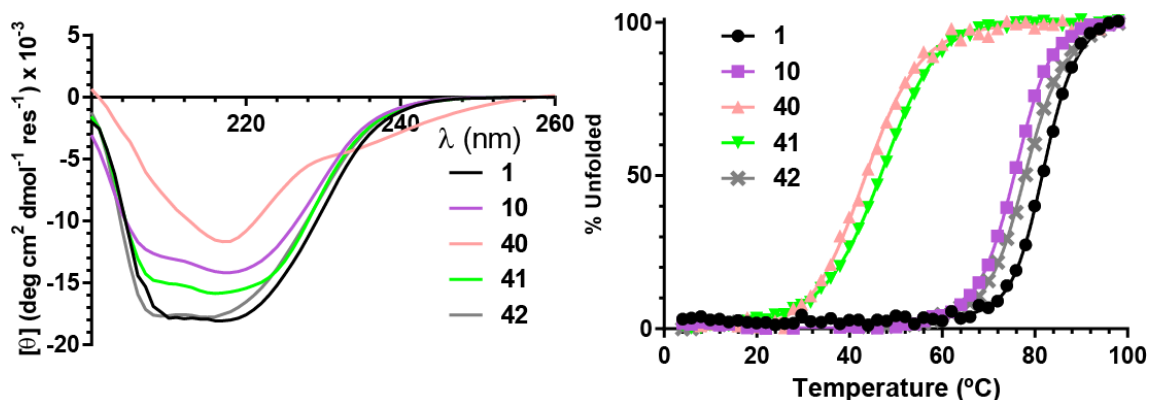




**Figure 69.** Sequences of wild-type **1**, *N*-methyl sheet mutant **10**, and sheet mutants **40-42**.

Underlined residues indicate side-chain changes relative to **1**.

CD scans of **40** and **41** suggested that both the vinylogous  $\gamma^4$  and  $\gamma^{\text{cyc}}$ -residues had an altered fold relative to GB1 (Figure 70). In particular, **40** had a CD spectrum that differed significantly in magnitude and shape compared to the other mutants and the wild-type protein. However, as seen in earlier work (Chapter 2), changes in CD scans from unnatural backbone modifications do not necessarily correlate with alterations in folded structure. Unnatural secondary structures such as  $\alpha/\beta^3$ -helices have intrinsically different CD signatures relative to natural analogues. Similarly,  $\gamma$ -residues may change the CD spectra of natural proteins. Both  $\alpha \rightarrow \gamma^4$  and  $\alpha \rightarrow \gamma^{\text{cyc}}$  substitutions destabilized the fold of GB1 by  $\sim 1.5$  kcal/mol per replacement (Table 19). This was a surprising result for **41** as the  $\gamma^{\text{cyc}}$ -residues were expected to be stabilizing to the sheet fold. The loss of polar side-chains as in **42** only partially accounted for the observed difference in free energies of **41** and wild-type **1**.



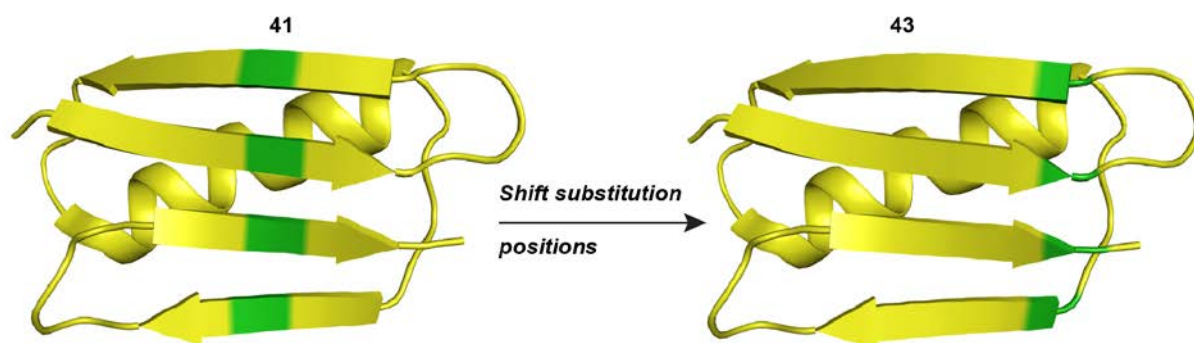
**Figure 70.** CD scans and thermal melts of wild-type **1** and sheet mutants **10** and **40-42** in 20 mM phosphate buffered water pH 7.

**Table 19.** Folding Thermodynamics of Proteins **1**, **10**, and **40-42** from CD Measurements.

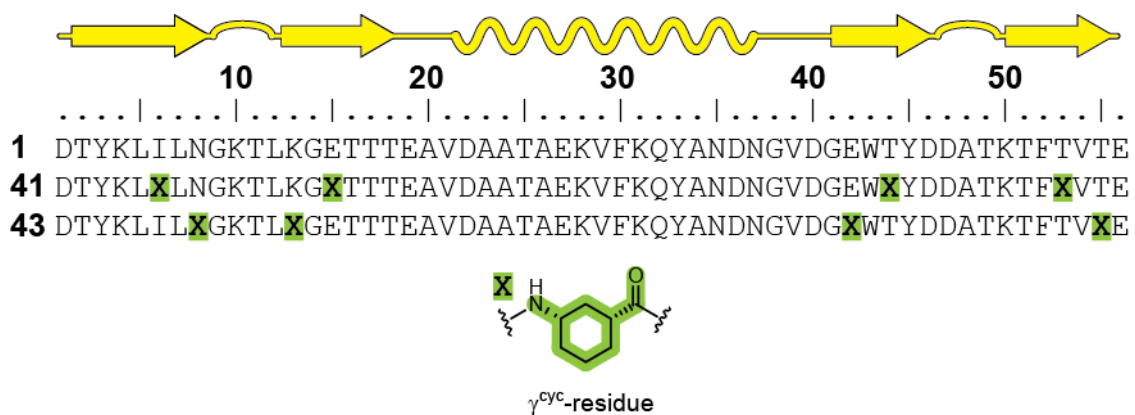
Sequence	sheet modification	T <sub>m</sub> (°C)	$\Delta\Delta G_{\text{fold vs. 1}}$ (kcal/mol)	$\Delta\Delta G_{\text{fold vs. 1}}$ (kcal/mol per substitution)
<b>1</b>	-	81.4 ± 0.1	-	-
<b>10</b>	2 $\alpha \rightarrow N\text{-Me}$	75.6 ± 0.1	+1.0	0.5 ± 0.1
<b>40</b>	4 $\alpha \rightarrow \gamma^{\text{A}}$	43.5 ± 0.4	+6.3	1.6 ± 0.4
<b>41</b>	4 $\alpha \rightarrow \gamma^{\text{cyc}}$	46.7 ± 0.3	+5.9	1.5 ± 0.3
<b>42</b>	Point mutations	78.0 ± 0.2	+0.7	0.2 ± 0.2

We performed simple modelling analysis of the GB1 structure focusing on the sites of backbone modification in **40** and **41**. One explanation for the large (~5 kcal/mol) destabilization of the folded state was that the protein backbone is lengthened considerably from  $\alpha \rightarrow \gamma$  replacement. The central location of the  $\gamma$ -residues in the sheet may significantly displace key core residues relative to the wild-type structure. The longer sheet may be less posed to make crucial tertiary contacts with the central  $\alpha$ -helix necessary to maintain the compact fold of GB1.

To test this hypothesis, a mutant of GB1 where four  $\alpha$ -residues, one at the end of the each strand, were replaced with  $\gamma^{\text{cyc}}$ -residues (**43**) was synthesized (Figures 71 and 72).

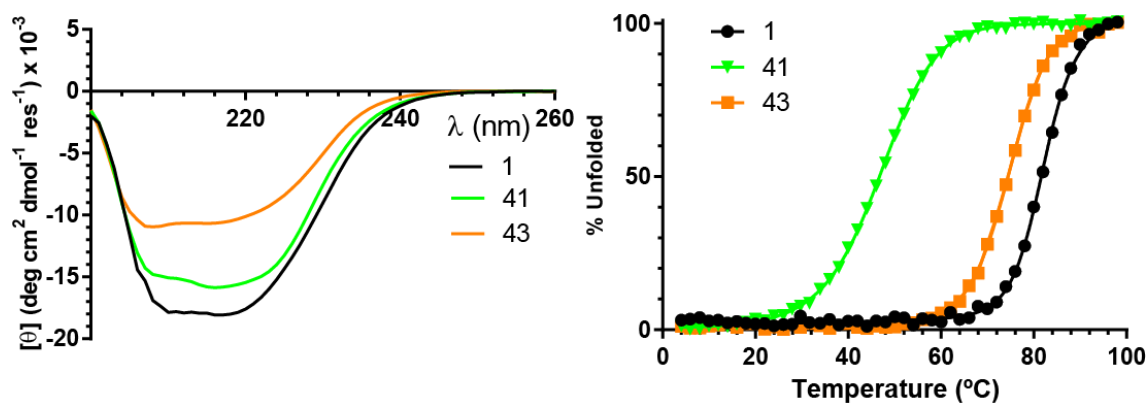


**Figure 71.** Model based on wild-type structure (2QMT) with location of unnatural residues in GB1 sheet mutants **41** and **43** colored green.



**Figure 72.** Sequences of wild-type **1** and sheet mutants **41** and **43**.

Supporting the rationale of our design, **43** had a  $T_m$  closer to that of **1** by CD-monitored thermal denaturation (Figure 73). The  $\Delta G_{\text{fold}}$  of **43** was 4.9 kcal/mol more favorable than that of **40** and **41**. This large energetic difference is reasonable given the large number of tertiary contacts restored between the sheet and helix of GB1. On a per-residue basis,  $\alpha \rightarrow \gamma^{\text{cyc}}$ -substitution is  $\sim 1.5x$  more favorable towards folding if placed near the termini of the strands (Table 20). The slightly ( $\sim 1.0$  kcal/mol) destabilized fold of **43** relative to wild-type is likely due to the replacement of several polar and charged side-chains that may make polar contacts in the folded protein.



**Figure 73.** CD scans and thermal melts of wild-type **1** and sheet mutants **41** and **43** in 20 mM phosphate buffered water pH 7.

**Table 20.** Folding Thermodynamics of Proteins **1**, **10**, and **40-43** from CD Measurements.

Sequence	Sheet modification	$T_m$ ( $^{\circ}\text{C}$ )	$\Delta\Delta G_{\text{fold}}$ vs. <b>1</b> (kcal/mol)	$\Delta\Delta G_{\text{fold}}$ vs. <b>1</b> (kcal/mol per substitution)
<b>1</b>	-	$81.4 \pm 0.1$	-	-
<b>41</b>	4 $\alpha \rightarrow \gamma^{\text{cyc}}$	$46.7 \pm 0.3$	+5.9	$1.5 \pm 0.3$
<b>43</b>	4 $\alpha \rightarrow \gamma^{\text{cyc}}$	$74.3 \pm 0.2$	+1.3	$0.3 \pm 0.1$

### 4.3 CONCLUSIONS AND FUTURE DIRECTIONS

In summary, we used a small hairpin peptide to understand and predict the thermodynamic impact of backbone modifications made in a larger protein system. It was observed in the sheet of GB1 that replacement of  $\alpha$ -residues with *N*-methylated analogues was slightly destabilizing to the fold. We synthesized several mutants of a shorter hairpin peptide bearing *N*-methyl residues at core hydrophobic packing positions. From analyses of 2D NMR spectra, it was shown that *N*-methylation is less favorable towards folding at positions closer to the  $\beta$ -turn. The observed destabilization was attributed to *cis/trans* amide isomerization and the diminished ability of *N*-methyl residues to adopt conformations necessary for hairpin and turn formations.

Earlier work in our lab showed that incorporation of  $\gamma^{\text{cyc}}$ -residues into the same hairpin peptide used for *N*-methylation studies was highly stabilizing to the sequence-encoded fold. The resulting mutant had a folded stability greater than that of the all  $\alpha$ -backbone peptide. As the replacement of an  $\alpha$ -residue with a  $\gamma^{\text{cyc}}$ -residue removes the side-chain, we synthesized vinylogous  $\gamma^4$ -residues that retain the native side-chain which may be important in maintaining important contacts in the wild-type structure. In contrast to the cyclic analogues,  $\gamma^4$ -residues destabilized the hairpin fold in the model peptide. This likely resulted from the additional rotatable bond present in  $\gamma^4$ -residues relative to  $\gamma^{\text{cyc}}$ -residues.

When the above strategies were applied to the larger sheet system of GB1, the results somewhat differed from what had been observed in the smaller model peptide. Both  $\gamma^4$ -residues and  $\gamma^{\text{cyc}}$ -residues dramatically reduced the folded stability of GB1. Structural analysis of the wild-type protein suggested that insertion of  $\gamma$ -residues into the center of the strands of GB1 disrupted tertiary contacts between the sheet and helix. When the positions of  $\gamma^{\text{cyc}}$ -substitution were shifted towards the termini of the strands, the resulting mutant had a folded stability within

1 kcal/mol of the wild-type protein. We attributed the minor energetic penalty towards folding as loss of polar and charged side-chains that may make contacts with other residues in the tertiary fold.

The above results underscore the limitations of model systems. Although incorporation of  $\gamma^{\text{cyc}}$ -residues into the center of a short hairpin peptide was favorable, this same strategy, when applied to a larger protein, led to highly destabilized mutants. The sheet of GB1 has shown to a particularly challenging target for backbone modification, and because of this, we hypothesize that our sheet mimicry strategies would be applicable to more complex biological targets. Future challenges include the backbone modification of larger proteins or multimeric enzymes with interesting functions as well as exploration of unnatural residues that increase the stability of protein structure beyond that of natural  $\alpha$ -amino acids.

## 4.4 EXPERIMENTAL

### 4.4.1 General Information

Solvents and all other reagents were purchased from Aldrich, Baker, EMD, or Fisher and used without further purification. HOBt was purchased from Anaspec Inc. HCTU, NovaPEG Rink Amide Resin, and Fmoc-protected  $\alpha$ -amino acids were purchased from Novabiochem. Fmoc-protected  $\beta^3$ -amino acids were purchased from Aapptec. Microcleavages were taken after all pseudoproline dipeptide couplings. Fmoc-ACC-OH was synthesized from a published route.<sup>198</sup> All protected Fmoc- $\gamma^4$ -amino acids were synthesized using published protocols.<sup>198, 201</sup>

#### **4.4.2 Peptide Synthesis**

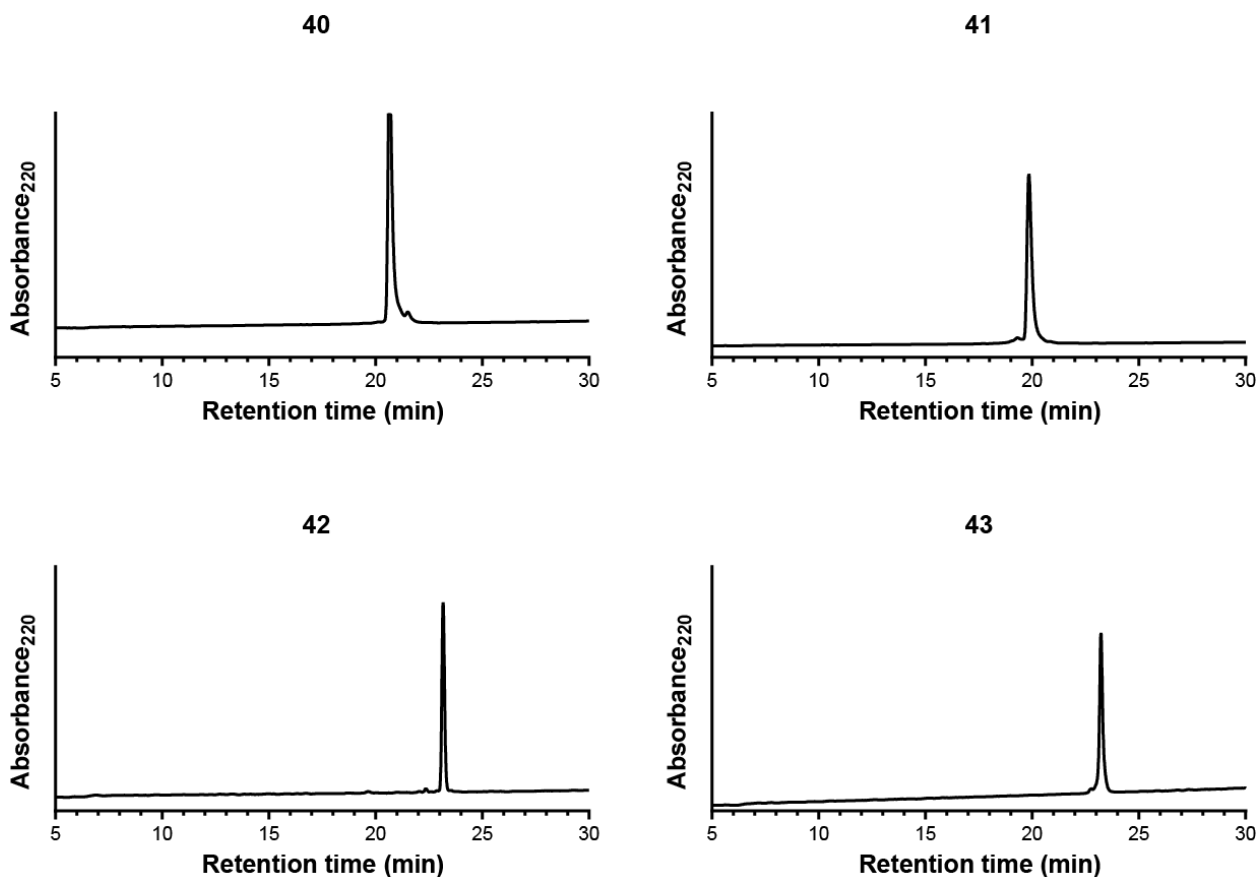
$\beta$ -Hairpin peptides **33-39** were synthesized using microwave-assisted Fmoc solid-phase synthesis techniques on a MARS microwave reactor (CEM) using NovaPEG Rink Amide resin. Couplings were carried out in NMP at 70 °C for 4 min using 4 equiv. of Fmoc-protected amino acid, 4 equiv. of HCTU, and 6 equiv. DIEA. PyAOP was used in place of HCTU for the coupling of N-methylated residues and residues immediately following them. Deprotections were performed using an excess of 20% 4-methylpiperidine in DMF at 80 °C for 2 min. After each coupling or deprotection cycle, the resin was washed three times with DMF. Double couplings were performed at sequence positions following proline or N-methylated residues. Prior to cleavage, the resin was washed three times each with DMF, dichloromethane, and methanol, and then dried. Peptide cleavage was accomplished using 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIS), and 2.5% water. Purification protocols were identical to those use for GB1 proteins.

#### **4.4.3 Protein Synthesis**

Proteins **40-43** were synthesized as described in Sections 2.5.2.

#### **4.4.4 Protein Purification and Characterization**

Proteins and peptides were purified as described in Section 2.5.3.



**Figure 74.** Analytical HPLC chromatograms of purified proteins **40-43**.

**Table 21.** MALDI-TOF MS Data for Peptides **34-39** and GB1 Mutants **40-43**.

#	[M+H] <sup>+</sup> <i>m/z</i> (average)	
	Calculated	Observed
<b>34</b>	1753.9	1753.8
<b>35</b>	1753.9	1753.8
<b>36</b>	1753.9	1753.7
<b>37</b>	1753.9	1753.8
<b>38</b>	1849.1	1849.1
<b>39</b>	1791.9	1791.4
<b>39<sub>cyc</sub></b>	2037.9	2037.5
<b>40</b>	6280.9	6281.8
<b>41</b>	6232.1	6231.3
<b>42</b>	6114.1	6114.7
<b>43</b>	6204.1	6203.7



#### 4.4.5 NMR Sample Preparation and Analysis

NMR samples were prepared by dissolving peptide in 750–850  $\mu\text{L}$  of degassed 50 mM phosphate, 9:1  $\text{H}_2\text{O}:\text{D}_2\text{O}$ , pH 6.3 (uncorrected for the presence of  $\text{D}_2\text{O}$ ) to a final concentration of 0.8–3 mM. 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS, 50 mM in water) was added to a final concentration of 0.2 mM. Each solution was passed through a 0.2  $\mu\text{m}$  syringe filter, and transferred to an NMR tube. The NMR tube headspace was purged with a stream of nitrogen prior to capping.

NMR experiments were performed on a Bruker Avance-700 spectrometer. Chemical shifts are reported relative to DSS (0 ppm). TOCSY, NOESY, and COSY pulse programs used excitation-sculpted gradient-pulse solvent suppression. For all 2D experiments, 2048 data points were collected in the direct dimension and 512 data points in the indirect dimension. The mixing times for TOCSY and NOESY were 80 ms and 200 ms, respectively. NMR measurements were performed at a temperature of 278 K for hairpin peptides **33-39** and at 293 K for cyclized hairpin peptide **39<sub>cyc</sub>**. The Sparky software package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) was used to analyze 2D NMR data. Backbone chemical shift assignments for peptides **33-39** were previously reported.<sup>120, 198, 201</sup>. Analysis of NMR data for **33-39** and estimation of folded populations followed previously published methods.<sup>120</sup> Tabulated NOEs for peptide **39<sub>cyc</sub>** were previously reported.<sup>198</sup> These data were applied to calculate an NMR solution structure of **39<sub>cyc</sub>** using the Crystallography and NMR system (CNS) software package according to published methods.<sup>120</sup>

#### **4.4.6 Circular Dichroism**

CD spectra and thermal melts were performed and analyzed for **1** and sheet mutants **40-43** as described in Section 2.5.5.

## 5.0 OUTLOOK

The above work described the development of heterogeneous-backbone oligomers with protein-like tertiary folds. We hope that this work will enable the design of analogues of more biologically interesting targets, including protein-based therapeutics. Unnatural oligomers can offer enhanced proteolytic resistance over natural peptides.<sup>187</sup> When designed to fold like natural proteins, such molecules could overcome a major drawback to peptide-based bioactive agents.<sup>202</sup> Before such capabilities can be fully realized, there are significant challenges that must be addressed.

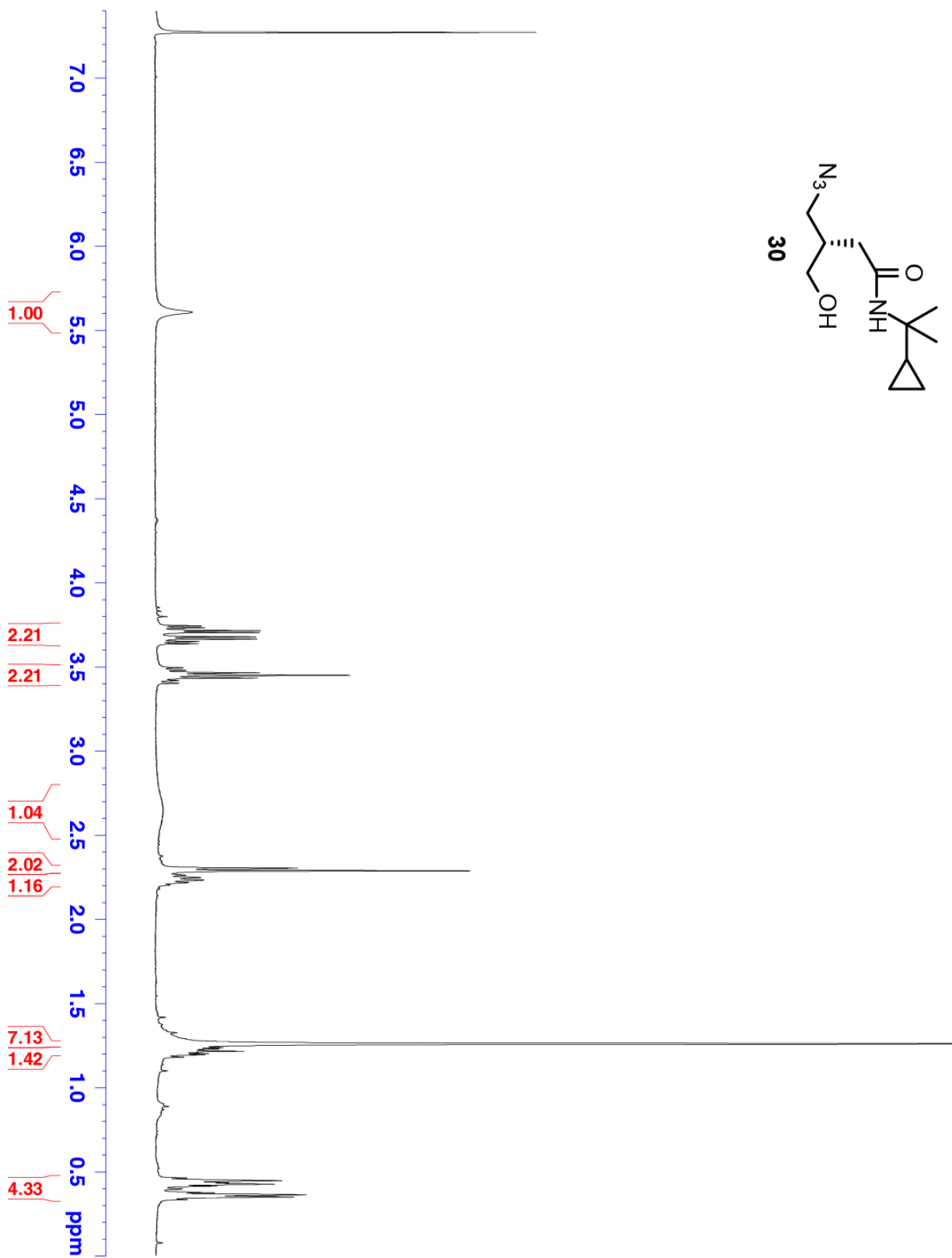
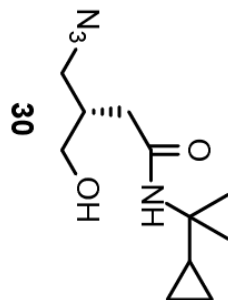
One limitation of protein-like foldamers is the scalability of synthesis. Many GB1 syntheses described earlier had overall yields of less than 1% after purification. Although GB1 contains 56 residues, a large portion of the sequence has high sheet-folding propensity, especially as a protected peptide on resin. This has the effect of promoting chain aggregation on resin which reduces yield over the 112+ steps required to synthesize the wild-type protein. This problem would be amplified in larger proteins of interest. One solution would be to biologically synthesize the desired mutants using unnatural amino acid expression techniques described in Chapter 1. Bacterial synthesis of proteins with significantly unnatural backbones would completely revolutionize the fields of peptide-based therapeutics and protein design. However, current technologies do not allow for several of the residue types explored in this work to be efficiently biologically incorporated into proteins. Furthermore, fine control over insertion of

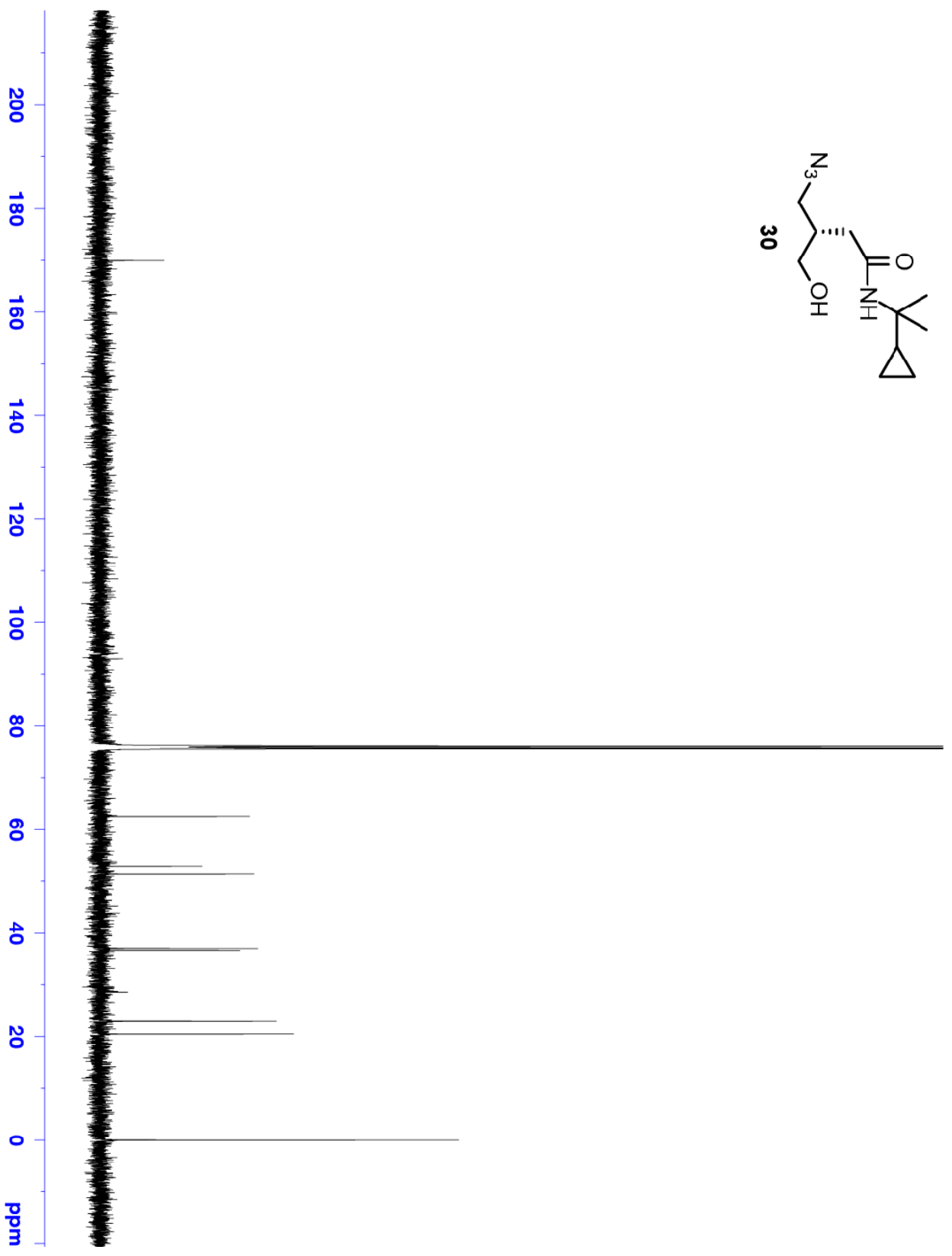
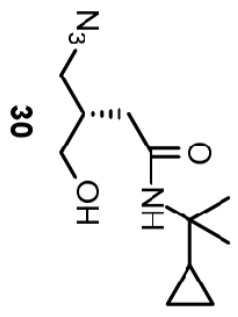
multiple unnatural residues of different types is not available. Methods expanding the chemical backbones amenable to ribosomal protein synthesis continue to be developed.<sup>203-204</sup> Perhaps in the future, recombinant synthesis of highly heterogeneous-backbone oligomers will be possible.

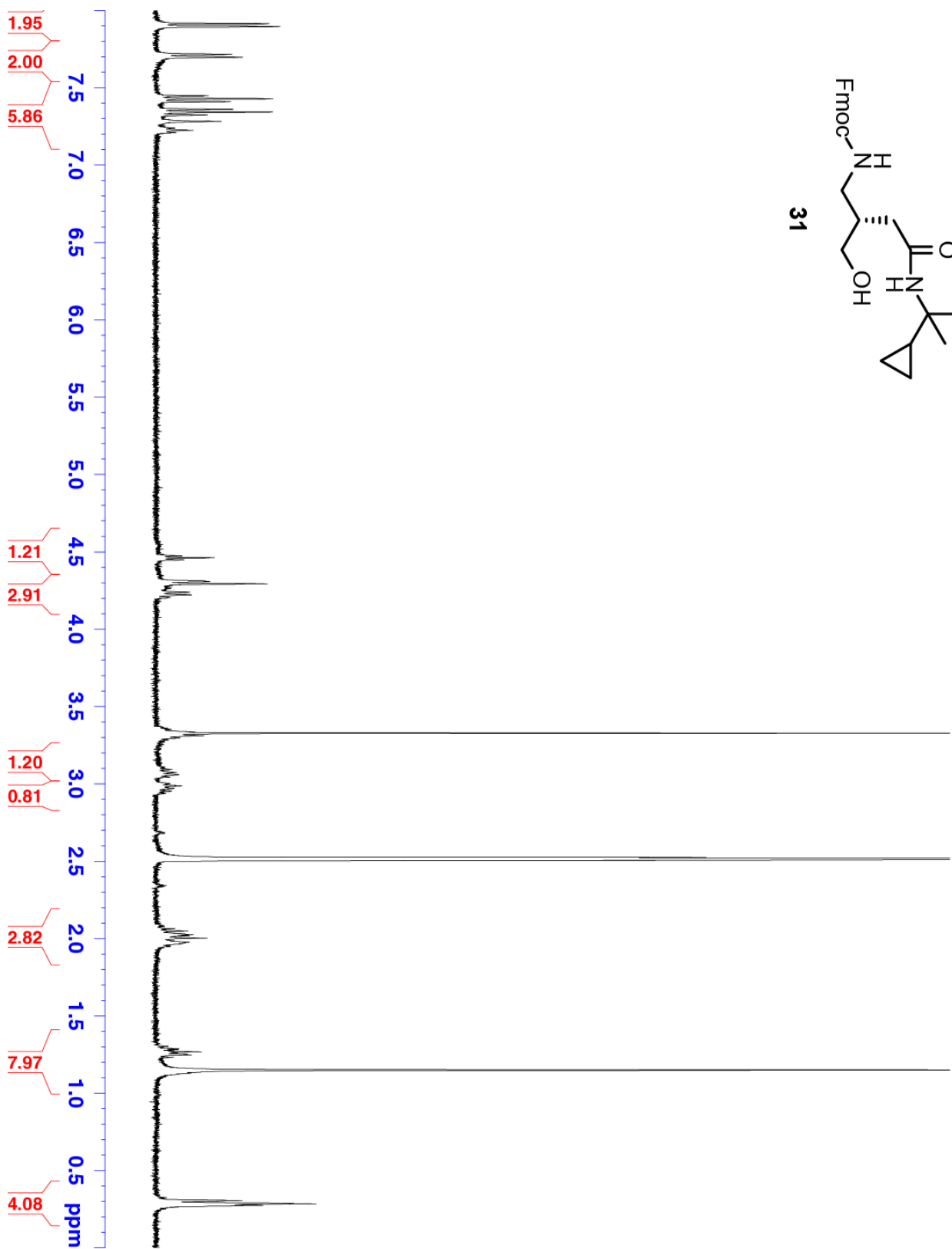
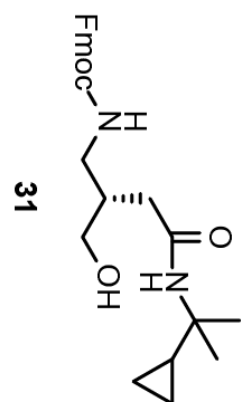
Alternatively, convergent syntheses of proteins allow for the production of unnatural backbone oligomers with higher yields than SPPS alone. In this method, NCL, EPL, and SPPS can be combined to generate designer proteins with a number of modifications in a scalable manner.<sup>205-206</sup> An improved version of this method using C-terminal acyl hydrazides has recently been applied to the synthesis of a 142 residue model protein through selective couplings between six peptide fragments.<sup>207</sup> Protein prosthesis strategies highlighted in Chapter 1 could also be applied to targets of therapeutic interest, especially in instances where only a small portion of the sequence would require enhanced proteolytic resistance.

Another challenge for foldamer research is the continued development of strategies for tertiary and quaternary protein structure mimicry. Because nearly all protein folds share the secondary structure elements present in GB1, our methods for mimicking tertiary folds on heterogeneous-backbone oligomers should be applicable to many other targets. The folded state of GB1 was tolerant of numerous backbone modifications in one sequence without compromising the sequence-encoded fold. This high folded stability is not representative of many proteins that are only marginally stable under physiological conditions. The design of backbone modifications that maintain or favorably change the folding free energy relative to  $\alpha$ -residues will expand the number of accessible biological targets and enhance our understanding of protein folding thermodynamics. Future work will examine the robustness of the tertiary structure mimicry methods developed here using other protein targets with interesting biological activities.

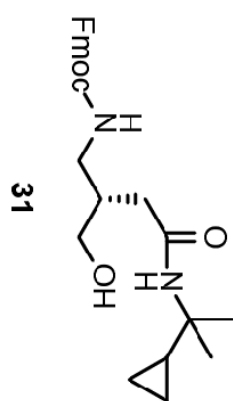
**$^1\text{H}$  AND  $^{13}\text{C}$  NMR FOR SYNTHESIZED SMALL MOLECULES 30-32**

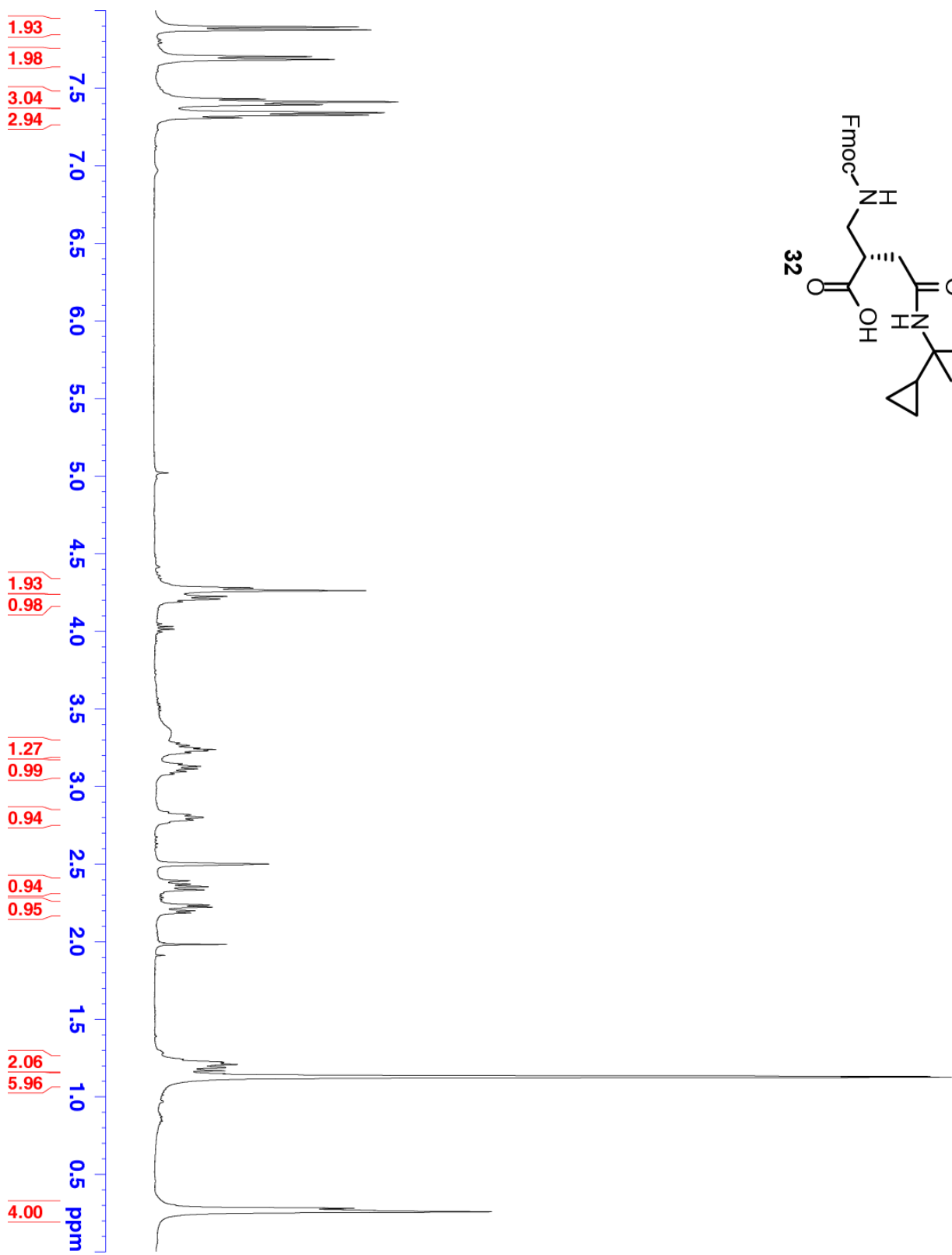
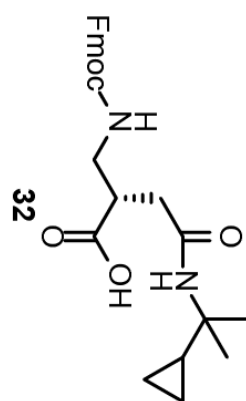


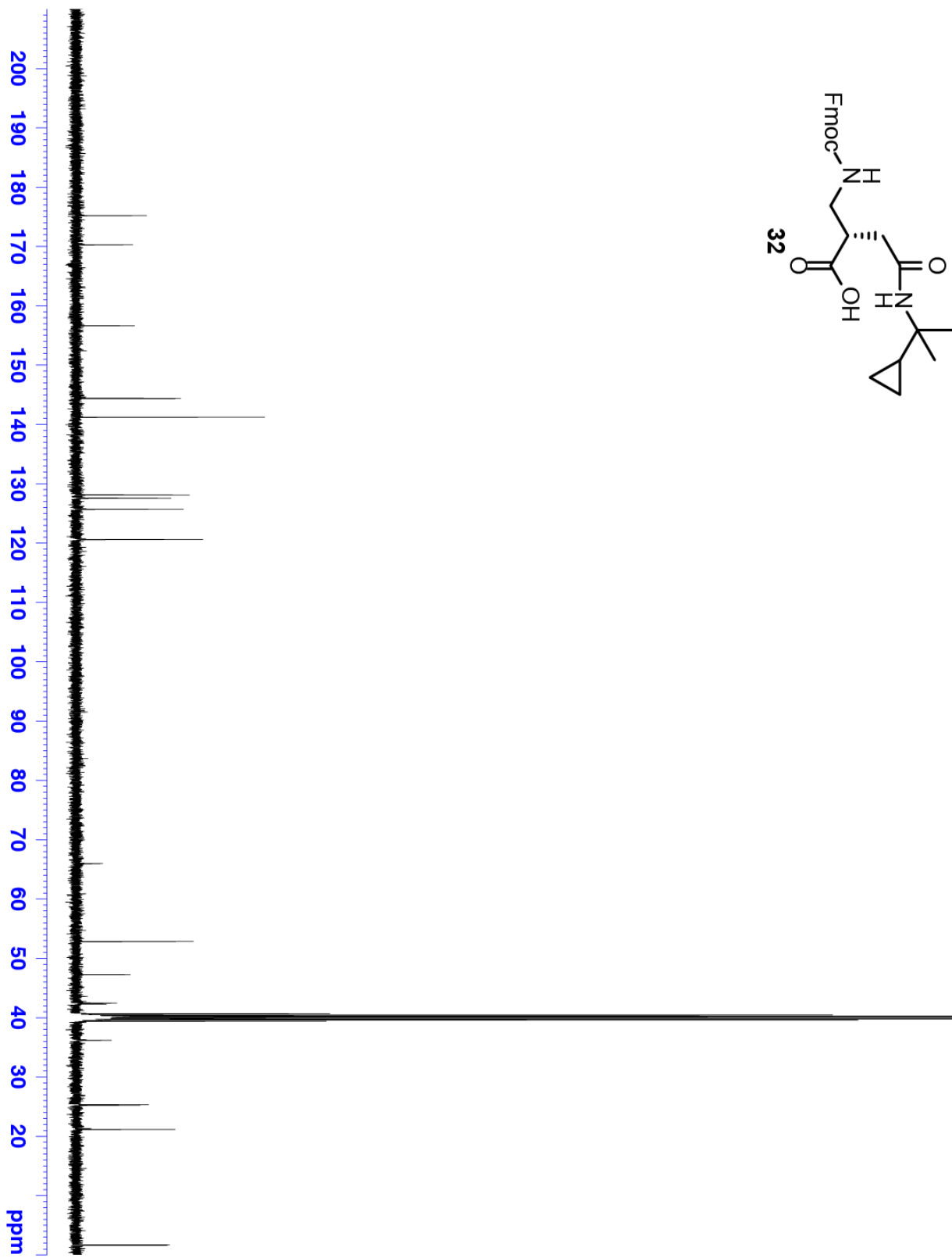
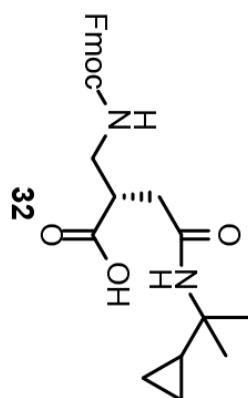












## REFERENCES

1. Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S., *Science*. **1994**, *266*, 776-779.
2. Muir, T. W.; Sondhi, D.; Cole, P. A., *Proc. Natl. Acad. Sci. USA*. **1998**, *95*, 6705-6710.
3. Ellman, J. A.; Mendel, D.; Schultz, P. G., *Science*. **1992**, *255*, 197-200.
4. Hackenberger, C. P. R.; Schwarzer, D., *Angew. Chem. Int. Ed.* **2008**, *47*, 10030-10074.
5. Raibaut, L.; Ollivier, N.; Melnyk, O., *Chem. Soc. Rev.* **2012**, *41*, 7001-7015.
6. Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N.; Schultz, P. G.; Dougherty, D. A.; Lester, H. A., *Science*. **1995**, *268*, 439-442.
7. Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G., *Proc. Natl. Acad. Sci. USA*. **1994**, *91*, 2910-2914.
8. England, P. M.; Zhang, Y.; Dougherty, D. A.; Lester, H. A., *Cell*. **1999**, *96*, 89-98.
9. Koh, J. T.; Cornish, V. W.; Schultz, P. G., *Biochemistry*. **1997**, *36*, 11314-11322.
10. Steward, L. E.; Collins, C. S.; Gilmore, M. A.; Carlson, J. E.; Ross, J. B. A.; Chamberlin, A. R., *J. Am. Chem. Soc.* **1997**, *119*, 6-11.
11. Karginov, A. V.; Lodder, M.; Hecht, S. M., *Nucleic Acids Res.* **1999**, *27*, 3283-3290.
12. Xie, J.; Schultz, P. G., *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775-782.

13. Choudhary, A.; Raines, R. T., *ChemBioChem*. **2011**, *12*, 1801-1807.
14. Lu, W.; Qasim, M. A.; Laskowski, M.; Kent, S. B. H., *Biochemistry*. **1997**, *36*, 673-679.
15. Chapman, E.; Thorson, J. S.; Schultz, P. G., *J. Am. Chem. Soc.* **1997**, *119*, 7151-7152.
16. Shin, I.; Ting, A. Y.; Schultz, P. G., *J. Am. Chem. Soc.* **1997**, *119*, 12667-12668.
17. Beligere, G. S.; Dawson, P. E., *J. Am. Chem. Soc.* **2000**, *122*, 12079-12082.
18. Wales, T. E.; Fitzgerald, M. C., *J. Am. Chem. Soc.* **2001**, *123*, 7709-7710.
19. Blankenship, J. W.; Balambika, R.; Dawson, P. E., *Biochemistry*. **2002**, *41*, 15676-15684.
20. McFarland, B. J.; Katz, J. F.; Sant, A. J.; Beeson, C., *J. Mol. Biol.* **2005**, *350*, 170-183.
21. Goldberg, J. M.; Wissner, R. F.; Klein, A. M.; Petersson, E. J., *Chem. Commun.* **2012**, *48*, 1550-1552.
22. Goldberg, J. M.; Batjargal, S.; Petersson, E. J., *J. Am. Chem. Soc.* **2010**, *132*, 14718-14720.
23. Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J., *J. Am. Chem. Soc.* **2012**, *134*, 9172-9182.
24. Haney, C. M.; Wissner, R. F.; Petersson, E. J., *Curr. Opin. Chem. Biol.* **2015**, *28*, 123-130.
25. Newberry, R. W.; VanVeller, B.; Guzei, I. A.; Raines, R. T., *J. Am. Chem. Soc.* **2013**, *135*, 7843-7846.
26. Milton, R. C.; Milton, S. C.; Kent, S. B. H., *Science*. **1992**, *256*, 1445-1448.
27. Fitzgerald, M. C.; Chernushevich, I.; Standing, K. G.; Kent, S. B. H.; Whitman, C. P., *J. Am. Chem. Soc.* **1995**, *117*, 11075-11080.
28. Zawadzke, L. E.; Berg, J. M., *J. Am. Chem. Soc.* **1992**, *114*, 4002-4003.

29. Matthews, B. W., *Protein Sci.* **2009**, *18*, 1135-1138.
30. Yeates, T. O.; Kent, S. B. H., *Annu. Rev. Biophys.* **2012**, *41*, 41-61.
31. Wukovitz, S. W.; Yeates, T. O., *Nat. Struct. Mol. Biol.* **1995**, *2*, 1062-1067.
32. Okamoto, R.; Mandal, K.; Sawaya, M. R.; Kajihara, Y.; Yeates, T. O.; Kent, S. B. H., *Angew. Chem. Int. Ed.* **2014**, *53*, 5194-5198.
33. Banigan, J. R.; Mandal, K.; Sawaya, M. R.; Thammavongsa, V.; Hendrickx, A. P. A.; Schneewind, O.; Yeates, T. O.; Kent, S. B. H., *Protein Sci.* **2010**, *19*, 1840-1849.
34. Mandal, K.; Pentelute, B. L.; Tereshko, V.; Thammavongsa, V.; Schneewind, O.; Kossiakoff, A. A.; Kent, S. B. H., *Protein Sci.* **2009**, *18*, 1146-1154.
35. Marshall, G. R.; Hodgkin, E. E.; Langs, D. A.; Smith, G. D.; Zabrocki, J.; Leplawy, M. T., *Proc. Natl. Acad. Sci. USA.* **1990**, *87*, 487-491.
36. De Filippis, V.; De Antoni, F.; Frigo, M.; Polverino de Laureto, P.; Fontana, A., *Biochemistry.* **1998**, *37*, 1686-1696.
37. Torbeev, V. Y.; Raghuraman, H.; Hamelberg, D.; Tonelli, M.; Westler, W. M.; Perozo, E.; Kent, S. B. H., *Proc. Natl. Acad. Sci. USA.* **2011**, *108*, 20982-20987.
38. Hamada, Y.; Shioiri, T., *Chem. Rev.* **2005**, *105*, 4441-4482.
39. Clore, G. M.; Appella, E.; Yamada, M.; Matsushima, K.; Gronenborn, A. M., *Biochemistry.* **1990**, *29*, 1689-1696.
40. Rajarathnam, K.; Sykes, B.; Kay, C.; Dewald, B.; Geiser, T.; Baggiolini, M.; Clark-Lewis, I., *Science.* **1994**, *264*, 90-92.
41. Armand, P.; Kirshenbaum, K.; Goldsmith, R. A.; Farr-Jones, S.; Barron, A. E.; Truong, K. T. V.; Dill, K. A.; Mierke, D. F.; Cohen, F. E.; Zuckermann, R. N.; Bradley, E. K., *Proc. Natl. Acad. Sci.* **1998**, *95*, 4309-4314.
42. Horne, W. S., *Expert Opin. Drug Discov.* **2011**, *6*, 1247-1262.

43. Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuckermann, R. N., *Proc. Natl. Acad. Sci.* **1998**, *95*, 4303-4308.
44. Wender, P. A.; Mitchell, D. J.; Pattabiraman, K.; Pelkey, E. T.; Steinman, L.; Rothbard, J. B., *Proc. Natl. Acad. Sci.* **2000**, *97*, 13003-13008.
45. Chongsiriwatana, N. P.; Patch, J. A.; Czyzewski, A. M.; Dohm, M. T.; Ivankin, A.; Gidalevitz, D.; Zuckermann, R. N.; Barron, A. E., *Proc. Natl. Acad. Sci.* **2008**, *105*, 2794-2799.
46. Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A., *Science*. **1998**, *282*, 2088-2092.
47. Trader, D. J.; Simanski, S.; Kodadek, T., *J. Am. Chem. Soc.* **2015**, *137*, 6312-6319.
48. Seebach, D.; Abele, S.; Gademann, K.; Guichard, G.; Hintermann, T.; Jaun, B.; Matthews, J. L.; Schreiber, J. V.; Oberer, L.; Hommel, U.; Widmer, H., *Helv. Chim. Acta.* **1998**, *81*, 932-982.
49. Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H., *J. Am. Chem. Soc.* **1996**, *118*, 13071-13072.
50. David, R.; Günther, R.; Baumann, L.; Lühmann, T.; Seebach, D.; Hofmann, H.-J.; Beck-Sickinger, A. G., *J. Am. Chem. Soc.* **2008**, *130*, 15311-15317.
51. Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T., *J. Am. Chem. Soc.* **2002**, *124*, 8522-8523.
52. Purcell, W. P.; Singer, J. A., *J. Phys. Chem.* **1967**, *71*, 4316-4319.
53. Horne, W. S.; Yadav, M. K.; Stout, C. D.; Ghadiri, M. R., *J. Am. Chem. Soc.* **2004**, *126*, 15366-15367.
54. Tam, A.; Arnold, U.; Soellner, M. B.; Raines, R. T., *J. Am. Chem. Soc.* **2007**, *129*, 12670-12671.
55. Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T., *Org. Lett.* **2005**, *7*, 2619-2622.
56. Dai, N.; Wang, X. J.; Etzkorn, F. A., *J. Am. Chem. Soc.* **2008**, *130*, 5396-5397.

57. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M., *Drug Discov. Today*. **2010**, *15*, 40-56.
58. Gellman, S. H., *Acc. Chem. Res.* **1998**, *31*, 173-180.
59. Jiang, H.; Léger, J.-M.; Dolain, C.; Guionneau, P.; Huc, I., *Tetrahedron*. **2003**, *59*, 8365-8374.
60. Horne, W. S.; Gellman, S. H., *Acc. Chem. Res.* **2008**, *41*, 1399-1408.
61. Nagai, U.; Sato, K., *Tetrahedron Lett.* **1985**, *26*, 647-650.
62. Haque, T. S.; Little, J. C.; Gellman, S. H., *J. Am. Chem. Soc.* **1994**, *116*, 4105-4106.
63. Viles, J. H.; Patel, S. U.; Mitchell, J. B. O.; Moody, C. M.; Justice, D. E.; Uppenbrink, J.; Doyle, P. M.; Harris, C. J.; Sadler, P. J.; Thornton, J. M., *J. Mol. Biol.* **1998**, *279*, 973-986.
64. Wipf, P.; Henninger, T. C.; Geib, S. J., *J. Org. Chem.* **1998**, *63*, 6088-6089.
65. Odaert, B.; Jean, F.; Melnyk, O.; Tartar, A.; Lippens, G.; Boutillon, C.; Buisine, E., *Protein Sci.* **1999**, *8*, 2773-2783.
66. Liang, G.-B.; Gellman, S. H., *J. Am. Chem. Soc.* **1999**, *121*, 1806-1816.
67. Chung, Y. J.; Huck, B. R.; Christianson, L. A.; Stanger, H. E.; Krauthäuser, S.; Powell, D. R.; Gellman, S. H., *J. Am. Chem. Soc.* **2000**, *122*, 3995-4004.
68. Masterson, L. R.; Etienne, M. A.; Porcelli, F.; Barany, G.; Hammer, R. P.; Veglia, G., *Pept. Sci.* **2007**, *88*, 746-753.
69. Fuller, A. A.; Du, D.; Liu, F.; Davoren, J. E.; Bhabha, G.; Kroon, G.; Case, D. A.; Dyson, H. J.; Powers, E. T.; Wipf, P.; Gruebele, M.; Kelly, J. W., *Proc. Natl. Acad. Sci.* **2009**, *106*, 11067-11072.
70. Horne, W. S.; Price, J. L.; Keck, J. L.; Gellman, S. H., *J. Am. Chem. Soc.* **2007**, *129*, 4178-4180.



71. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H., *Proc. Natl. Acad. Sci.* **2009**, *106*, 14751-14756.
72. Daniels, D. S.; Petersson, E. J.; Qiu, J. X.; Schepartz, A., *J. Am. Chem. Soc.* **2007**, *129*, 1532-1533.
73. Petersson, E. J.; Schepartz, A., *J. Am. Chem. Soc.* **2008**, *130*, 821-823.
74. Wang, P. S. P.; Nguyen, J. B.; Schepartz, A., *J. Am. Chem. Soc.* **2014**, *136*, 6810-6813.
75. Price, J. L.; Hadley, E. B.; Steinkruger, J. D.; Gellman, S. H., *Angew. Chem. Int. Ed.* **2010**, *49*, 368-371.
76. Lee, B.-C.; Chu, T. K.; Dill, K. A.; Zuckermann, R. N., *J. Am. Chem. Soc.* **2008**, *130*, 8847-8855.
77. Collie, G. W.; Pulka-Ziach, K.; Lombardo, C. M.; Fremaux, J.; Rosu, F.; Decossas, M.; Mauran, L.; Lambert, O.; Gabelica, V.; Mackereth, C. D.; Guichard, G., *Nat. Chem.* **2015**, *7*, 871-878.
78. Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M., *J. Phys. Chem. B.* **2007**, *111*, 14362-14369.
79. Gronenborn, A.; Filpula, D.; Essig, N.; Achari, A.; Whitlow, M.; Wingfield, P.; Clore, G., *Science.* **1991**, *253*, 657-661.
80. Byeon, I.-J. L.; Louis, J. M.; Gronenborn, A. M., *J. Mol. Biol.* **2003**, *333*, 141-152.
81. Li, H.; Wang, H. C.; Cao, Y.; Sharma, D.; Wang, M., *J. Mol. Biol.* **2008**, *379*, 871-80.
82. Minor, D. L.; Kim, P. S., *Nature.* **1994**, *367*, 660-663.
83. Minor, D. L.; Kim, P. S., *Nature.* **1994**, *371*, 264-267.
84. Smith, C. K.; Withka, J. M.; Regan, L., *Biochemistry.* **1994**, *33*, 5510-5517.

85. Kievit, F. M.; Veiseh, O.; Fang, C.; Bhattarai, N.; Lee, D.; Ellenbogen, R. G.; Zhang, M., *ACS Nano*. **2010**, *4*, 4587-4594.
86. Price, J. L.; Powers, E. T.; Kelly, J. W., *ACS Chem. Biol.* **2011**.
87. Harris, J. M.; Chess, R. B., *Nat. Rev. Drug Discov.* **2003**, *2*, 214-221.
88. Alconcel, S. N. S.; Baas, A. S.; Maynard, H. D., *Polym. Chem.* **2011**, *2*, 1442-1448.
89. Trauger, J. W.; Kohli, R. M.; Walsh, C. T., *Biochemistry*. **2001**, *40*, 7092-7098.
90. Green, B. R.; Catlin, P.; Zhang, M.-M.; Fiedler, B.; Bayudan, W.; Morrison, A.; Norton, Raymond S.; Smith, B. J.; Yoshikami, D.; Olivera, B. M.; Bulaj, G., *Chem. Biol.* **2007**, *14*, 399-407.
91. Reinert, Z. E.; Musselman, E. D.; Elcock, A. H.; Horne, W. S., *ChemBioChem*. **2012**, *13*, 1107-1111.
92. Kelly, S. M.; Jess, T. J.; Price, N. C., *BBA Protein Proteom.* **2005**, *1751*, 119-139.
93. Cregut, D.; Serrano, L., *Protein Sci.* **1999**, *8*, 271-282.
94. Betz, S. F.; Raleigh, D. P.; DeGrado, W. F., *Curr. Opin. Struct. Biol.* **1993**, *3*, 601-610.
95. Alexander, P.; Fahnestock, S.; Lee, T.; Orban, J.; Bryan, P., *Biochemistry*. **1992**, *31*, 3597-3603.
96. Shortle, D.; Meeker, A. K.; Freire, E., *Biochemistry*. **1988**, *27*, 4761-4768.
97. Becktel, W. J.; Schellman, J. A., *Biopolymers*. **1987**, *26*, 1859-1877.
98. Bjorck, L.; Kronvall, G., *J. Immunol.* **1984**, *133*, 969-974.
99. Sauer-Eriksson, A. E.; Kleywegt, G. J.; Uhlén, M.; Jones, T. A., *Structure*. **1995**, *3*, 265-278.

100. Frick, I. M.; Wikström, M.; Forsén, S.; Drakenberg, T.; Gomi, H.; Sjöbring, U.; Björck, L., *Proc. Natl. Acad. Sci.* **1992**, *89*, 8532-8536.
101. Tripathi, S.; Portman, J. J., *J. Chem. Phys.* **2008**, *128*, 205104-9.
102. Tripathi, S.; Portman, J. J., *Proc. Natl. Acad. Sci.* **2009**, *106*, 2104-2109.
103. Stites, W. E., *Chem. Rev.* **1997**, *97*, 1233-1250.
104. Schmitt, M. A.; Choi, S. H.; Guzei, I. A.; Gellman, S. H., *J. Am. Chem. Soc.* **2005**, *127*, 13130-13131.
105. Mándity, I. M.; Wéber, E.; Martinek, T. A.; Olajos, G.; Tóth, G. K.; Vass, E.; Fülöp, F., *Angew. Chem. Int. Ed.* **2009**, *48*, 2171-2175.
106. Boersma, M. D.; Haase, H. S.; Peterson-Kaufman, K. J.; Lee, E. F.; Clarke, O. B.; Colman, P. M.; Smith, B. J.; Horne, W. S.; Fairlie, W. D.; Gellman, S. H., *J. Am. Chem. Soc.* **2011**, *134*, 315-323.
107. Sawada, T.; Gellman, S. H., *J. Am. Chem. Soc.* **2011**, *133*, 7336-7339.
108. Wu, C. W.; Sanborn, T. J.; Huang, K.; Zuckermann, R. N.; Barron, A. E., *J. Am. Chem. Soc.* **2001**, *123*, 6778-6784.
109. Horne, W. S.; Boersma, M. D.; Windsor, M. A.; Gellman, S. H., *Angew. Chem.* **2008**, *120*, 2895-2898.
110. Horne, W. S.; Price, J. L.; Gellman, S. H., *Proc. Natl. Acad. Sci.* **2008**, *105*, 9151-9156.
111. Bansal, M.; Kumar, S.; Velavan, R., *J. Biomol. Struct. Dyn.* **2000**, *17*, 811-9.
112. Smith, B. J., *J. Mol. Graphics Modell.* **2012**, *33*, 52-60.
113. Martinek, T. A.; Mándity, I. M.; Fülöp, L.; Tóth, G. K.; Vass, E.; Hollósi, M.; Forró, E.; Fülöp, F., *J. Am. Chem. Soc.* **2006**, *128*, 13539-13544.

114. Martinek, T. A.; Tóth, G. K.; Vass, E.; Hollósi, M.; Fülöp, F., *Angew. Chem. Int. Ed.* **2002**, *41*, 1718-1721.
115. Rúa, F.; Boussert, S.; Parella, T.; Díez-Pérez, I.; Branchadell, V.; Giralt, E.; Ortuño, R. M., *Org. Lett.* **2007**, *9*, 3643-3645.
116. Nowick, J. S., *Acc. Chem. Res.* **2008**, *41*, 1319-1330.
117. Pham, J. D.; Chim, N.; Goulding, C. W.; Nowick, J. S., *J. Am. Chem. Soc.* **2013**, *135*, 12460-12467.
118. Pham, J. D.; Spencer, R. K.; Chen, K. H.; Nowick, J. S., *J. Am. Chem. Soc.* **2014**, *136*, 12682-12690.
119. Lengyel, G. A.; Frank, R. C.; Horne, W. S., *J. Am. Chem. Soc.* **2011**, *133*, 4246-4249.
120. Lengyel, G. A.; Horne, W. S., *J. Am. Chem. Soc.* **2012**, *134*, 15906-15913.
121. Hughes, E.; Burke, R. M.; Doig, A. J., *J. Biol. Chem.* **2000**, *275*, 25109-25115.
122. Kapurniotu, A.; Schmauder, A.; Tenidis, K., *J. Mol. Biol.* **2002**, *315*, 339-350.
123. Koo, B.-K.; Park, C.-J.; Fernandez, C. F.; Chim, N.; Ding, Y.; Chanfreau, G.; Feigon, J., *J. Mol. Biol.* **2011**, *411*, 927-942.
124. Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E., *J. Am. Chem. Soc.* **2013**, *135*, 3186-3192.
125. Hutchinson, E. G.; Thornton, J. M., *Protein Sci.* **1994**, *3*, 2207-2216.
126. Chung, Y. J.; Christianson, L. A.; Stanger, H. E.; Powell, D. R.; Gellman, S. H., *J. Am. Chem. Soc.* **1998**, *120*, 10555-10556.
127. Oh, K.; Guan, Z., *Chem. Commun.* **2006**, 3069-3071.
128. Haque, T. S.; Little, J. C.; Gellman, S. H., *J. Am. Chem. Soc.* **1996**, *118*, 6975-6985.

129. Stanger, H. E.; Gellman, S. H., *J. Am. Chem. Soc.* **1998**, *120*, 4236-4237.
130. Burke, D. F.; Deane, C. M.; Blundell, T. L., *Bioinformatics.* **2000**, *16*, 513-519.
131. Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H., *J. Am. Chem. Soc.* **2004**, *126*, 7238-7243.
132. Karle, I. L.; Balaram, P., *Biochemistry.* **1990**, *29*, 6747-6756.
133. Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D., *Protein Sci.* **2002**, *11*, 2924-2931.
134. Nauli, S.; Kuhlman, B.; Baker, D., *Nat. Struct. Mol. Biol.* **2001**, *8*, 602-605.
135. Wells, J. A.; McClendon, C. L., *Nature.* **2007**, *450*, 1001-1009.
136. Merrifield, R. B., *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
137. White, P.; Keyte, J. W.; Bailey, K.; Bloomberg, G., *J. Pept. Sci.* **2004**, *10*, 18-26.
138. Yang, Y.; Sweeney, W. V.; Schneider, K.; Thörnqvist, S.; Chait, B. T.; Tam, J. P., *Tetrahedron Lett.* **1994**, *35*, 9689-9692.
139. Palasek, S. A.; Cox, Z. J.; Collins, J. M., *J. Pept. Sci.* **2007**, *13*, 143-148.
140. Wöhr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M., *J. Am. Chem. Soc.* **1996**, *118*, 9218-9227.
141. Pukin, A. V.; Branderhorst, H. M.; Sisu, C.; Weijers, C. A. G. M.; Gilbert, M.; Liskamp, R. M. J.; Visser, G. M.; Zuilhof, H.; Pieters, R. J., *ChemBioChem.* **2007**, *8*, 1500-1503.
142. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Comput. Phys. Commun.* **1995**, *91*, 43-56.
143. Lindahl, E.; Hess, B.; van der Spoel, D., *J. Mol. Model.* **2001**, *7*, 306-317.

144. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., *J. Comput. Chem.* **2005**, *26*, 1701-1718.
145. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
146. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., *J. Phys. Chem. B.* **2001**, *105*, 6474-6487.
147. Fischer, J.; Paschek, D.; Geiger, A.; Sadowski, G., *J. Phys. Chem. B.* **2008**, *112*, 2388-2398.
148. Lee, H.; Venable, R. M.; MacKerell, A. D.; Pastor, R. W., *Biophys. J.* **2008**, *95*, 1590-1599.
149. Goutev, N.; Ohno, K.; Matsuura, H., *J. Phys. Chem. A.* **2000**, *104*, 9226-9232.
150. M. P. Allen, D. J. T., *Computer Simulations of Liquids*. Oxford Science Publications: Oxford, 1987.
151. Jorgensen, W. L., *J. Phys. Chem.* **1986**, *90*, 1276-1284.
152. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983**, *79*, 926-935.
153. Aqvist, J., *J. Phys. Chem.* **1990**, *94*, 8021-8024.
154. Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981**, *52*, 7182-7190.
155. Hoover, W. G., *Phys. Rev. A.* **1985**, *31*, 1695-1697.
156. Nose, S., *J. Chem. Phys.* **1984**, *81*, 511-519.
157. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., *J. Chem. Phys.* **1995**, *103*, 8577-8593.
158. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., *J. Comput. Chem.* **1997**, *18*, 1463-1472.

159. Collaborative, *Acta Crystallog. Sect. D.* **1994**, *50*, 760-763.
160. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., *Acta Crystallog. Sect. D.* **2010**, *66*, 213-221.
161. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
162. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta Crystallog. Sect. D.* **1997**, *53*, 240-255.
163. Emsley, P.; Cowtan, K., *Acta Crystallog. Sect. D.* **2004**, *60*, 2126-2132.
164. Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A., *Nat. Protoc.* **2008**, *3*, 1171-9.
165. Baker, D.; Agard, D. A., *Biochemistry.* **1994**, *33*, 7505-7509.
166. Makhatadze, G. I.; Privalov, P. L., Energetics of Protein Structure. In *Adv. Protein Chem.*, C.B. Anfinsen, F. M. R. J. T. E.; David, S. E., Eds. Academic Press: 1995; Vol. Volume 47, pp 307-425.
167. Kuhlman, B.; Raleigh, D. P., *Protein Sci.* **1998**, *7*, 2405-2412.
168. Buer, B. C.; Levin, B. J.; Marsh, E. N. G., *J. Am. Chem. Soc.* **2012**, *134*, 13027-13034.
169. Dill, K. A., *Biochemistry.* **1990**, *29*, 7133-7155.
170. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N., *Nat. Chem. Biol.* **2010**, *6*, 615-620.
171. Shortle, D., *FASEB J.* **1996**, *10*, 27-34.
172. Prabhu, N. V.; Sharp, K. A., *Annu. Rev. Phys. Chem.* **2005**, *56*, 521-548.

173. Möhle, K.; Günther, R.; Thormann, M.; Sewald, N.; Hofmann, H.-J., *Biopolymers*. **1999**, *50*, 167-184.
174. Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N., *Nat. Chem. Biol.* **2010**, *6*, 615-620.
175. Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N., *J. Am. Chem. Soc.* **2013**, *135*, 18682-18688.
176. Appella, D. H.; Christianson, L. A.; Klein, D. A.; Powell, D. R.; Huang, X.; Barchi, J. J.; Gellman, S. H., *Nature*. **1997**, *387*, 381-384.
177. Price, J. L.; Horne, W. S.; Gellman, S. H., *J. Am. Chem. Soc.* **2007**, *129*, 6376-6377.
178. LePlae, P. R.; Fisk, J. D.; Porter, E. A.; Weisblum, B.; Gellman, S. H., *J. Am. Chem. Soc.* **2002**, *124*, 6820-6821.
179. Price, J. L.; Horne, W. S.; Gellman, S. H., *J. Am. Chem. Soc.* **2010**, *132*, 12378-12387.
180. Burgess, A. W.; Leach, S. J., *Biopolymers*. **1973**, *12*, 2599-2605.
181. Crisma, M.; Toniolo, C., *Pept. Sci.* **2015**, *104*, 46-64.
182. Crisma, M.; Formaggio, F.; Pantano, M.; Valle, G.; Bonora, G. M.; Toniolo, C.; Schoemaker, H. E.; Kamphuis, J., *J. Chem. Soc., Perkin Trans. 2*. **1994**, 1735-1742.
183. Toniolo, C.; Polese, A.; Formaggio, F.; Crisma, M.; Kamphuis, J., *J. Am. Chem. Soc.* **1996**, *118*, 2744-2745.
184. Crisma, M.; Moretto, A.; Rainaldi, M.; Formaggio, F.; Broxterman, Q. B.; Kaptein, B.; Toniolo, C., *J. Pept. Sci.* **2003**, *9*, 620-637.
185. O'Neil, K. T.; DeGrado, W. F., *Science*. **1990**, *250*, 646-651.
186. Ratnaparkhi, G. S.; Awasthi, S. K.; Rani, P.; Balaram, P.; Varadarajan, R., *Protein Eng.* **2000**, *13*, 697-702.



187. Werner, H. M.; Cabaltega, C. C.; Horne, W. S., *ChemBioChem*. **2015**, n/a-n/a.
188. Lelais, G.; Campo, M. A.; Kopp, S.; Seebach, D., *Helv. Chim. Acta*. **2004**, *87*, 1545-1560.
189. Liu, L.; Yang, C.; Guo, Q.-X., *Biophys. Chem*. **2000**, *84*, 239-251.
190. Lumry, R.; Rajender, S., *Biopolymers*. **1970**, *9*, 1125-1227.
191. Sharp, K., *Protein Sci*. **2001**, *10*, 661-667.
192. Ok, T.; Jeon, A.; Lee, J.; Lim, J. H.; Hong, C. S.; Lee, H.-S., *J. Org. Chem*. **2007**, *72*, 7390-7393.
193. Mazzini, C.; Lebreton, J.; Alphan, V.; Furstoss, R., *J. Org. Chem*. **1997**, *62*, 5215-5218.
194. Cunningham, T. F.; Putterman, M. R.; Desai, A.; Horne, W. S.; Saxena, S., *Angew. Chem. Int. Ed*. **2015**, *54*, 6330-6334.
195. Sui, Q.; Borchardt, D.; Rabenstein, D. L., *J. Am. Chem. Soc*. **2007**, *129*, 12042-12048.
196. Manavalan, P.; Momany, F. A., *Biopolymers*. **1980**, *19*, 1943-1973.
197. Hovmoller, S.; Zhou, T.; Ohlson, T., *Acta Crystallog. Sect. D*. **2002**, *58*, 768-776.
198. Lengyel, G. A.; Eddinger, G. A.; Horne, W. S., *Org. Lett*. **2013**, *15*, 944-947.
199. Hagihara, M.; Anthony, N. J.; Stout, T. J.; Clardy, J.; Schreiber, S. L., *J. Am. Chem. Soc*. **1992**, *114*, 6568-6570.
200. Bandyopadhyay, A.; Mali, S. M.; Lunawat, P.; Raja, K. M. P.; Gopi, H. N., *Org. Lett*. **2011**, *13*, 4482-4485.
201. Lengyel, G. A.; Reinert, Z. E.; Griffith, B. D.; Horne, W. S., *Org. Biomol. Chem*. **2014**, *12*, 5375-5381.

202. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D., *Chem. Biol. Drug Des.* **2013**, *81*, 136-147.
203. Maini, R.; Nguyen, D. T.; Chen, S.; Dedkova, L. M.; Chowdhury, S. R.; Alcala-Torano, R.; Hecht, S. M., *Biorg. Med. Chem.* **2013**, *21*, 1088-1096.
204. Hartman, M. C. T.; Josephson, K.; Lin, C.-W.; Szostak, J. W., *PLoS ONE*. **2007**, *2*, e972.
205. Casi, G.; Hilvert, D., *Curr. Opin. Struct. Biol.* **2003**, *13*, 589-594.
206. Levine, P. M.; Craven, T. W.; Bonneau, R.; Kirshenbaum, K., *Org. Lett.* **2014**, *16*, 512-515.
207. Fang, G.-M.; Wang, J.-X.; Liu, L., *Angew. Chem. Int. Ed.* **2012**, *51*, 10347-10350.