Chemical sequence control of crystallization in periodic polypeptides of the sequence poly(AG)XEG/

Mark Thomas Krejchi
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_1

Recommended Citation
https://doi.org/10.7275/j6p5-r403

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
CHEMICAL SEQUENCE CONTROL OF CRYSTALLIZATION IN
PERIODIC POLYPEPTIDES OF THE SEQUENCE POLY(AG)×EG

A Dissertation Presented
by
MARK THOMAS KREJCHI

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY
September 1993
Polymer Science and Engineering
CHEMICAL SEQUENCE CONTROL OF CRYSTALLIZATION IN PERIODIC POLYPEPTIDES OF THE SEQUENCE POLY(AG)xEG

A Dissertation Presented
by
MARK THOMAS KREJCHI

Approved as to style and content by:

David A. Tirrell, Chair

Maurille J. Fournier, Member

Shaw Ling Hsu, Member

Robert W. Lenz, Member

William J. MacKnight, Department Head Polymer Science and Engineering
DEDICATION

This work is dedicated to my wife, Amy, whose friendship
I will always cherish.
ACKNOWLEDGEMENTS

The interdisciplinary nature of this work has made it necessary for me to pursue many collaborations. As a result, I am indebted to a number of people whose help has made this work possible. First, I would like to thank Professor Tirrell for providing me with the opportunity and resources to work on such an interesting project and for his valuable assistance and guidance during my tenure at University of Massachusetts.

I would also like to thank Professor Edward Atkins for the stimulating late night discussions and for help regarding the x-ray data.

I would like to acknowledge and thank the members of my committee, Professors Hsu, Fournier and Lenz, for their advice during execution of this work.

I would like to thank my colleagues, Kevin McGrath, Mike Dougherty, Rich Lempicki, Katherine Bakeev-Antolin, and Scott Nitzsche for the help they provided in their particular field of expertise. In addition, I would like to thank Geni Dessipri, Ajay Parkhe and Susan Dawson for their friendship and entertaining discussions.

I would like to thank my children, Alex and Nick for being excellent distractions from the sometimes serious business of science.

Finally, I would like thank my wife, Amy, for her last minute help putting the final figures together and for her love, support and patience during this entire exercise.
ABSTRACT

CHEMICAL SEQUENCE CONTROL OF CRYSTALLIZATION IN
PERIODIC POLYPEPTIDES OF THE SEQUENCE POLY(AG)xEG
SEPTEMBER 1993
MARK THOMAS KREJCHI, B.S., SAN JOSE STATE UNIVERSITY
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: Professor David A. Tirrell

Biomolecular synthesis provides a pathway to polymeric materials in which absolute control over molecular weight, sequence, stereochemistry and composition can be achieved. By utilizing biomolecular synthesis to acquire control over molecular architecture, it should be possible to direct by design, the chain conformation and interchain organization in polymeric solids. Specifically, we are interested in exploiting the known sequence-dependent secondary structures observed in polypeptides to construct chain folded lamellar crystals of specified thickness and surface functionality. To utilize this technique for the production of polymeric materials, artificial oligonucleotide (DNA) monomers encoding two repeats of the oligopeptide sequence (AlaGly)3GluGly (1.4) were synthesized chemically. Polymerization of the DNA, and cloning and expression in a bacterial host, resulted in a polypeptide with the expected composition.

Solid state structural characterization of poly(AG)3EG demonstrates that this material adopts two crystal modifications that are isomorphic with those observed in poly-L-alanylglycine (PLAG) and Bombyx mori silk fibroin. Infrared, Raman, and cross polarization/magic angle spinning (CP/MAS) $^{13}$C NMR spectroscopic analysis of poly(AG)3EG-I (type I crystalline modification prepared by gelation of aqueous formic acid solutions) shows features characteristic of antiparallel $\beta$-sheets (ap$\beta$-
sheets). Wide angle x-ray diffraction experiments performed on mats of poly(AG)\textsubscript{3}EG, demonstrate the crystalline nature of this material, support the conclusion of the antiparallel arrangement of chains within the \(\beta\)-sheet structure, and indicate that the crystals are oriented with cylindrical symmetry about the \(a\) axis. These crystals index on an orthorhombic unit cell with dimensions \(a = 0.948\) nm, \(b = 1.060\) nm, and \(c\) (chain axis) = 0.695 nm (errors \(\pm 0.002\) nm). Selective line broadening of the wide angle diffraction signals with \(l = 1\) argues for a coherence length of approximately 2-4 nm in the \(c\) direction. Low angle x-ray diffraction measurements show that regular fluctuations in electron density, perpendicular to \(a\), occur at approximately 3 nm which are consistent with chain-folded lamellae. The results are commensurate with a model constructed from stacked, regularly chain-folded crystalline lamellae composed of polypeptide chains that reverse polarity in register with the sequence periodicity of poly(AG)\textsubscript{3}EG. Solid state structural analysis on poly(AG)\textsubscript{4}EG-I, poly(AG)\textsubscript{5}EG-I, poly(AG)\textsubscript{6}EG-I shows similar crystalline architectures and supports a model in which chain folding is directed by the chemical sequence periodicity. Although experimental evidence collected on poly(AG)\textsubscript{3}EG-I and other members in the series does not provide detailed information regarding the turn geometry, evidence collected on similar systems favors a model that exploits \(\beta\)-turns to reverse chain polarity.

Comparison of the infrared and Raman spectra of poly(AG)\textsubscript{3}EG-II (type II crystalline modification prepared by dialysis of aqueous LiBr solutions) and poly(AG)\textsubscript{3}EG-I shows that substantial differences exist in the frequencies and intensities of the observed bands, suggesting that these materials have different crystal structures. The CP/MAS \(^{13}\)C NMR spectrum of poly(AG)\textsubscript{3}EG-I reinforces this conclusion and provides evidence that the solid state structure of poly(AG)\textsubscript{3}EG-II is isomorphic with PLAG-II and \(B.\) \textit{mori} silk-I. Wide angle x-ray diffraction patterns of poly(AG)\textsubscript{3}EG-II exhibit discrete Bragg diffraction signals that index on an
orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.734$ nm, and $c$ (chain axis) = 0.940 nm (errors ±0.002 nm). These unit cell dimensions are consistent with those observed in PLAG-II and B. mori silk-I. Based on the detailed structural analysis of poly(AG)3EG-I, the experimental evidence collected on poly(AG)3EG-II supports a crystalline structure composed of chain folded lamellae constructed from the lateral stacking of sheets, in which the peptide backbone is highly contracted in comparison with the apβ-sheet structure. In this model, the proposed contraction results from the glycine residues adopting a left handed α-helical conformation similar to that proposed by Lotz and coworkers for the structure of PLAG-II. As a result of the absence of any low angle signals in the x-ray diffraction patterns of poly(AG)3EG-II, no statements regarding the specific details of chain folding can be made.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Historical Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The Synthetic Method</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 Manipulation of DNA Fragments</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Introduction of Vector DNA into a Host</td>
<td>6</td>
</tr>
<tr>
<td>1.2.3 Potential Problems</td>
<td>8</td>
</tr>
<tr>
<td>1.2.4 Expression Systems</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Materials applications</td>
<td>11</td>
</tr>
<tr>
<td>1.4 Project Description</td>
<td>12</td>
</tr>
<tr>
<td>1.5 Type II Crystal Structure of <em>B. mori</em> Silk Fibroin</td>
<td>15</td>
</tr>
<tr>
<td>1.6 Type I Crystal Structure of Poly-L-alanylglcyline</td>
<td>17</td>
</tr>
<tr>
<td>1.7 The Other Crystalline Modification Observed in <em>B. mori</em> Silk Fibroin and PLAG</td>
<td>26</td>
</tr>
<tr>
<td>1.8 Polymer Sequence Design</td>
<td>28</td>
</tr>
<tr>
<td>2. EXPERIMENTAEN</td>
<td>32</td>
</tr>
<tr>
<td>2.1 Materials</td>
<td>32</td>
</tr>
<tr>
<td>2.1.1 Reagents</td>
<td>32</td>
</tr>
<tr>
<td>2.1.2 Enzymes</td>
<td>35</td>
</tr>
<tr>
<td>2.1.3 Stock Solutions</td>
<td>35</td>
</tr>
</tbody>
</table>
2.1.4 Buffer Solutions ........................................................................ 36
  2.1.4.1 10X TBE Buffer ................................................................. 36
  2.1.4.2 Stacking Gel Buffer .......................................................... 37
  2.1.4.3 Separation Gel Buffer ....................................................... 37
  2.1.4.4 Running Buffer ................................................................. 37
  2.1.4.5 GTE Buffer ....................................................................... 37
  2.1.4.6 TFB 1 Buffer ..................................................................... 38
  2.1.4.7 TFB 2 Buffer ..................................................................... 38
  2.1.4.8 Lysis Buffer ....................................................................... 38

2.1.5 Additional Solutions .................................................................. 39
  2.1.5.1 25X Amino Acid Solution ................................................ 39
  2.1.5.2 Acrylamide Solution .......................................................... 39
  2.1.5.3 Phenol Solution ................................................................. 39
  2.1.5.4 Staining Solution ............................................................... 39
  2.1.5.5 Destaining Solution ........................................................... 40

2.1.6 Media ....................................................................................... 40
  2.1.6.1 2X YT Medium ................................................................. 40
  2.1.6.2 YT Medium ....................................................................... 40
  2.1.6.3 M9AA Medium ................................................................. 40

2.2 Methods ....................................................................................... 41
  2.2.1 General Procedures ............................................................... 41
    2.2.1.1 Preparation of Competent Cells ...................................... 41
    2.2.1.2 Preparation of Frozen Primary Stock .............................. 41
    2.2.1.3 Cell Titering ................................................................. 42
    2.2.1.4 Transformation of Host Cells with Plasmid DNA .......... 42
    2.2.1.5 Isolation of Plasmid DNA ............................................... 43
    2.2.1.6 Double-strand Plasmid DNA Sequencing ...................... 43
2.2.2 Preparation of Target Monomer DNA ........................................... 44
  2.2.2.1 Synthesis and Purification of the Single-stranded 64-mer Oligonucleotides .......................................................................................... 44
  2.2.2.2 Phosphorylation and Annealing of the Oligonucleotides ...................................................................................................................... 46
  2.2.2.3 Insertion of the Oligonucleotide Duplex into pUC18 ............................................................................................................................ 46
  2.2.2.4 Transformation of Host Cells with pUC18-oligonucleotide Duplex ........................................................................................................... 47
  2.2.2.5 Identification of Transformants by Restriction Analysis ...................................................................................................................... 47
  2.2.2.6 Isolation of the 48 Base Pair Monomer .................................... 48

2.2.3 Preparation of Ban I Multimers .................................................... 49
  2.2.3.1 Polymerization of the Ban I Monomer Fragment .................... 49
  2.2.3.2 Insertion of the Ban I Multimers into p937.51 ....................... 49
  2.2.3.3 Transformation of Host Cells with p937-multimer Ligation Mix .......................................................................................................... 50
  2.2.3.4 Identification of Transformants by Bam HI Digestion .......... 51
  2.2.3.5 Isolation of the Bam HI Fragment from p937.M18 ................ 51

2.2.4 Preparation of the Expression System ........................................ 52
  2.2.4.1 Insertion of Octadecamer into pET3-b .................................. 52
  2.2.4.2 Transformation of Host Cells with pET3-b-octadecamer Ligation Mix .................................................................................................................. 53
  2.2.4.3 Identification of Transformants by Bam HI and Ava I Digestion ........................................................................................................ 53
  2.2.4.4 Transformation of Host with pET1-M18 ................................ 54

2.2.5 Preparation of Poly(AG)3EG-fusion (-F) ..................................... 55
  2.2.5.1 Expression of Poly(AG)3EG-F in Host Cells ........................ 55
2.2.5.2  Purification of Poly(AG)3EG-F .................................. 56
2.2.5.3  Large Scale Protein Expression of Poly(AG)3EG-F .................... 57
2.2.5.4  Large Scale Purification of Poly(AG)3EG-F ................................ 57
2.2.6  Preparation of Poly(AG)3EG .................................................. 58
  2.2.6.1  Cyanogen Bromide Cleavage of Poly(AG)3EG-F .......................... 58
2.2.7  Crystallization of Recombinant Polypeptides .................................. 59
  2.2.7.1  Preparation of Poly(AG)3EG-I .............................................. 59
  2.2.7.2  Preparation of Poly(AG)4EG-I .............................................. 59
  2.2.7.3  Preparation of Poly(AG)5EG-I .............................................. 60
  2.2.7.4  Preparation of Poly(AG)6EG-I .............................................. 60
  2.2.7.5  Preparation of Type I Crystal Mats ....................................... 60
  2.2.7.6  Preparation of Poly(AG)3EG-II ............................................. 61

2.3  Measurements ............................................................................. 61

3.  RESULTS AND DISCUSSION ........................................................... 63
  3.1  Construction of the Bacterial Expression System ............................... 63
  3.2  Expression of Poly(AG)3EG-F ................................................... 68
  3.3  Purification of Poly(AG)3EG-F .................................................. 72
  3.4  Preparation of Poly(AG)3EG ...................................................... 75
  3.5  Thermal Analysis of Poly(AG)3EG .............................................. 81
  3.6  Solid-state Structural Characterization of Poly(AG)3EG-I .................. 84
     3.6.1  Sample Preparation ............................................................ 84
     3.6.2  Infrared Spectroscopy ......................................................... 85
     3.6.3  Raman Spectroscopy ........................................................... 91
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6.4</td>
<td>CP/MAS Solid-state $^{13}$C NMR Spectroscopy</td>
<td>93</td>
</tr>
<tr>
<td>3.6.5</td>
<td>X-ray Diffraction</td>
<td>95</td>
</tr>
<tr>
<td>3.6.6</td>
<td>Thermal Treatment</td>
<td>101</td>
</tr>
<tr>
<td>3.7</td>
<td>Solid-state Structural Characterization of Proteins of the Sequence Poly(AG)$_x$EG-I</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>3.7.1 Sample Preparation</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>3.7.2 Infrared Spectroscopy</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>3.7.3 X-ray Diffraction</td>
<td>109</td>
</tr>
<tr>
<td>3.8</td>
<td>Model Proposed for the Crystal Structures of Proteins of the Sequence Poly(AG)$_x$EG-I</td>
<td>115</td>
</tr>
<tr>
<td>3.9</td>
<td>Alternative Models</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>3.9.1 Variations in Folding in the Model-I Structure</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>3.9.2 Variations in Intersheet Stacking in the Model-I Structure</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>3.9.3 Variations in Turn Conformation in the Model-I Structure</td>
<td>121</td>
</tr>
<tr>
<td>3.10</td>
<td>Solid-state Structural Characterization of Poly(AG)$_3$EG-II</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>3.10.1 Sample Preparation</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>3.10.2 Infrared Spectroscopy</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>3.10.3 Raman Spectroscopy</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>3.10.4 CP/MAS Solid-state $^{13}$C NMR Spectroscopy</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>3.10.5 X-ray Diffraction</td>
<td>133</td>
</tr>
<tr>
<td>3.11</td>
<td>Model Proposed for the Crystal Structure of Poly(AG)$_3$EG-II</td>
<td>136</td>
</tr>
<tr>
<td>3.12</td>
<td>Effect of Chemical Sequence on the Relative Stability of the Poly(AG)$_3$EG-II Crystal Structure</td>
<td>137</td>
</tr>
<tr>
<td>4.</td>
<td>CONCLUSIONS AND FUTURE WORK</td>
<td>139</td>
</tr>
<tr>
<td>4.1</td>
<td>Conclusions</td>
<td>139</td>
</tr>
<tr>
<td>4.2</td>
<td>Future Work</td>
<td>141</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>74</td>
</tr>
<tr>
<td>3.2</td>
<td>75</td>
</tr>
<tr>
<td>3.3</td>
<td>79</td>
</tr>
<tr>
<td>3.4</td>
<td>88</td>
</tr>
<tr>
<td>3.5</td>
<td>88</td>
</tr>
<tr>
<td>3.6</td>
<td>93</td>
</tr>
<tr>
<td>3.7</td>
<td>95</td>
</tr>
<tr>
<td>3.8</td>
<td>98</td>
</tr>
<tr>
<td>3.9</td>
<td>107</td>
</tr>
<tr>
<td>3.10</td>
<td>112</td>
</tr>
<tr>
<td>3.11</td>
<td>113</td>
</tr>
</tbody>
</table>
3.12 Comparison of observed diffraction signal spacings \( (d_0) \) in poly(AG)\(_6\)EG\(_4\)-I with those calculated \( (d_c) \) for an orthorhombic unit cell with dimensions \( a = 0.964 \) nm, \( b = 0.962 \) nm, \( c = 0.695 \) nm (errors \( \pm 0.002 \) nm)................................. 114

3.13 Frequencies and assignments for selected vibrations observed in the infrared spectrum of poly(AG)\(_3\)EG-II............................................. 125

3.14 Frequencies and assignments for selected vibrations observed in the Raman spectrum of poly(AG)\(_3\)EG-II................................. 129

3.15 Chemical shifts and assignments for peaks observed in the CP/MAS \(^{13}\)C NMR spectra of poly(AG)\(_3\)EG-II, PLAG-II (70) and silk-I (70)........................................................................... 132

3.16 Comparison of the observed diffraction signal spacings \( (d_0) \) in poly(AG)\(_3\)EG-II with those calculated \( (d_c) \) for an orthorhombic unit cell with dimensions \( a = 0.948 \) nm, \( b = 1.734 \) nm, \( c = 0.940 \) nm (errors \( \pm 0.002 \) nm)........................................135
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A schematic illustration of the methodological approach toward protein biosynthesis</td>
<td>4</td>
</tr>
<tr>
<td>1.2</td>
<td>Potential problems that may arise during key steps in protein biosynthesis</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>A schematic illustration of the approach to crystal engineering with polypeptides</td>
<td>13</td>
</tr>
<tr>
<td>1.4</td>
<td>A schematic illustration of the structure proposed for <em>Chrysopa</em> silk (43)</td>
<td>14</td>
</tr>
<tr>
<td>1.5</td>
<td>A schematic illustration of the structure proposed for the silk-II modification observed in <em>B. mori</em> silk fibroin (46)</td>
<td>16</td>
</tr>
<tr>
<td>1.6</td>
<td>Antiparallel B-sheet structures (64). The rippled sheet structure illustrated as projections of the (a) a c and the (b) b c planes. The pleated sheet structure illustrated as projections of the (c) a c and the (d) b c planes.</td>
<td>19</td>
</tr>
<tr>
<td>1.7</td>
<td>An illustration of the vibrational modes of a peptide bond in an antiparallel B-sheet structure as predicted by Miyazawa (53)</td>
<td>21</td>
</tr>
<tr>
<td>1.8</td>
<td>Models of four possible sheet structures formed by hydrogen bonding of polypeptide chains (66)</td>
<td>25</td>
</tr>
<tr>
<td>1.9</td>
<td>Proposed lamellar structure for proteins of sequence 1.4</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Identification of clones (C1-C6) containing recombinant plasmid by restriction analysis with <em>Ban</em> I (even lanes) and <em>Eco Rl/Bam</em> HI (odd lanes). Lanes M1 and M2 contain pBR322-msp I and λ-<em>Hind</em> III digested DNA as molecular weight standards. Lane X is uncut pUC-18 as a digestion control.</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Analysis of coding sequence of four recombinant plasmids by double strand DNA sequencing. The lanes are identified by the nucleotide used to terminate the chain extension reaction. The sequence shown is the anticoding strand.</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>Screening recombinant multimers of p937.51 for size identification by <em>Bam</em> HI digestion of clones 1-11. Lanes M1 and M2 contain the <em>Hind</em> III and BstE II digests of λ-DNA as molecular weight standards.</td>
<td>67</td>
</tr>
</tbody>
</table>
3.4 Cell growth curves for (1) BL21(DE3) pLysS pET1-M18, (2) BL21(DE3) pLysS, (3) BL21(DE3) pET3-b and (4) BL21(DE3) pLysS pET1-M18 (no induction). Cell growth is measured as a change in the optical density of the culture as a function of time. Time zero indicates the point of addition of IPTG ................................................................. 70

3.5 Electrophoretic analysis of protein expression by in vivo labeling of proteins derived from; (1) BL21(DE3) pLysS, (2) BL21(DE3) pET3-b and (3) BL21(DE3) pLysS pET1-M18 (no induction), (4) BL21(DE3) pLysS pET1-M18 at various time points after the addition of IPTG .................................................................................. 71

3.6 Electrophoretic analysis of poly(AG)3EG-F initial purification. Proteins were visualized by (a) Coomassie Blue staining and (b) radiolabeling. Lanes M, 1, 2, 3 and 4 are protein molecular weight standards, crude lysate, pH 5.0 supernatant, pH 4.5 supernatant, pH 4.0 supernatant .......................................................... 73

3.7 Selective peptide cleavage at methionine with cyanogen bromide ........................................................................................................... 77

3.8 300 MHz $^1$H NMR spectrum of poly(AG)3EG in formic acid-d$_1$ Chemical shifts were referenced to the acyl proton (*) in the formic acid ........................................................................... 80

3.9 TGA thermogram of poly(AG)3EG .................................................................................................................. 82

3.10 DSC thermogram of poly(AG)3EG ................................................................................................................. 83

3.11 Infrared spectrum of poly(AG)3EG-I ............................................................................................................... 87

3.12 Expansion of the amide I and II region of the infrared spectrum of poly(AG)3EG-I .................................................... 89

3.13 Expansion of the amide III-V region of the infrared spectrum of poly(AG)3EG-I ................................................... 90

3.14 Raman spectrum of poly(AG)3EG-I .................................................................................................................. 92

3.15 CP/MAS $^{13}$C NMR spectrum of poly(AG)3EG-I. Chemical shifts are reported downfield from 1,4-di-t-butylbenzene. ssb; spinning side band .................................................................................. 94
X-ray diffraction patterns from oriented samples of poly(AG)3EG-I: (a) Wide angle x-ray pattern taken using a Statton camera with the beam parallel to the plane of the mat. (b) Wide angle x-ray pattern taken using a Debye-Scherrer camera (Ni filter removed) while rotating the sample through 360° continuously. (c) Schematic illustration of the prominent diffraction signals observed with indexing details. The symbol LX refers to the low angle signal. (d) Low angle diffraction pattern taken using a Statton camera with the beam parallel to the plane of a stack of mats.

X-ray diffraction patterns from oriented samples of poly(AG)3EG-I swollen in glycerol: (a) Wide angle and (b) low angle x-ray patterns taken using a Statton camera with the beam parallel to the plane of the stacked mats.

X-ray diffraction patterns from thermally treated oriented samples of poly(AG)3EG-I: (a) Wide angle and (b) low angle x-ray patterns taken using a Statton camera with the beam parallel to the plane of the stacked mats.

Overlay of the infrared spectra of: (1) thermally treated poly(AG)3EG-I and (2) as matted poly(AG)3EG-I.

Infrared spectra of the protein series: (1) poly(AG)3EG-I, (2) poly(AG)4EG-I, (3) poly(AG)5EG-I and (4) poly(AG)6EG-I.

Expansion of the amide I, II and III regions in the infrared spectra of the protein series: (1) poly(AG)3EG-I, (2) poly(AG)4EG-I, (3) poly(AG)5EG-I and (4) poly(AG)6EG-I.

Wide (upper case) and low angle (lower case) x-ray diffraction patterns from the protein series: (a) poly(AG)3EG-I, (b) poly(AG)4EG-I, (c) poly(AG)5EG-I and (d) poly(AG)6EG-I.

A schematic illustration of the model-I crystal structure.

Overlay of the infrared spectrum of (1) poly(AG)3EG-I and (2) poly(AG)3EG-II.

Overlay of the amide I and II region of the infrared spectrum of (1) poly(AG)3EG-I and (2) poly(AG)3EG-II.

Overlay of the Raman spectrum of (1) poly(AG)3EG-II and (2) poly(AG)3EG-I.

Expansion of the amide I and II region of the Raman spectrum of poly(AG)3EG-II.

CP/MAS $^{13}$C NMR spectrum of poly(AG)3EG-II.
Wide angle x-ray diffraction patterns from powders of (a) poly(AG)3EG-II and (b) poly(AG)3EG-I taken using a Statton camera.
CHAPTER 1

INTRODUCTION

1.1 Historical Background

The central focus of synthetic chemistry is to create new materials through control of molecular architecture. This involves implementing systematic chemical strategies that allow absolute control of chemical connectivity and stereochemistry. Chemists have been quite successful in controlling reactions of small molecules to achieve this objective (1-7). However, the development of synthetic routes to polymeric materials providing an equivalent level of architectural control has been relatively unsuccessful. Based on the polymerization mechanism, the conventional synthetic routes to polymeric materials can be divided into two main categories: (i) step growth polymerization or (ii) chain growth polymerization (8). The materials prepared by either of these pathways are characterized by substantial heterogeneity in chain length, sequence, composition, and stereochemistry. The statistical nature of these materials simply reflects the chemistry of the process used to prepare them. Under these conditions, chemical control can be exercised in a statistical sense only and considerable skill is required to tailor even the average properties of the chain population. If one is to acquire the ability to manipulate polymer molecules in more specific ways, obtaining absolute control over the variables that define molecular architecture is of critical importance.

To date, the most powerful of the synthetic approaches for controlling polymer chain architecture are living polymerization (9, 10) and Ziegler-Natta catalysis (11, 12). Living polymerizations proceed via the chain growth mechanism but, in the absence of any spontaneous termination events. This method provides an effective means for
controlling polymer molecular weight and dispersity. The living nature of the propagating species provides a route to macromolecular architectures (13) such as, block, comb, and star copolymers in addition to telechelic polymers, that are inaccessible by other methods. Ziegler-Natta catalysis initiates a polymerization which proceeds via a chain growth, coordination-type mechanism where the configuration of the coordination complex directs the insertion of the monomer molecules into the growing polymer chain. Under these conditions, propagation occurs in a stereospecific fashion thus providing a means for exercising stereochemical control within the chain. The importance of this catalyst system is largely due to its ability to polymerize \( \alpha \)-olefins in a stereoregular manner and is primarily responsible for the annual domestic production of four billion pounds of polypropylene with a market value of approximately one billion dollars (14).

Even given their successes, each of these synthetic approaches constitutes only a partial solution to the problems encountered in achieving architectural control. To circumvent these problems we have initiated a program exploiting the synthetic machinery of the cell for preparing polymeric materials with precisely defined molecular architectures. The basis of our approach is founded in a process referred to as genetic recombination, the breakage and rejoining of DNA molecules. The recent advances in the synthesis, cloning, and expression of artificial genes have provided the ability to construct DNA molecules containing genes and controlling elements from virtually any source, and clone, propagate and express these new genes in suitable host organisms, particularly bacteria (15). The ability to construct, disseminate and express quite novel genetic combinations has created an opportunity for the integration of biomolecular synthesis into the mainstream of polymer materials science. An approach which exploits this synthetic pathway, provides an unprecedented opportunity to achieve absolute control over polymer molecular weight, sequence, composition and stereochemistry.
1.2. The Synthetic Method

To harness the synthetic machinery of the cell for the preparation of new materials one must employ the techniques available through genetic engineering. The fundamental concern of genetic engineering is the creation and manipulation of new combinations of genetic material and the transfer of that DNA to suitable host organisms where it can be amplified and selected for or expressed as proteins.

The methodological approach toward the application of this gene technology for the preparation of new protein polymers is outlined in Figure 1.1. The process is initiated by identifying specific issues concerning the polymer materials science community and then, utilizing our knowledge of structural biology, chemistry, and physics, we design target polymers that address those issues. The information regarding these issues is encoded into a sequence of primary amino acids comprising the protein polymer and is then converted into a corresponding sequence of DNA. This oligonucleotide, prepared chemically via solid-phase organic synthesis (16), is assembled into an artificial gene that encodes the repetitive protein polymer of interest.

1.2.1 Manipulation of DNA Fragments

A fundamental step in the assembly of synthetic genes involves the cutting and splicing of DNA fragments. These procedures are conducted in vitro, through enzymatic reactions performed by restriction nucleases and DNA ligases, respectively (17). Restriction enzymes are divided into three categories (types I, II and III), according to their recognition site cleavage patterns (18).
Figure 1.1. A schematic illustration of the methodological approach toward protein biosynthesis.
The most commonly used are type II restriction endonucleases. These restriction enzymes recognize and cleave DNA within a particular sequence. Generally, these sequences are four to seven base pairs in length and have an axis of rotational symmetry. Some type II enzymes, upon cleavage of the DNA, generate fragments with protruding 3' or 5'-termini (sticky ends); others give rise to flush termini (blunt ends). Table 1.1 lists some of the more commonly used type II restriction enzymes with their recognition sequences and cleavage sites (18).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Target Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ava I</td>
<td>C*TCGCGG</td>
</tr>
<tr>
<td>Bam HI</td>
<td>G*GATCC</td>
</tr>
<tr>
<td>Bgl II</td>
<td>A*GATCT</td>
</tr>
<tr>
<td>Eco RI</td>
<td>G*AATTC</td>
</tr>
<tr>
<td>Eco RV</td>
<td>GAT*ATC</td>
</tr>
<tr>
<td>Hae III</td>
<td>GG*CC</td>
</tr>
<tr>
<td>Hha I</td>
<td>GCG*C</td>
</tr>
<tr>
<td>Hind III</td>
<td>A*AGCTT</td>
</tr>
<tr>
<td>Pst I</td>
<td>CTGCA*G</td>
</tr>
<tr>
<td>Sma I</td>
<td>CCC*GGG</td>
</tr>
<tr>
<td>Xma I</td>
<td>C*CCGGGG</td>
</tr>
</tbody>
</table>

When a restriction endonuclease produces DNA fragments with sticky ends, these fragments can associate by hydrogen bonding between overlapping termini. The resulting joint between the fragments contains two nicks a few base pairs apart in
opposite strands. DNA ligase can be used to repair these nicks to form an intact recombinant molecule. This process is relatively efficient and has been used extensively to create recombinant molecules. Although less efficient, this procedure can also be used to catalyze blunt end ligation.

1.2.2 Introduction of Vector DNA into a Host

Once the synthetic gene is assembled and the recombinant vector is prepared, a method is needed for introducing this vector into the bacterial cell were it can be selected for and amplified. The method utilized for accomplishing this task is determined by the cloning vehicle used to prepare the recombinant vector. An extremely efficient method for transport of recombinant DNA into a host cell is to package the recombinant molecule into the head of a phage (19). The phage is a bacterial virus that, under these circumstances, acts as a transport vehicle that injects the recombinant DNA into the bacterial cell during transfection. The in vitro packaging requires a minimal length of DNA and the presence of the cos sites, unique endonucleolytic cleavage sites that are used for the insertion of passenger DNA. Several vector systems utilizing bacteriophage lambda or cosmid cloning vehicles have been developed for this purpose and can yield as many as $2(10)^{10}$ plaques from the equivalent of one $\mu$g of lambda DNA.

If these vectors cannot be used, the technique of choice is bacterial transformation (20). This procedure involves the introduction of recombinant plasmid DNA into competent cells by transport across the cell membrane. Plasmids are circular extrachromosomal DNA molecules that have the ability to replicate independently of the host chromosome. The typical transformation protocol yields approximately $1(10)^{7}$ transformants per $\mu$g of DNA. A transformation efficiency of this magnitude means only 1 out of $1(10)^{3}$ cells actually takes up a recombinant molecule. As a result, highly
Efficient screening procedures have been developed for selecting cells containing recombinant plasmids.

A particularly elegant system, developed for screening purposes, is based on the insertional inactivation of a gene encoding the $\alpha$-domain of $\beta$-galactosidase (21). In this system, active $\beta$-galactosidase is composed of two fragments. Cells and plasmids are constructed such that the gene encoding the $\alpha$-fragment is located within the plasmid and the gene encoding the $\beta$-fragment is located on the host genome. These plasmids generally carry an antibiotic resistance gene and are designed so insertion of the passenger DNA into the plasmid inactivates the gene encoding the $\alpha$-domain. Cells harboring recombinant DNA are selected for on the basis of antibiotic resistance and color (22).

The procedure involves plating freshly transformed cells on a growth medium containing a selective antibiotic and 5-bromo-4-chloro-3-indolyl-$\beta$-D-galactopyranoside (X-gal). Growth of cells unsuccessfully transformed with plasmid will be prevented by the presence of antibiotic. In cells transformed with native plasmid, the active enzyme is formed and catalyzes the hydrolysis of X-gal to 5-bromo-4-chloroindoxyl which is subsequently oxidized in air to the blue dye 5,5'-dibromo-4,4-dichloro-indigo. Thus, bacterial colonies containing active enzyme are easily recognized by their blue color. Clones harboring recombinant plasmid are prevented from synthesizing the $\alpha$-fragment; this precludes the formation of active enzyme and results in the formation of white colonies. Therefore, positive clones (white colonies) can be identified against a background of blue colonies.

Once identified, the cells containing recombinant plasmid are purified and the recombinant plasmid is isolated and sequenced to verify the coding region of the synthetic gene. If this recombinant plasmid contains the appropriate controlling elements for protein expression, it is transferred directly to a suitable host organism where protein expression is carried out. After product isolation and purification, the
protein polymers are processed and experiments are conducted to determine if the characteristics they were engineered to possess have been realized.

1.2.3 Potential Problems

The production of pure materials required for the success of this approach relies on the \textit{in vivo} expression of artificial genes. As Figure 1.2 illustrates, there are several key steps involved in the biosynthesis of proteins.

Figure 1.2. Potential problems that may arise during key steps in protein biosynthesis.
The perpetuation of the DNA coding sequence within the subsequent generations of the host organism relies on the faithful replication of these DNA sequences during cell growth. The transfer of the information encoded in the DNA through the expression of protein requires the DNA to be accurately transcribed into the corresponding messenger RNA and this mRNA to be translated into the protein of interest. Biological systems have developed elaborate synthetic machinery to accomplish this task for their survival. However, as Figure 1.2 illustrates, potential problem areas exist when converting artificial genes into protein-like polymers. These include genetic instability, mRNA synthesis and stability, translational efficiency, and protein toxicity and instability.

Foreign DNA sequences, particularly the highly repetitive sequences that result from the approach outlined above, are especially susceptible to these problems (23). In general, these sequences impose a metabolic burden on the bacterial host without conferring any competitive growth advantage; as a result, the rearrangement or complete elimination of such DNA sequences from a dividing cell population can occur quite readily (24). In addition, protein synthesis requires adequate accumulation of nascent mRNA and efficient translation of that mRNA. The specific sequence requirements for efficient processing during these steps are poorly understood. A particular mRNA sequence may not accumulate to adequate levels or may result in inhibition of initiation or may prevent elongation during translation (25). Finally, foreign proteins may be susceptible to rapid proteolysis by host enzymes or be highly toxic to the cell resulting in slow growth or cell death; thus preventing high levels of protein accumulation (26).

1.2.4 Expression Systems

In response to the issues stated above, a special class of vectors, termed expression vectors, has been designed and developed (15). These vectors not only contain the structural features required for cloning, transfer, and replication of DNA but
also provide the structural elements necessary for RNA to be translated into protein. In most cases, these vector constructions are designed to maximize protein expression through optimization of regulatory sequences such as promoters, operators, and ribosomal binding sites and increase protein accumulation through the production of translational fusion proteins that result in the formation of inclusion bodies or contain signal sequences to direct transport of the fusion protein outside the cell (27).

The T7 host-vector expression system developed by Studier and coworkers (28) is an excellent example of such a construction. In this system, transcription of the target DNA is driven by T7 RNA polymerase. This enzyme actively transcribes mRNAs under the control of a T7 promoter at a rate five times that of Escherichia coli RNA polymerase. The bacterial host for this expression system is a B strain E. coli that is deficient in the lon and ompt proteases and carries a chromosomal copy of the gene for T7 RNA polymerase under control of the inducible lacUV5 promoter. This configuration offers the potential advantage of high levels of protein expression and accumulation while allowing reasonably good control over protein induction. The vectors, developed to express the target DNAs, were constructed from pBR322, a multicopy plasmid that confers ampicillin resistance to its host. Upstream from the unique Bam HI cloning site, these vectors contain the φ10 promoter region, one of six strong promoters in T7 DNA. In addition, synthetic genes inserted into the Bam HI cloning site can utilized the natural translational start signal for the major capsid protein of bacteriophage T7, one of the most efficient start sequences identified in bacteriophage T7 (29). Synthetic genes, lacking a translational start, inserted into the Bam HI site can utilize this start signal for initiation of protein synthesis from the target mRNA. Downstream from the unique Bam HI cloning site, the Tφ fragment was inserted to direct efficient termination of transcription by T7 RNA polymerase.

The synthesis of T7 RNA polymerase is induced by the presence of isopropyl-β-D-thiogalactopyranoside (IPTG). This enzyme transcribes actively and selectively only
DNA under the control of a T7 promoter. Under optimal conditions, most of the cellular resources are devoted to the production of target mRNA and protein, and as a result, target protein can constitute over 50% of the total cellular protein. In the cases where suppression of basal level T7 RNA polymerase activity is required, this host-vector expression system can be equipped with a plasmid that will provide a low level of T7 lysozyme, an inhibitor of T7 RNA polymerase.

1.3 Materials Applications

Evolution has resulted in the complex biochemical machinery required to prepare very precisely defined materials. In fact, living systems are a complex assembly of these materials performing a remarkable variety of mechanical, chemical, electrical, and optical functions at the macroscopic and the microscopic levels. By exercising control over the biological processes responsible for the preparation of these materials, we can exploit the high fidelity of these processes to prepare materials that are inaccessible using conventional synthetic methodologies. The essence of our approach involves the exploitation of a biological system to prepare materials that have no biological analog, but where existing biological materials may serve as paradigms in their design.

At the forefront of research involving this approach, groups are exploiting the principles of genetic engineering, protein chemistry, and protein design to develop materials with potentially interesting chemical (30-33), physical (34, 35), and mechanical properties (36, 37). In many cases, natural proteins such as the silks, collagens, elastins and bioadhesives have served as the paradigms for the design of these materials. The characteristic physical and chemical properties of these extracellular matrix proteins are conferred by repetitive amino acid sequence elements which makes them excellent candidates for an approach as outlined in Figure 1.1. In general, the focus of these research efforts has been to exploit this sequence periodicity to duplicate
the desired physicochemical properties in materials that are more easily fabricated into articles that are of technological interest.

An excellent example of just such a research effort is the work being conducted by Cappello and coworkers (38). These researchers have prepared copolymers constructed from blocks of a silk-like protein periodically disrupted by a cell-binding region of fibronectin. These hybrid materials combine the chemical and physical resistance to denaturation of the silk-like structure with specific biological activity, provided by the cell binding region of fibronectin. Even after heat treatment in an autoclave at 120°C for 20 minutes, these materials retain their high biological activity and demonstrate the ability to specifically bind mammalian cells.

1.4 Project Description

The focus of this research project is the design and expression of recombinant proteins with the potential to form chain folded lamellar crystals of predetermined thickness and surface functionality. Since the preparation of single crystals of polyethylene by Keller (39), it has been known that many semicrystalline polymers adopt a chain folded architecture in the solid state. However, there are no general methods available to control the lamellar thickness, which is most often determined by the kinetics of the crystallization process rather than the thermodynamics of that process (40). In principle, by controlling the chemical periodicity within the primary structure of the polymer, it may be possible to induce a spatial periodicity in the solid and direct crystallization. As illustrated in Figure 1.3, by integrating regions of specific secondary structures (41), such as helices, β-sheets and reverse turns in target polypeptides, and the utilization of genetic engineering for their preparation, it should be possible to exercise the level of control required to enable us to define the lamellar thickness and surface functionality of the resulting crystals.
To successfully engineer lamellar crystals, it is necessary to control the intra- and inter-molecular noncovalent connectivity between molecules and overcome the unfavorable entropy associated with assembling molecules into a compact properly folded structure. This can be accomplished through the development of hydrogen bonding networks. Dreyfuss and coworkers (42) have observed that the characteristic fold periodicity within nylon lamellar crystals contains 16 hydrogen bonds. Proteins can be considered chemically decorated analogs of nylon-2, suggesting that a polypeptide designed with a chemical periodicity constructed from at least 8 amino acids should be sufficient to induce the spatial periodicity required to direct crystallization.
Experimental evidence obtained on the crystalline structure of the protein silk in the egg stalks of the green lace-wing fly *Chrysopa flava* supports this conclusion (43). A schematic representation of the model proposed for the crystalline structure of this material is illustrated in Figure 1.4.

![Figure 1.4](image-url)
This structure, referred to as cross-β, is constructed by the regular folding of a single polypeptide chain such that the chain axis between the folds lies in a direction perpendicular to the fibre axis. The short extended segments between the folds have adopted an antiparallel β-sheet (αβ-sheet) architecture and the observed lamellar thickness of this material is approximately 2.5 nm, suggesting that the polypeptide chains are folding at every eighth amino acid.

To design the secondary structural elements for directing the intracrystalline chain organization, advantage was taken of the sequence specific secondary structures observed in natural proteins and their homologs (44). The design of these chain folded lamellae is based on the periodic disruption of a stem segment by a sequence of amino acids that has a high propensity to form reverse turns. The amino acid sequence that will constitute the secondary structural element used to construct the stem portion of these crystals is founded in the polypeptide sequence proposed for the type I crystalline modification observed in poly-L-alanylglucose (PLAG-I). A related crystalline modification is observed in Bombyx mori silk fibroin (referred to as type II).

1.5 Type II Crystal Structure of B. mori Silk Fibroin

In general, the primary sequence of B. mori silk fibroin can be characterized by a Gly-X repeat where X is Ala or Ser in a ratio of two to one, respectively (45). The crystalline domains in B. mori silk fibroin can exist in two structures referred to as silk-I and silk-II (45). A detailed model for silk-II has been proposed by Marsh and coworkers (46). They proposed an orthogonal pseudo unit cell with dimensions \( a = 0.940 \text{ nm}, b = 0.920 \text{ nm}, \) and \( c = 0.697 \text{ nm}, \) where \( a, b, \) and \( c \) are defined as the directions parallel to the hydrogen bonds, the intersheet spacing and the chain axis, respectively. The derivation of this model is based on the interplanar spacings observed
for two strong meridional diffraction signals, indexed 200 and 210, a weak diffraction signal indexed 006 and the relative intensities of the three equatorial diffraction signals indexed 010, 020, and 030. A schematic of the proposed model is illustrated in Figure 1.5.

![Figure 1.5. A schematic illustration of the structure proposed for the silk-II modification observed in B. mori silk fibroin (46).](image)

In general, this model focuses on the packing of B-pleated sheets. These sheets are constructed from polypeptide chains in an antiparallel conformation oriented such that the sheet surfaces are preferentially decorated with the side chains of alanine or glycine. The sheets pack with like surfaces in contact giving rise to an alternation of
intersheet spacing of 0.57 nm across the alanyl-alanyl contact and 0.35 nm across the glycyl-glycyl contact. In this model, the chain axes in adjacent sheets are displaced by $a/4$ to allow the interleaving of the alanyl residues of neighboring sheets.

The intersheet values observed in silk II are slightly larger than the intersheet distances of 0.535 nm and 0.344 nm observed in β-polyalanine (47) and polyglycine-I (48), respectively. Marsh and coworkers (46) proposed that the distances between adjacent sheets are determined primarily by the sizes and shapes of the side chains of the amino acid residues present in the crystalline portion and that these values reflect the incorporation of amino acid residues with side chains larger than alanine, glycine, or serine into the crystal lattice. This is consistent with the work by Warwicker on the crystal structures of various fibroins (49).

1.6 Type I Crystal Structure of Poly-L-alanylglycine

The structure proposed for silk-II was later refined by Fraser and coworkers (50) using PLAG as a model polypeptide for the crystalline regions of B. mori silk fibroin. This polypeptide is also observed in two crystalline forms that are proposed to be isomorphous with crystalline forms of silk-II and silk-I (45). They have been termed PLAG-I and PLAG-II, respectively. For the crystalline component of PLAG-I, an orthogonal unit cell with dimensions $a = 0.942$ nm, $b = 0.887$ nm and $c = 0.695$ nm was proposed. In this model, it is supposed the polypeptide chains have adopted a similar arrangement to that observed in silk-II and that the resulting antiparallel pleated sheets are oriented such that like surfaces are in contact; thus, leading to an alternation of intersheet spacing as proposed for the structure of silk-II. The main difference between the two models arises from the sharper diffraction signals observed in PLAG-I that are indexed 0k0. The spacing and intensity of the 010 and 030 diffraction signals observed in PLAG-I strongly support the alternation of intersheet spacing, however upon
quantitative evaluation of these signals, Fraser and coworkers proposed intersheet spacings of 0.517 nm across the alanyl-alanyl contact and 0.389 nm across the glycyl-glycyl contact for PLAG-I. Reinterpretation of the intensity data obtained by Marsh and coworkers (46) on silk II, based on the adjustments required for PLAG-I, led Fraser and coworkers (50) to propose a refined model for silk II that increased the intersheet distance across the glycyl-glycyl contact to 0.387 nm and decreased the intersheet distance across the alanyl-alanyl contact to 0.533 nm.

The intersheet distance across the glycyl-glycyl contact observed in PLAG-I is significantly larger than observed in polyglycine-I. This has been attributed to differences in packing resulting from the slightly different crystal structures. The structure of polyglycine-I is proposed to be an antiparallel rippled sheet (51). This structure is related to the antiparallel pleated sheet, from which it can be derived, by reflection in the plane of the sheet of alternate chains into their enantiomers (Figure 1.6). It is centrosymmetric and is only permitted for sheets constructed from chains composed of achiral residues, such as polyglycine, or when alternating chains in the sheet are composed entirely of D-residues and of L-residues, respectively. Conformational energy calculations (51) on the two structures suggest that the packing of rippled sheets is more compact than that of pleated sheets. Therefore, the intersheet distances from PLAG-I and silk-II should be compared, not to the intersheet distance observed in the rippled sheet structure of polyglycine-I; but, to the hypothetical pleated sheet structure of polyglycine. By comparing this structure with the β-pleated sheet structures of β-polyalanine and PLAG-I, Colonna-Cesari and coworkers (52) were able to account for the differences in intersheet distances observed in PLAG-I from those expected based on the distances observed in the corresponding homopolymers. In addition, Lotz and coworkers (45) suggested that extension of these ideas to the crystal structure of B. mori silk fibroin strongly supports the refinement proposed by Fraser and coworkers (50).
Figure 1.6. Antiparallel β-sheet structures (64). The rippled sheet structure illustrated as projections of the (a) $a \ c$ and the (b) $b \ c$ planes. The pleated sheet structure illustrated as projections of the (c) $a \ c$ and the (d) $b \ c$ planes.
Although the crystal structures proposed for PLAG-I and silk-II are primarily based on electron and x-ray diffraction data, a considerable amount of spectroscopic evidence has been collected supporting these crystal structures. Miyazawa (53) analyzed the normal vibrational modes of the amide groups in this structure and predicted that the unperturbed amide I vibrational frequency is split into the four components illustrated in Figure 1.7. In this Figure, the arrows represent the transition moments associated with the in-plane vibrations of the peptide group, and the phase angles in parentheses represent the phase angles between adjacent peptides in the same chain and neighboring peptides on adjacent chains, respectively.

Of the four possible amide I modes illustrated in Figure 1.7, three are infrared active and two are Raman active. The \( v(0,0) \) mode is Raman active but infrared inactive because the transition dipole moments of the four peptides add to zero. The \( v(0,\pi) \) mode is Raman inactive and weak in the infrared because the transition dipole moments of the two carbonyls point in opposite directions and cancel, but the contribution from the N-H in-plane vibrations do not. The \( v(\pi,0) \) mode is Raman inactive but strong in the infrared. The carbonyl transition moments in this mode also point in opposite directions but oscillate out of phase and therefore do not cancel. Finally, the \( v(\pi,\pi) \) mode is infrared and Raman active but very weak in intensity. This splitting gives rise to the most characteristic feature observed in the vibrational spectrum of \( \alpha \beta \)-sheets (54). Two amide I bands are observed in the infrared, a strong band near 1630 cm\(^{-1}\) and a weak band near 1690 cm\(^{-1}\), corresponding to the \( v(\pi,0) \) and \( v(0,\pi) \) modes, respectively. In the Raman spectrum, a single strong amide I band, corresponding to the \( v(0,0) \) mode, is observed near 1670 cm\(^{-1}\).

The observed frequencies and assignments for these and other vibrational modes for PLAG-I and silk-II (50) are listed in Table 1.2. The calculated frequencies for selected amide vibrational modes are included for comparison (55). The close agreement between the calculated and observed frequencies of the amide vibrational
modes in PLAG-I and silk-II clearly reinforce the structural models proposed for the crystalline components of these materials.

Figure 1.7. An illustration of the vibrational modes of a peptide bond in an antiparallel β-sheet structure as predicted by Miyazawa (53).
Table 1.2

Observed frequencies, assignments and calculated (55) values for selected vibrations in the infrared spectra of PLAG-I and silk-II (50).

<table>
<thead>
<tr>
<th>Frequency (cm⁻¹)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLAG-I</strong></td>
<td><strong>silk-II</strong></td>
<td></td>
</tr>
<tr>
<td><strong>observed</strong></td>
<td><strong>calculated</strong></td>
<td><strong>observed</strong></td>
</tr>
<tr>
<td>3292</td>
<td>3292</td>
<td>Amide A</td>
</tr>
<tr>
<td>3075</td>
<td>3084</td>
<td>Amide B</td>
</tr>
<tr>
<td>1702</td>
<td>1702</td>
<td>Amide I</td>
</tr>
<tr>
<td>1630</td>
<td>1630</td>
<td>Amide I</td>
</tr>
<tr>
<td>1535</td>
<td>1535</td>
<td>Amide II</td>
</tr>
<tr>
<td>1266</td>
<td>1265</td>
<td>Amide III</td>
</tr>
<tr>
<td>1230</td>
<td>1230</td>
<td>Amide III</td>
</tr>
<tr>
<td>1050</td>
<td>1050</td>
<td>Seryl residues</td>
</tr>
<tr>
<td>1450</td>
<td>1450</td>
<td>Alanyl residues</td>
</tr>
<tr>
<td>1447</td>
<td>1446</td>
<td>Alanyl residues</td>
</tr>
<tr>
<td>1171</td>
<td>1166</td>
<td>Alanyl residues</td>
</tr>
<tr>
<td>998</td>
<td>999</td>
<td>glycyllalanyl</td>
</tr>
<tr>
<td>977</td>
<td>977</td>
<td>glycyllalanyl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OH stretching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>methyl deformation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>methyl deformation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>methyl rocking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skeletal modes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skeletal modes</td>
</tr>
</tbody>
</table>


v(0,π) of αβ-sheet

v(π,0) of αβ-sheet

v(0,π) of αβ-sheet

v(0,π) of αβ-sheet

Seryl residues OH stretching

Seryl residues OH stretching

Alanyl residues methyl deformation

Alanyl residues methyl deformation

Alanyl residues methyl rocking

glycyllalanyl Skeletal modes

glycyllalanyl Skeletal modes
The infrared active band at 1656 cm\(^{-1}\) observed in the spectrum of silk-II, but not in PLAG-I, is not the result of amide I splitting but has been attributed to polypeptide chains which are not regularly arranged in the pleated sheets. This band was previously considered to be the result of an \(\alpha\)-helical component present in the amorphous region of silk-II (56); however, the absence of a strong band near 900 cm\(^{-1}\) in the Raman spectrum, in conjunction with the theoretical work conducted by Miyazawa and coworkers (54) suggest that it is more reasonably assigned to a random conformation. Krimm and Abe (57) suggest that the presence of this band is a consequence of parallel chain hydrogen bond defects rather than a random chain component. Work conducted by Krimm and Bandekar (58) on the calculated vibrational frequencies for \(\beta\)-turns predicts that the characteristic vibrational frequencies for \(\beta\)-turn structures show substantial overlap with those of \(\beta\)-sheets and \(\alpha\)-helical structures. This work suggests this band may be assigned to chain folding; a phenomenon not unfamiliar to synthetic polymers and proteins. In fact, this type of disorder had been envisaged as a possible explanation for the large intersheet distance between the glycyl-glycyl contacts observed in PLAG-I and silk-II (45).

As mentioned earlier, the amide I mode of ap\(\beta\)-sheet polypeptides is split into four components as a result of intermolecular and intramolecular chain coupling. This type of splitting is also proposed for the amide III modes and has been observed in several model polypeptides (55, 59-62). The Raman active band appearing in the spectrum of PLAG-I at 1271 cm\(^{-1}\) has been associated with this type of coupling (59). The band at 1263 cm\(^{-1}\) in the Raman spectrum of silk-II is within the frequency range assignable to transition dipole coupling; however, the origin of this band is not entirely clear. Since the amide III mode for disordered proteins appears near 1255 cm\(^{-1}\) in the Raman and the amide I mode of silk-II shows a band at 1656 cm\(^{-1}\) in the infrared, it is possible this band results from a disordered component. Similar bands are observed in the amide I and III regions of polyserylglycine (63). Frushour and Koenig (59) have
suggested that this band is a consequence of parallel chain hydrogen bonding defects, similar to those proposed by Krimm and Abe (57) to account for the presence of the 1656 cm⁻¹ band in the infrared spectrum of silk-II. The amide III vibrational frequencies for β-turns, as calculated by Krimm and Bandekar (64), suggest the frequency of this band is too low to be associated with chain folding.

Although the model for the structure of silk-II, as refined by Fraser and coworkers, is generally accepted, other model structures have been proposed. A parallel β-sheet model constructed from chains composed of amino acid residues adopting alternating left and right handed alpha helical conformations was proposed by Lim and Steinburg (65). However, no quantitative evaluation of structure with the experimentally observed data was performed. Takahashi and coworkers (66) believe that the anisotropy in the line broadening observed in the x-ray diffraction patterns suggests there is disorder perpendicular to the sheets and one may account for this intersheet disorder by the statistical coexistence of four sheets with different orientations at a crystal site with different probabilities. The four possible model sheet structures are illustrated in Figure 1.8. By refinement of x-ray signal intensities observed in doubly oriented samples obtained by rolling the silk gland, Takahashi and coworkers (66) propose that the crystalline regions of B. mori silk fibroin are composed of irregularly stacked sheet structures constructed predominantly from antiparallel chains where adjacent chains within the sheets are rotated by 180°. This rotation produces sheet surfaces equally decorated with alanine, glycine, and serine side chain residues resulting in an antipolar sheet packing arrangement in contrast to the polar sheet packing arrangement proposed by Marsh and coworkers (46) (Figure 1.5). However, in their analysis, Takahashi and coworkers (66) fail to explain the intensity of the diffraction signals indexed 010 and 030 in addition to the presence of the strong reflection corresponding to 4.45Å observed in x-ray diffraction patterns of PLAG-I. These
reflections would be absent or extremely weak if packing arrangements other than the type proposed by Marsh and coworkers (46) were present.

Figure 1.8. Models of four possible sheet structures formed by hydrogen bonding of polypeptide chains (66).
Earlier it was stated that *B. mori* silk fibroin adopts two different crystalline modifications referred to as silk-I and silk-II, the latter of which is the more stable crystal structure. These two crystalline modifications were shown to be isomorphous with the two crystalline modifications observed in PLAG. There has been considerable controversy surrounding the exact structure assigned to the silk-I and PLAG-II modifications and several model structures have been proposed for this crystalline structure. In general, there are two strong diffraction signals characteristic of silk-I and PLAG-II that correspond to 0.45 nm and 0.72 nm spacings. The indexing of these two signals depends on which model one chooses. However, the 0.45 nm spacing is generally considered to be a consequence of the inter-chain (hydrogen bond) dimension and the 0.72 nm spacing is generally considered to be a consequence of the inter-sheet dimension. Lotz and Keith (67) proposed a crank-shaft model that consists of polypeptide chains in an antiparallel sheet conformation with orthorhombic unit cell dimensions \( a = 0.944 \) nm, \( b = 1.44 \) nm, and \( c = 0.960 \) nm. This model is constructed from glycine and alanine residues adopting dihedral angles that approximate \( \alpha \)-helical and extended conformations, respectively. A chain constructed from amino acid residues adopting these conformations results in a contracted antiparallel pleated sheet structure with a residue translation of 0.24 nm along the chain axis. Konishi and Kurokawa (68) proposed a four fold helical structure with orthorhombic unit cell dimensions \( a = 0.459 \) nm, \( b = 0.720 \) nm, and \( c = 0.908 \) nm. This structure is similar to that observed in polyglycine-II with its hydrogen bonds projecting out normal to the chain axis. The residue translation for this model is 0.227 nm along the chain axis. The main point of contention between these two models arises from the specific indexing of the 0.45 nm diffraction signal, indexed 110 by Lotz and Keith (67) and 100 by Konishi and Kurokawa (68). Both of these models are constructed from polypeptide chains.
containing a helical twist. Neither model explains how the helical twist is relieved during chain extension while making the transition from silk-I to silk-II.

The difficulty in characterizing the silk-I structure arises from an inability to obtain more highly ordered crystalline material in the silk-I form. Attempts to accomplish this have led to the conversion from the silk-I form to the silk-II form (69). In an effort to more completely characterize the conformation of the silk-I structure, Saito and coworkers (70) have examined the two crystalline modifications of silk fibroin and PLAG by CP/MAS $^{13}$C NMR spectroscopy. By observing the conformationally dependent $^{13}$C chemical shifts of various model polypeptides, silk fibroin and PLAG, they concluded the crank shaft model proposed by Lotz and Keith (67) is not an acceptable conformational model for the silk-I structure. Instead, they found that the four fold helical structure proposed by Konishi and Kurokawa (68) could account for the observed $^{13}$C chemical shifts.

Conformational energy calculations conducted by Fossey and coworkers (71) have resulted in a new proposal for the structure of silk-I with orthorhombic unit cell dimensions of $a = 0.894$ nm, $b = 1.126$ nm, and $c = 0.646$ nm. This hydrogen bonded antiparallel sheet structure is constructed from an alternating left-hand and right-hand $3_1$ helix with alanine residues configured as a right-handed helix and glycine residues configured as a left-handed helix. This architecture results in sheet surfaces equally decorated with the side chains of alanine and glycine. Since this model is constructed from alternating left and right-handed helices, there is no net helical twisting of the polypeptide chains. Therefore, as the chain extends, there is no helical twist to be relieved and only a simple rotational adjustment is required during the transition from silk-I to silk-II. This model accounts for most of the spacings observed in the x-ray diffraction of silk-I and PLAG-II and is consistent with the NMR work conducted by Saito and coworkers (70) as well as the infrared work conducted by Brack and Spach (72).
1.8 Polymer Sequence Design

By choosing to construct the stem portion of the lamellar crystals from an amino acid sequence which prefers to adopt an antiparallel extended chain architecture, it is necessary to construct the turn from amino acid residues capable of disrupting this chain extended architecture in such a way that chain folding is favored. Consensus sequences compiled by Chou and Fasman for sequence dependent secondary structures observed in globular proteins suggest: (i) that proline frequently initiates β-turns (73) and (ii) that proline and glutamic acid are poor β-sheet formers (74). These characteristics suggest that a sequence of alanylglutamic acid dyads periodically disrupted by these two amino acids may provide a driving force for chain reversal, ultimately resulting in the formation of the crystalline architecture illustrated in Figure 1.3.

This hypothesis suggests that polypeptide sequence 1.1 should be expected to form crystalline lamellae composed of stacked, folded sheets constructed from polypeptide chains adopting an α-sheet arrangement in which the periodic proline and glutamic acid residues define the lamellar thickness and decorate the lamellar surfaces.

\[-[(\text{AlaGly})_3\text{ProGluGly}]_n^-\] 1.1

In the design of 1.1, the expected lamellar thickness of ca. 3 nm is characteristic of similar cross-β structures (43), and the use of an odd number of amino acids in the repeating sequence leads to a chain arrangement which mimics that proposed for the crystal structures of silk-II (46) and PLAG-I (50). However, McGrath and coworkers (75) recently reported that solid state structural characterization of four chain length variants of nonapeptide sequence 1.1 failed to show any evidence for the anticipated lamellar structure.
We believe that the inability of polymers of sequence 1.1 to crystallize into the anticipated structure results from the use of an odd number of amino acids in the repeat sequence. Molecular modeling conducted on the folded chain of sequence 1.1 suggests that the chain trajectory at the termini of the β-strands is such that the normal β-turns can be accommodated at only one edge of a sheet constructed from repeating units consisting of odd numbers of amino acids. As a result, the array of hydrogen bonds necessary to stabilize the β-conformation cannot be maintained over the full extent of the sheet. Consequently, the folded conformation is destabilized to such an extent that the entropically favored glassy state of the solid is preferred under the experimental conditions.

In simplified terms, the development of chain folded crystals represents the energetic balance between the destabilizing effect of the fold and the stabilizing effect resulting from the juxtaposition of the stems. In support of this view, alternative designs to induce the formation of the desired chain folded architecture have proceeded along two lines: (i) changing the composition of the turn sequence to minimize its destabilizing effect on the crystallization of the stem segments and permit the formation of the desired chain folded architecture, and (ii) increasing the length of the proposed crystalline stem segment to overcome the destabilizing effect that the turn geometry has on the formation of the desired chain folded architecture. The families of protein sequences 1.2 and 1.3 have been designed to address these issues separately (76).

\[-[(\text{AlaGly})_3\text{XY}]_n^- \quad 1.2\]
where \(\text{XY}=\text{SerGly}, \text{ValGly}, \text{MetGly}, \text{GlySer}, \text{AspGly} \text{ and } \text{TyrGly}.\)

\[-[(\text{AlaGly})_x\text{ProGluGly}]_n^- \quad 1.3\]
where \(x=4, 5\text{ and } 6\)
Both of these approaches have demonstrated variable levels of success in
generating materials with architectures characteristic of crystalline structures constructed
from polypeptides adopting an apβ-sheet type conformation (77, 78). However, since
the objective of our work is to control the crystallization process such that we can define
not only the intracrystalline chain organization but also dictate the position and
periodicity of the fold defining the crystal lamellae, it is necessary to construct a series
of proteins that incorporates both of these approaches. This idea is embodied in protein
sequence 1.4.

\[-[(\text{AlaGly})_x\text{GluGly}]_n^-\] 1.4

In light of the observation that proline is rarely found to occur in four residue β-
turns connecting two apβ-strands (79); we propose that polypeptides of the sequence
1.4 are more reasonable candidates, compared to sequence 1.1, for preparing stable
lamellar structures constructed from polypeptide chains folded regularly at every glutamic
acid residue and adopting an apβ-conformation.

The predicted lamellar structure for proteins of sequence 1.4 (where x=3) is
illustrated in Figure 1.9. If crystallization occurs as proposed, the resulting crystals will
be composed of a stem sequence constructed from alanylglucose dyads disrupted by a
turn sequence containing glutamic acid and glycine residues. Regular folding is
expected to occur at every glutamic acid residue resulting in an apβ-sheet architecture
with β-sheet surfaces equally decorated by alanine and glycine side chains. If folding
can be directed as proposed, the fold periodicity (ie. lamellar thickness) will be
determined by the length of the stem segment in polymers of sequence 1.4.

We have chosen to investigate the crystallization behavior of polymers of
sequence 1.4 constructed from stems containing 3, 4, 5 or 6 alanylglucose dyads.
Figure 1.9. Proposed lamellar structure for proteins of sequence 1.4.

Conceptually, the following discussion of this work will be divided into three parts: (i) the synthesis and cloning of the synthetic gene used to encode 36 repeats of polymer sequence 1.4 (x = 3), (ii) the expression of this synthetic gene and the purification of the resulting polymer and, (iii) the solid state structural characterization of this protein as well as proteins of sequence 1.4 where x = 4, 5 and 6. The expression systems for the latter three proteins (x = 4, 5 and 6) were developed by Yoshikuni Deguchi and obtained by the author for the purpose of structural characterization. The synthesis, cloning, expression and purification of these materials will be presented elsewhere (80).
CHAPTER 2

EXPERIMENTAL

2.1 Materials

2.1.1 Reagents

The following is a list of all reagents and solvents used during the course of the work described herein. All materials were used as received. Letter codes are used to indicate the source of the material.

- Acetic acid, glacial (F)
- Acrylamide (S)
- Adenosine triphosphate (ATP) (S)
- Agarose (BRL)
- Ammonium chloride (F)
- Ammonium persulfate (BRL)
- Ammonium sulfate (F)
- Ampicillin, sodium salt (S)
- β-Isopropylthiogalactosidase (IPTG) (S)
- 2-Mercaptoethanol (A)
- Bacto-Tryptone (D)
- Bisacrylamide (S)
- Boric acid (S)
- Bovine serum albumin (S)
- 5-Bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-gal) (S)
<table>
<thead>
<tr>
<th>Chemical Name</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromophenol Blue</td>
<td>(A)</td>
</tr>
<tr>
<td>Calcium chloride</td>
<td>(F)</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>(S)</td>
</tr>
<tr>
<td>Coomassie Brilliant Blue R</td>
<td>(S)</td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>(S)</td>
</tr>
<tr>
<td>Dimethylformamide</td>
<td>(F)</td>
</tr>
<tr>
<td>Dithiothreitol (DTT)</td>
<td>(S)</td>
</tr>
<tr>
<td>Ethanol</td>
<td>(F)</td>
</tr>
<tr>
<td>Formic acid</td>
<td>(B)</td>
</tr>
<tr>
<td>Formic acid-d₁</td>
<td>(C)</td>
</tr>
<tr>
<td>Glucose</td>
<td>(A)</td>
</tr>
<tr>
<td>Glycerol</td>
<td>(F)</td>
</tr>
<tr>
<td>Glycine</td>
<td>(A)</td>
</tr>
<tr>
<td>^H-glycine</td>
<td>(DN)</td>
</tr>
<tr>
<td>Hydrogen chloride</td>
<td>(F)</td>
</tr>
<tr>
<td>L-Alanine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Arginine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Aspartic acid</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Cysteine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Glutamic acid</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Glutamine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Histidine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Isoleucine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Leucine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Lysine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Methionine</td>
<td>(S)</td>
</tr>
<tr>
<td>L-Phenylalanine</td>
<td>(S)</td>
</tr>
</tbody>
</table>
L-Proline
L-Serine
L-Threonine
L-Tryptophan
L-Tyrosine
L-Valine
Lithium bromide
Magnesium chloride
Magnesium sulfate
(2-[N-Morpholino]ethanesulfonic acid), MES
(3-[Morpholino]propanesulfonic acid), MOPS
n-Butanol
Phenol
Poly(ethylene glycol), MW 8000 (PEG)
2-Propanol
Sodium chloride
Sodium dodecylsulfate (SDS)
Sodium hydroxide
Sodium phosphate
N,N,N',N' Tetramethylethylenediamine, TEMED
Ethylenediaminetetraacetic acid (EDTA) disodium salt
(Tris[hydroxymethyl]aminomethane), Tris
Urea
Vitamin B1
Yeast extract
Zinc chloride
2.1.2 Enzymes

- Alkaline phosphatase (N)
- *Ava* I (N)
- *Bam* HI (N)
- *Ban* I (N)
- Calf intestinal alkaline phosphatase (U)
- *Eco* RI (N)
- *Sac* I (N)
- T4 DNA ligase (N)
- T4 polynucleotide kinase (N)

Aldrich = A; Baker = B; GIBCO BRL = BRL; Difco = D; Du Pont NEN = DN;
New England Biolabs = N; Sigma = S; U.S. Biochemical = U; Cambridge Isotope Labs = C.

2.1.3 Stock Solutions

All stock solutions were prepared in distilled, deionized water unless otherwise noted. An asterisk (*) indicates the solution was autoclaved at 120 °C for 20 minutes following preparation. A pound (#) indicates the solution was filter sterilized through a 0.2 µm filter following preparation.

- Adenosine triphosphate 10 mM
- Ammonium persulfate 10 %
- Ampicillin, sodium salt 200 mg/mL#
Chloramphenicol
Calcium chloride
DTT
EDTA, disodium salt
Glucose
Glycerol
IPTG
Magnesium chloride
Magnesium sulfate
PEG
Sodium acetate
Sodium chloride
SDS
Sodium hydroxide
Tris-HCl pH 8.0
Tris-HCl pH 7.5
Vitamin B1
Zinc chloride
X-gal

25 mg/ml in 95 % ethanol
1 M*
0.4 M#
0.5 M, pH to 8.0 with NaOH
1 M#
50 % (v/v)*
0.2 M in DMF#
1 M*
1 M*
40 %*
3 M, pH to 4.8 with acetic acid*
5 M*
20 %*
10 M*
1 M, pH to 8.0 with HCl
1 M, pH to 7.5 with HCl
0.2 %#
1 M*
40 mg/mL in DMF

2.1.4 Buffer Solutions

2.1.4.1 10X TBE Buffer

A solution containing Tris (216 gm), boric acid (110 gm) and disodium EDTA was prepared by the addition of deionized, distilled water (2 L).
2.1.4.2 Stacking Gel Buffer

A solution containing Tris (6 gm), disodium EDTA (1.6 mL, 0.5 M) and SDS (2 mL, 20 %) was prepared by the addition of deionized, distilled water (80 mL). The pH was adjusted to 6.8 by the addition of HCl and the volume was increased to 100 mL by the addition of deionized, distilled water.

2.1.4.3 Separation Gel Buffer

A solution containing Tris (90.75 gm), disodium EDTA (8 mL, 0.5 M) and SDS (10 mL, 20 %) was prepared by the addition of deionized, distilled water (450 mL). The pH was adjusted to 8.8 by the addition of HCl and the volume was increased to 500 mL by the addition of deionized, distilled water.

2.1.4.4 Running Buffer

A solution containing glycine (230.4 gm), disodium EDTA (32 mL, 0.5 M) and SDS (40 mL, 20 %) was prepared by the addition of deionized, distilled water (7 L). The pH was adjusted to 8.6 by the addition of Tris and the volume was increased to 8 L by the addition of deionized, distilled water.

2.1.4.5 GTE Buffer

A solution containing glucose (5 mL, 1 M), Tris-HCl (2.5 mL, 1 M, pH 7.5), and EDTA (2 mL, 0.5 M) was prepared by the addition of deionized, distilled water to a final volume of 100 mL. The solution was filter sterilized and stored at room temperature until use.
2.1.4.6 TFB 1 Buffer

A solution containing MES (1 mL, 1 M, pH 6.5), RbCl2 (10 mL, 1 M), CaCl2 (1 mL, 1 M), and MnCl2 (5 mL, 1 M) was prepared by the addition of deionized, distilled water (80 mL). The pH was adjusted to 5.8 by the addition of acetic acid and the volume was increased to 100 mL by the addition of deionized, distilled water. The solution was filter sterilized and stored at 4 °C until use.

2.1.4.7 TFB 2 Buffer

A solution containing MOPS (1 mL, 1 M, pH 6.5), RbCl2 (10 mL, 1 M), CaCl2 (7.5 mL, 1 M), and glycerol (15 mL) was prepared by the addition of deionized, distilled water (60 mL). The pH was adjusted to 6.5 by the addition of acetic acid and the volume was increased to 100 mL by the addition of deionized, distilled water. The solution was filter sterilized and stored at 4 °C until use.

2.1.4.8 Lysis Buffer

A solution containing Tris-HCl (5 mL, 1 M, pH 7.5), SDS (10 mL, 20 %), EDTA (0.400 mL, 0.5 M) and glycerol (15 mL) was prepared by the addition of deionized, distilled water to a final volume of 50 mL.
2.1.5 Additional Solutions

2.1.5.1 25X Amino Acid Solution

L-Amino acids (50 mg) were dissolved in distilled, deionized water (85 mL). The pH of the resulting solution was adjusted to 7.0 with NaOH (2 M), and the volume increased to 100 mL with distilled, deionized water. The solution was filter sterilized and stored at 4 °C in the dark.

2.1.5.2 Acrylamide Solution

Bisacrylamide (20 gm) and acrylamide (380 gm) were dissolved in distilled, deionized water (1 L). The solution was filter sterilized and stored at 4 °C in the dark.

2.1.5.3 Phenol Solution

Phenol was melted at 65 °C and extracted with an equal volume of Tris-HCl (1 M, pH 7.5) until the pH of the aqueous layer was 7.5. The resulting two phase solution was stored at 4 °C in the dark.

2.1.5.4 Staining Solution

A solution containing Coomassie Brilliant Blue R (12.5 gm), methanol (400 mL) and acetic acid (70 mL) was prepared by the addition of distilled, deionized water to a final volume of 1 L.
2.1.5.5 Destaining Solution

A solution containing methanol (200 mL) and acetic acid (100 mL) was prepared by the addition of distilled, deionized water (700 mL).

2.1.6 Media

2.1.6.1 2X YT Medium

A solution containing Bacto-Tryptone (16 gm), Yeast extract (10 gm) and NaCl (5 gm) was prepared by the addition of deionized, distilled water (1 L). The solution was autoclaved and stored at room temperature until use.

2.1.6.2 YT Medium

A solution containing Bacto-Tryptone (8 gm), Yeast extract (5 gm) and NaCl (5 gm) was prepared by the addition of deionized, distilled water (1 L). The solution was autoclaved and stored at room temperature until use.

2.1.6.3 M9AA Medium

A solution containing Na$_2$HPO$_4$ (6 gm), NaH$_2$PO$_4$ (3 gm), NH$_4$Cl (1 gm) and NaCl (0.5 gm) was prepared by the addition of deionized, distilled water (800 mL). The pH of this solution was adjusted to 7.4 by the addition of NaOH and the volume increased to 1 L by the addition of deionized, distilled water. This solution was autoclaved and stored at room temperature until use. Immediately prior to use, glucose
(20 mL, 1 M), CaCl₂ (0.1 mL, 1 M), MgSO₄ (2 mL, 1 M), 25X amino acid solution (40 mL) and Vitamin B1 (1 mL, 0.2 %) were added.

2.2 Methods

2.2.1 General Procedures

Unless otherwise noted all DNA manipulations were performed using procedures cited in "Current Protocols in Molecular Biology" (81).

2.2.1.1 Preparation of Competent Cells

Cells were prepared by growing to saturation in sterile 2X YT at 37 °C overnight. Fresh 2X YT (20 mL) was inoculated with a portion (400 μL) of the saturated culture and incubated at 37 °C for 2 hours. The cells were harvested by centrifugation at 3000 x g for 5 minutes at 4 °C and gently resuspended in cold TFB1 buffer (10mM Mes-Tris pH 6.5, 100 mM RbCl₂, 10 mM CaCl₂, 50 mM MnCl₂, final solution pH 5.8). The cells were incubated on ice for 20 minutes and pelleted by centrifugation at 3000 x g for 5 minutes at 4 °C then, gently resuspended in 2 mL of TFB2 buffer (10 mM MOPS, 75 mM CaCl₂, 10 mM RbCl₂, 15 % glycerol, final solution pH 6.5). Aliquots (100 μL) of cells were stored at -80 °C.

2.2.1.2 Preparation of Frozen Primary Stock

A colony from a fresh transformation plate was used to inoculate 2 mL of M9AA with appropriate antibiotic selection and incubated at 37 °C until the culture became lightly turbid. An aliquot of this culture (150 μL, 10⁶ dilution of original
concentration) was streaked on a plate containing antibiotics and incubated at 37 °C. As soon as the colony developed (ca. 10 hours) it was used to inoculate 5 mL of M9AA medium and incubated at 37 °C for 8 hours. Aliquots of this culture (1.5 mL) were mixed with 150 μL of 80 % glycerol in a cryovial, frozen in dry ice/ethanol and stored at -80 °C.

2.2.1.3 Cell Titering

A single colony was used to inoculate 5 mL of medium and incubated at 37 °C for 8 hours. An aliquot (100 μL, $10^6$ dilution of original concentration) was plated onto YT and YT + ampicillin (200 μg/mL). A second aliquot (100 μL, $10^5$ dilution of original concentration) was plated onto YT + ampicillin (200 μg/mL) + IPTG (1 mM) and YT + IPTG (1 mM). These plates were incubated at 37 °C for 12 hours and the colonies were counted to determine the total fraction of cells that contain inducible plasmid. The reproducibility ($±10\%$) of the experiment was estimated by plating the YT plates in triplicate.

2.2.1.4 Transformation of Host Cells with Plasmid DNA

An aliquot (100 μL) of competent cells was thawed on ice, 2 μL of ligation mix was added and the mixture was incubated on ice for 2.5 hours. The cells were heat shocked at 42 °C for 90 seconds then, 1 mL of YT was added. After incubation at 37 °C for 45 minutes the transformation mix was plated onto YT plates containing selection, if necessary, and incubated at 37 °C for 24 hours. This procedure resulted in transformation efficiencies of approximately $5 \times 10^5$ transformants per microgram of pBR322 DNA.
2.2.1.5 Isolation of Plasmid DNA

Cells containing plasmid were harvested from 1.5 mL of overnight culture by centrifugation at 13,500 x g for 45 seconds. The supernatant was discarded and the cells were resuspended in 100 μL of GTE (50 mM glucose, 25 mM Tris pH 7.5 and 10 mM EDTA). The cells were lysed by the addition of 200 mM NaOH containing 1% SDS. After gentle mixing, the solution was incubated on ice for 10 minutes and the chromosomal DNA was precipitated by the addition of NaOAc (150 μL, 3 M). After incubation on ice for an additional 10 minutes, the mixture was centrifuged at 13,500 x g for 5 minutes and the supernatant was transferred to a new tube. RNase (1 μL, 10 μg/μL, DNAse free) was added to the solution and incubated at 37 °C for 40 minutes. The solution was extracted with an equal volume of phenol/chloroform 1:1 and the plasmid DNA was precipitated by the addition of 2 volumes of 100 % ethanol. The precipitated DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 20 μL of distilled deionized water. Typical yields of plasmid DNA were 1-2 μg.

2.2.1.6 Double-strand Plasmid DNA Sequencing

Plasmid was sequenced using the Pharmacia T7 Sequencing™ kit. Denaturation of the plasmid DNA (8 μL, 2 μg) was accomplished by addition of NaOH (2 μL, 2 M). After gentle mixing, the solution was incubated at room temperature for 10 minutes, after which, NaOAc (3 μL, 3 M) and 7 μL of distilled deionized water were added. The denatured DNA template was precipitated by the addition of ethanol (60 μL, 100 %) and incubated at -20 °C for 1 hour. The precipitated DNA template was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 10 μL of distilled deionized water. Annealing of the -40 universal
primer (New England Biolabs) was accomplished by the addition of annealing buffer (2 μL) and primer solution (2 μL, 5 ng/μL) to the solution containing template DNA (10 μL). After incubation at 37°C for 20 minutes, labeling mix (3 μL), [α-35S]dATPαS (1 μL, 10 μCi/μL) and T7 DNA polymerase (2 μL, 1.5 units/μL) were added. The components were mixed gently on ice and incubated at room temperature for 5 minutes. Termination of the primer extension reaction was completed by the addition of 4.5 μL of the labeling reaction mix to four prewarmed (37 °C for at least 1 minute) sequencing mixes containing ddNTP nucleotides. After incubation at 37 °C for 5 minutes, 5 μL of stop solution was added and the solutions were stored at -80 °C until further use.

Analysis of the sequencing reactions was conducted by gel electrophoresis using a 6 % polyacrylamide gel containing 7 M urea. Aliquots (3 μL) of the sequencing reactions were heated at 95 °C for 5 minutes and loaded on a gel (prerun at 60 W for 30 minutes). Electrophoresis was carried out at 60 W until the bromophenol blue tracking dye ran off the bottom of the gel. The gel was dried down onto Whatman 3MM filter paper and exposed to a photographic film overnight at -80 °C.

2.2.2 Preparation of Target Monomer DNA

2.2.2.1 Synthesis and Purification of the Single-stranded 64-mer Oligonucleotides

The oligonucleotides were synthesized on a Biosearch Model 8700 DNA synthesizer using CED-phosphoramidite chemistry (16). Oligonucleotide synthesis was carried out 3' to 5' on a 1.0 μmole scale using controlled pore size glass supports functionalized with the first base of the 3' end of the DNA sequence. After synthesis was complete, the oligos were removed from the glass support by incubation in concentrated NH₄OH at 65 °C for 12 hours. The solutions were cooled to room temperature, centrifuged at 13,500 x g for 5 minutes, and the supernatant lyophilized to
dryness in vacuo. The pellet obtained was resuspended in 200 μL of distilled deionized water. The amount of DNA recovered was estimated by measuring the optical absorbances of the solutions at 260 nm (#181, 38.9 μg/μL, 36.8 % yield ; #182, 29.0 μg/μL, 27.3 % yield). Oligonucleotide #181: 5’- ATT TCG TAA GTT GCC GGC GCT GGT GCG GGC GAA GGT GCC G; Oligonucleotide #182: 5’ GAT CCG GCA CCT TCG CCC GCA CCA GCG CCT GCA CCT TCG CCC GCA CCA GCG CCG GCA CCT TCA G

Purification of the crude oligos was accomplished by preparative gel electrophoresis using 13 % polyacrylamide gels containing 7 M urea. Aliquots of the crude oligo solutions, (5 optical density units, 185 μg DNA) were transferred to an equal volume of formamide loading solution (90 % formamide, 0.05 % bromophenol blue, 0.5 % xylene cyanol ) and heated at 90 °C for 3 minutes. Electrophoresis was carried out at 400 V for 1.5 hours. The product bands were visualized by shadowing and excised from the gel. The gel slices were crushed and the oligos eluted in 1 mL of 500 mM NH₄OAc, 1 % SDS, 10 mM MgOAc₂ and 1 mM EDTA at 37 °C for 9 hours. The elution mixture was centrifuged at 13,500 x g for 5 minutes and the supernatant was lyophilized to dryness in vacuo. The pellet was resuspended in 200 μL of distilled deionized water and lyophilized to dryness. The white pellet was resuspended in 200 μL of distilled deionized water and NaOAc  pH 5.2 (20 μL, 3 M) was added. The oligos were precipitated by the addition of oyster glycogen (20 μL, 10 μg/μL) and 1 mL of 100 % ethanol, followed by incubation at -20 °C overnight. The DNA was collected by centrifugation at 13,500 x g for 15 minutes, washed with ethanol (500 μL, 70 %) and dried in vacuo. The DNA was resuspended in 100 μL of distilled deionized water. The quantity was estimated by measuring the optical absorbance at 260 nm (#181, 4.2 μg, 2.3 % recovery: #182, 4.2 μg, 2.3 % recovery).
2.2.2.2 Phosphorylation and Annealing of the Oligonucleotides

The purified oligos (14 μL, 27.8 pmoles) were resuspended in 80 mM Tris pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP. T4 polynucleotide kinase (2 μL, 20 units) was added and the reactions were incubated at 37 °C for 55 minutes. The enzyme was inactivated by heating at 65 °C for 10 minutes and the complementary oligonucleotides were mixed together. The resulting solution was adjusted to a final concentration of 100 mM NaCl. Annealing was completed by heating the mixture at 95 °C for 10 minutes, 65 °C for 30 minutes, 37 °C for 30 minutes and 15 °C for 60 minutes. The reaction mixture was extracted with an equal volume of phenol/chloroform 1:1 and the annealed DNA was precipitated by the addition of 2 volumes of 100 % ethanol, followed by incubation at -20°C overnight. The DNA was collected by centrifugation at 13,500 x g for 15 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 20 μL of distilled deionized water.

2.2.2.3 Insertion of the Oligonucleotide Duplex into pUC18

The plasmid pUC18 (5 μg) was dissolved in 100 μL of 10 mM Tris pH 7.5, 10 mM MgCl₂, 100 mM NaCl and 1 μg BSA (DNAse free). *Eco* RI (3 μL, 20 units/μL) and *Bam* HI (3 μL, 20 units/μL) were added and the reaction mixture was incubated at 37 °C overnight. A portion (7 μL, 0.2 pmoles) of this digest was added to the annealed duplex (4 μL, 5 pmoles) and the resulting solution was made 50 mM Tris pH 7.5, 10 mM MgCl₂, 10 mM DTT and 1 mM ATP. T4 DNA ligase (2 μL, 400 Biolab units) was added and the reaction was incubated at 15 °C overnight.
2.2.2.4 Transformation of Host Cells with pUC18-oligonucleotide Duplex

An aliquot (100 μL) of competent *E. coli* strain DH5αF' cells was thawed on ice and 2 μL of ligation mix was added. After incubation on ice for 2 hours, the cells were heat shocked at 42 °C for 90 seconds and 1 mL of 2X YT was added. The cells were incubated at 37 °C for 1 hour, plated on a YT plate containing Ampicillin (200 μg/mL), IPTG (100 μg/mL) and X-gal (200 ng/mL). The plate were incubated at 37 °C overnight. A total of 50 colonies were observed on the plate, of which 47 were colorless.

2.2.2.5 Identification of Transformants by Restriction Analysis

Plasmid DNA was isolated by the method described in section 2.1.5 from saturated liquid cultures (5 mL, 2X YT containing Ampicillin at 200 μg/mL) prepared by inoculation with white colonies chosen from the plate containing ligation transformants. Insert verification was preformed by digestion of plasmid DNA with *Ban* I (48 bp fragment) and *Eco RI/Bam HI* (64 bp fragment) endonucleases. Four out of six colonies showed the presence of insert; four were chosen for further analysis.

Fresh cultures were prepared by inoculated 2X YT media (50 mL, containing Ampicillin at 200 μg/mL) with aliquots (200 μL) of the overnight cultures. After incubation overnight at 37 °C, the plasmids were isolated by scaling up the method described in section 2.1.5. The resulting pellet was further purified by resuspending in 30 μL of distilled deionized water and precipitated by the addition of 10 μL of 25 % poly(ethylene glycol) containing 2.5 M NaCl. After incubation at 4 °C for 2 hours, the precipitated DNA was collected by centrifugation at 13,500 x g at 4 °C for 15 minutes. The supernatant was carefully removed by aspiration using a drawn-out pipette, and the pellet was washed with ethanol (500 μL, 70 %) and resuspended in 1 mL of distilled
deionized water. The target DNA sequence in each of these recombinant plasmids was verified by double strand DNA sequencing as described in section 2.1.6 and one recombinant plasmid termed pUC18-M was chosen for further analysis.

2.2.2.6 Isolation of the 48 Base Pair Monomer

Recombinant plasmid pUC18-M (500 μg) was resuspended in 1 mL of 10 mM Tris pH 7.5, 10 mM MgCl₂ containing Ban I (100 μL, 10 units/μL) endonuclease. The reaction was incubated at 37 °C overnight. The digestion mixture was extracted with an equal volume of phenol/chloroform 1:1 and NaOAc (250 μL, 3 M) was added. The plasmid DNA was precipitated by the addition of 2 volumes of 100 % ethanol. After incubation at -20°C for 1 hour, the precipitated DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (1000 μL, 70 %), and resuspended in 300 μL of distilled deionized water.

Isolation of the Ban I monomer fragment was carried out by preparative gel electrophoresis using a 10 % polyacrylamide gel. A portion of the digest (60 μL) was loaded on the gel and electrophoresed at 300 V for 2 hours. The bands were visualized by staining with ethidium bromide, and excised from the gel, and the Ban I fragments were recovered by electrolution at 100 V into NaOAc pH 4.8 (120 μL, 3 M). Oyster glycogen (3 μg, 10 μg/μL) was added to the eluent and the DNA was precipitated by the addition of 2 volumes of 100 % ethanol. After incubation at -20°C for 1 hour, the precipitated DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 50 μL of distilled, deionized water. The amount of DNA recovered (1.3 μg) was estimated by running 2 μL on a 2 % agarose gel and comparing the ethidium bromide fluorescence intensity of the monomer band with the ethidium bromide fluorescence intensity of a λ–Hind III DNA standard.
2.2.3 Preparation of Ban I Multimers

2.2.3.1 Polymerization of the Ban I Monomer Fragment

Purified Ban I monomer fragment (370 ng, 11.6 pmoles) was resuspended in 18 μL of ice cold 50 mM Tris pH 7.5, 10 mM MgCl₂, 10 mM DTT and 1 mM ATP. After incubation on ice for 5 minutes, T4 DNA ligase (2 μL, 800 Biolab units) was added and the reaction was incubated at 15 °C overnight.

2.2.3.2 Insertion of the Ban I Multimers into p937.51

The cloning vector p937.51 was isolated by the method described in section 2.1.5. Plasmid p937.51 (5 μg) was resuspended in 100 μL of 10 mM Tris pH 7.5, 10 mM MgCl₂ containing 50 units of Ban I endonuclease. The reaction mixture was incubated at 37 °C overnight. The resulting digestion mixture was extracted with an equal volume of phenol/chloroform 1:1 and NaOAc pH 5.2 (50 μL, 3 M) was added. The digested plasmid was precipitated by the addition of two volumes of 100 % ethanol followed by incubation at -20 °C for 1 hour. The DNA was pelleted by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 500 μL of 100 mM glycine pH 9.6, 1 mM MgCl₂ and 1 mM ZnCl₂. Calf intestinal alkaline phosphatase (0.05 units) was added and the reaction mixture was incubated at 37 °C for 30 minutes. The reaction was extracted with an equal volume of phenol/chloroform 1:1 and precipitated by the addition of oyster glycogen (2 μL, 10 μg/μL) and 0.5 volumes of 100 % isopropanol. After incubation at -20 °C for 2 hours, the DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 20 μL of distilled deionized water.
The Ban I digested and dephosphorylated vector was isolated by preparative gel electrophoresis using a 1.5 % agarose gel. The vector solution was loaded on the gel and electrophoresed at 100 V for 2 hours. The linearized vector was excised and isolated by electroelution at 100 V into NaOAc (120 μL, 3 M). The linearized vector was precipitated by the addition of oyster glycogen (2 μL, 10 μg/μL) and 2 volumes of 100 % ethanol. After incubation for 1 hour at -20 °C, linearized vector was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 100 μL of distilled deionized water. The amount of linearized vector recovered (2.5 μg, 25 ng/μL) was estimated by loading a portion of the vector solution on a 1.5 % agarose gel and comparing the ethidium bromide fluorescence intensity of the linearized vector band with the ethidium bromide fluorescence intensity of a λ-Hind III DNA standard.

To the Ban I monomer ligation solution was added linearized, dephosphorylated p937.51 (4 μL, 25 ng/μL), ATP (1 μL, 10 mM), T4 DNA ligase (1 uL, 400 units/μL) and ligase buffer (1 μL, 500 mM Tris pH 7.5, 100 mM MgCl2, 10 mM DTT). The reaction mixture was incubated at 15 °C for 10 hours.

2.2.3.3 Transformation of Host Cells with p937-multimer Ligation Mix

Competent cells of E. coli strain HB101 were made by the method described in section 2.1.1. An aliquot (100 μL) of cells was thawed on ice and 2 μL of the p937.M multimer ligation mix was added; the mixture was incubated on ice for 6 hours. The cells were heat shocked at 42 °C for 90 seconds, then 1 ml of YT was added. After incubation at 37 °C for 45 minutes, the transformation mix was plated onto YT plates (containing 25 μg/mL Chloramphenicol) and incubated at 37 °C. After incubation for 24 hours approximately 300 transformants were obtained.
2.2.3.4 Identification of Transformants by *Bam* HI Digestion

Plasmid DNA from 160 transformants was isolated using the method described in section 2.1.5. The plasmids isolated were resuspended in 20 µL of 10 mM Tris pH 7.5, 10 mM MgCl₂ and 100 mM NaCl containing *Bam* HI endonuclease (2 µL, 20 units/µL) and incubated at 37 ºC for 6 hours. The digests were analyzed by gel electrophoresis using 1.5 % agarose gel. Of the 160 transformants, 115 contained monomer, 38 contained dimer, 3 contained trimer, 2 contained pentamer, 1 contained octamer and 1 contained octadecamer. The plasmid, p937.M18, containing 18 repeats of the monomer was chosen for further analysis. Plasmid DNA containing this insert was isolated from 100 mL of saturated culture by a scaled up version of the method described in section 2.1.5. The plasmid DNA collected was resuspended in 500 µL of distilled deionized water.

2.2.3.5 Isolation of the *Bam* HI Fragment from p937.M18

p937.M18 (100 µL) was added to 100 µL of 100 mM Tris pH 7.5, 100 mM MgCl₂ and 1M NaCl, *Bam* HI endonuclease (100 µL, 20 units/µL) was added. The reaction mixture was brought to a final volume of 1 mL with distilled deionized water and incubated at 37 ºC for 6 hours. The digestion mixture was extracted with an equal volume of phenol/chloroform 1:1 and NaOAc pH 4.8 (500 µL, 3 M) was added. The DNA was precipitated by the addition of 2 volumes of 100 % ethanol and incubated at -20 ºC for 1 hour. After washing with ethanol (1000 µL, 70 %) the pellet was resuspended in 50 µL of distilled deionized water. The octadecamer insert was isolated by preparative gel electrophoresis using a 1.5 % agarose gel. A portion (20 µL) of this digest was loaded on the gel and electrophoresed at 100 V for 2 hours. The band corresponding to the octadecamer insert was excised and electroeluted at 100 V into
NaOAc pH 4.8 (120 μL, 3 M). The DNA was precipitated by the addition of 2 volumes of 100 % ethanol and incubated at -20 °C for 1 hour. The DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 μL, 70 %) and resuspended in 100 μL of distilled deionized water.

2.2.4 Preparation of the Expression System

2.2.4.1 Insertion of Octadecamer into pET3-b

The plasmid pET3-b (5 μg) was resuspended in 100 μL of 10 mM Tris pH 7.5, 10 mM MgCl2 containing 50 units of Bam HI endonuclease. The reaction mixture was incubated at 37 °C overnight. The digestion mixture was extracted with an equal volume of phenol/chloroform 1:1 and NaOAc pH 4.8 (50 μL, 3 M) was added. The digested plasmid was precipitated by the addition of 2 volumes of ethanol and incubated at -20 °C for 1 hour. The DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (1000 μL, 70 %) and resuspended in 500 μL of 100 mM glycine pH 9.6, 1 mM MgCl2 and 1 mM ZnCl2. Calf intestinal alkaline phosphatase (0.5 units) was added and the mixture was incubated at 37 °C for 30 minutes. The reaction mixture was extracted with an equal volume of phenol/chloroform 1:1 and precipitated by the addition of oyster glycogen (2 μL, 10 μg/μL) and 0.5 volumes of 100 % isopropanol. After incubation at -20°C for 2 hours, the DNA was collected by centrifugation at 13,500 x g, washed with ethanol (500 μL, 70 %) and resuspended in 20 μL of distilled deionized water. Linearized Bam HI digested and dephosphorylated vector was isolated by preparative gel electrophoresis using a 1.5 % agarose gel. The vector solution was loaded on a gel and electrophoresed at 100 V for 2 hours. The linearized vector was excised from the gel and electroeluted at 100 V into NaOAc pH 4.8 (120 μL, 3 M). The linearized vector was precipitated by addition of oyster glycogen (2 μL, 10 μg/μL) and
2 volumes of 100 % ethanol. After incubation at -20 °C for 1 hour, the DNA was collected by centrifugation at 13,500 x g for 5 minutes, washed with ethanol (500 µL, 70 %) and resuspended in 100 µL of distilled deionized water. The amount of linearized vector (80 µg, 0.8 µg/µL) was estimated by loading a portion of the vector solution on a 1.5 % agarose gel and after ethidium bromide staining, the fluorescence intensity of the vector band was compared to the fluorescence intensity of a DNA standard.

The octadecamer (2 µL) was added to linearized, dephosphorylated pET3-b (4 µL, 24 µg), ATP (2 µL, 10 mM), ligase buffer (2 µL, 500 mM Tris-HCl pH 7.5, 100 mM MgCl₂, 10 mM DTT) containing T4 DNA ligase (2 µL, 400 units/µL). The reaction mixture was incubated at 15 °C overnight.

2.2.4.2 Transformation of Host Cells with pET3-b-octadecamer Ligation Mix

Competent cells of E. coli strain HB101 were made by the method described in section 2.1.1. An aliquot (100 µL) of cells was thawed on ice and 2 µL of the pET3-b/ligation mix was added; the mixture was incubated on ice for 6 hours. The cells were heat shocked at 42 °C for 90 seconds, then 1 ml of YT was added. After incubation at 37 °C for 45 minutes, the transformation mix was plated onto YT plates (containing 200 µg/mL Ampicillin) and incubated at 37 °C. After incubation overnight approximately 100 transformants were obtained.

2.2.4.3 Identification of Transformants by BamHI and AvaI Digestion

Plasmid DNA from 10 transformants was isolated using the method described in section 2.1.5. The plasmid DNA isolated was resuspended in 50 µL of distilled deionized water and a portion was digested with BamHI endonuclease to determine the
Seven out of ten colonies screened contained octadecamer. The orientation of the insert was determined by digestion with Ava I endonuclease. Upon digestion with Ava I, plasmids containing insert in the correct orientation will yield 2 fragments, 3453 base pairs (bp) and 2114 bp in length. Inserts in the opposite orientation yield two fragments of 4247 bp and 1320 bp in length, respectively. Two out of the seven colonies showed the presence of the insert in the correct orientation. One of these, pET1-M18, was chosen for further analysis.

2.2.4.4 Transformation of Host Cells with pET1-M18

Competent cells of *E. coli* strain BL21(DE3) transformed with pLysS were made by the method described in section 2.1.1. pET1-M18 was isolated by the method described in section 2.1.5 and 2 μL of the resulting DNA solution was diluted to a final volume of 10 μL. An aliquot (100 μL) of competent cells was thawed on ice, after which, 2 μL of the diluted pET1-M18 solution was added to the competent cells. The reaction mixture was incubated on ice overnight. The cells were heat shocked at 42 °C for 90 seconds, after which 1 mL of YT was added. The cells were incubated at 37 °C for 45 minutes, then plated onto YT plates (containing 200 μg/mL Ampicillin and 25 μg/mL Chloramphenicol) and incubated at 37 °C. After incubation overnight approximately 150 transformants were obtained and ten were screened for the presence of insert by Bam HI digestion. All 10 showed the presence of insert. One of these was selected for further analysis.

54
2.2.5 Preparation of Poly(AG)$_3$EG-fusion (-F)

2.2.5.1 Expression of Poly(AG)$_3$EG-F in Host Cells

The expression of poly(AG)$_3$EG-F was monitored by the incorporation of $^3$H-glycine into the target protein. M9AA medium, lacking glycine, was used to grow overnight cultures. Single colonies obtained from frozen primary stock or YT plates containing BL21(DE3) pLysS pET1-M18, BL21(DE3) pLysS pKF526 (positive control), BL21(DE3) pLysS pET3-b (no insert control) and BL21(DE3) pLysS (no plasmid control) were used to inoculate 5 mL of fresh M9AA, without glycine. Incubation, with appropriate antibiotic selection, was carried out at 37 °C for 12 hours. Portions of these overnight cultures (90-100 μL, 1.455 O.D. units/μL, as measured in a 1 cm cell) were used to inoculate fresh M9AA medium (10 mL) with selection. These cultures were incubated, with vigorous agitation at 37 °C to a cell density corresponding to $A_{600} = 0.4-0.5$. $^3$H-glycine (50 μCi, 1 μCi/μL) were added and the cells were incubated at 37 °C for 15 minutes. IPTG (15 μL, 0.2 M) and an additional volume of $^3$H-glycine (100 μL, 1 μCi/μL) was added. Periodically, aliquots (1 mL) of cells were removed and the cell density at $A_{600}$ was measured to monitor cell growth rates. At several time points during cell growth, aliquots (1 mL) of cells were removed and harvested by centrifugation at 13,500 x g for 45 seconds. The cells were washed with fresh medium, resuspended in 50 μL of 10 mM Tris pH 7.5, 10 % glycerol and lysed by the addition of 50 μL of 100 mM Tris pH 7.5, 4 % SDS, 4 mM EDTA, 2 % β-mercaptoethanol and 20 % glycerol. Lysates were analyzed on a 10 % discontinuous polyacrylamide gel. The lysates (20-30 μL, 1.007 O.D. units/μL) were heated at 95 °C for 5-10 minutes and electrophoresed at 10 mA for 14 hours. Gels were stained with Coomassie Brilliant Blue R (0.125 % Coomassie, 40 % methanol, 7 % acetic acid) for 10 hours and destained overnight in an aqueous solution containing 20 % methanol, 10
% acetic acid. The gel was soaked in autoradiography enhancer (Enlighten™) for 30 minutes and dried down on Whatman 3MM filter paper at 65 °C for 60 minutes. Photographic film was exposed to the dried gel at -80 °C for 2.5 days.

2.2.5.2 Purification of Poly(AG)3EG-F

The remaining fraction of induced BL21(DE3) pLysS pET3-M18 cell culture from the in vivo labeling expression experiment was pelleted by centrifugation at 13,000 x g for 2 minutes. The cells were washed with fresh media and resuspended in 1 mL of TE (100 mM Tris, pH 7.5, 1 mM EDTA). The cells were placed on ice and cell lysis was accomplished by sonication at 40 units power (Branson Sonifier) for 30 seconds. The lysate was clarified by centrifugation at 13,000 x g for 10 minutes. The clarified supernatant was acidified to pH 5.0, pH 4.5 and pH 4.0 sequentially, by the addition of aqueous acetic acid (1%). The solution was allowed to stand at 4 °C for at least 4 hours before centrifugation (13,000 x g for 10 minutes) between pH adjustments. At each pH adjustment an aliquot (50 μL) of the supernatant was recovered and an equal volume of a solution containing 100 mM Tris pH 7.5, 4 % SDS, 4 mM EDTA, 2 % β-mercaptoethanol and 20 % glycerol was added. Samples were heated at 95°C for 5-10 minutes and analyzed on a 10 % discontinuous polyacrylamide gel. The samples were electrophoresed at 10 mA for 14 hours. The gel was stained with Coomassie Brilliant Blue R (0.125 % Coomassie, 40 % methanol, 7 % acetic acid) for 10 hours and destained overnight in an aqueous solution containing 20 % methanol, 10 % acetic acid. The gel was soaked in autoradiography enhancer (Enlighten™) for 30 minutes and dried down on Whatman 3MM filter paper at 65 °C for 60 minutes. Photographic film was exposed to the dried gel at -80 °C for 2.5 days.
2.2.5.3 Large Scale Protein Expression of Poly(AG)₃EG-F

A culture was prepared by inoculating 5 mL of YT medium containing antibiotics (200 µg/mL ampicillin, 25 µg/mL chloramphenicol) from frozen primary stock and incubated for 9 hours at 37 °C. The entire culture was used to inoculate 250 mL of YT medium containing antibiotics, and incubated with vigorous agitation overnight at 37°C. This culture was used to inoculate 12 L of YT containing antibiotics (2.4 gm Ampicillin/12 L and 0.3 gm Chloramphenicol/12 L), in a Microferm fermenter (New Brunswick Scientific Inc) and incubated at 37 °C with vigorous mixing and aeration. Antifoam A concentrate (1 mL) was added to the culture to prevent excessive foaming during incubation. After the culture reached a cell density corresponding to A₆₀₀ 0.8-1.0 (ca. 3 hours of growth), IPTG (1.14 gm/12L, 0.4 mM) was added. Growth was allowed to continue for an additional 2.5-3 hours, and the cells were harvested by centrifugation at 4000 x g for 30 minutes at 4 °C. The cell pellet was resuspended in 250mL of TE (100mM Tris pH 7.0, 1 mM EDTA) and stored at -80 °C.

The cells were placed on ice and lysed by sonication at 40 units power (Branson Sonifier) for 15 minutes. The lysate was clarified by centrifugation at 13,000 x g for 20 minutes at 4 °C followed by subsequent centrifugation of the supernatant at 19,000 x g for 20 minutes at 4 °C.

2.2.5.4 Large Scale Purification of Poly(AG)₃EG-F

The clarified supernatant obtained from protein expression was acidified to pH 5.0, pH 4.5 and pH 4.0 sequentially, by the addition of glacial acetic acid. The solution was allowed to stand at 4 °C for at least 4 hours before centrifugation (13,000 x g for 20 minutes) at 4°C between pH adjustments. The pH 4.0 supernatant was adjusted to pH 7.0 by the addition of NaOH (10M, ca.2-3 mL) and MgCl₂.
(0.26 gm/250mL) was added. DNase (1 mg) and RNAse (1 mg) were added and the solution was incubated at 37 °C with shaking for 2 hours. The resulting solution was adjusted to pH 4.0 with glacial acetic acid and made 40 % ethanol (167 mL ETOH/250 mL solution) and placed at -10 °C for 2 days. The precipitate was collected by centrifugation at 13,000 x g for 30 minutes at 4 °C, washed with distilled water for 2 days, washed with acetone twice and dried overnight in vacuo at 78 °C over P2O5. The average yield of water insoluble protein for a 12 L expression was 0.3-0.5 gm. The protein was stored desicated in the freezer.

2.2.6 Preparation of Poly(AG)3EG

2.2.6.1 Cyanogen Bromide Cleavage of Poly(AG)3EG-F

Cyanogen bromide cleavage of poly(AG)3EG was accomplished by the method of Smith (82). A portion of fusion protein (100 mg, 0.022 mmol methionine residues) was dissolved in formic acid (70 %, 25 mL) and filtered (0.4 μm). After flushing the solution with nitrogen for approximately 15 minutes, CNBr (0.9434 mmol, 100 mg) was added. The reaction mixture was allowed to stir at room temperature for 48 hours in the absence of light and dried down without heating under high vacuum. The protein poly(AG)3EG was washed several times with distilled, deionized water and acetone, and dried overnight in vacuo at room temperature.
2.2.7 Crystallization of Recombinant Polypeptides

2.2.7.1 Preparation of Poly(AG)₃EG-I

The preparation of poly(AG)₃EG-I was carried out by stirring poly(AG)₃EG (30 mg/mL) in aqueous formic acid (90 %) overnight at room temperature after which the concentration of formic acid was decreased to 70 % by the addition of deionized water. The solution was allowed to stir until gelation occurred (7 days) after which, the gelled material was washed with formic acid (70 %) and methanol, then resuspended in methanol (40 mL) and aged at -10°C for two days. Powder samples were obtained by centrifugation of the aged suspension to isolate the solid, followed by drying in vacuo overnight at room temperature. Alternatively, a more traditional approach to preparing microrytsalline powders of PLAG I was used to prepared unoriented samples (83). Protein solutions in aqueous formic acid (90 %) were precipitated by vapor diffusion of methanol. A precipitate resulted after several days. The precipitate was collected by centrifugation, washed with methanol and dried in vacuo overnight at room temperature.

2.2.7.2 Preparation of Poly(AG)₄EG-I

The preparation of poly(AG)₄EG-I was carried out by stirring poly(AG)₄EG (30 mg/mL) in aqueous formic acid (90 %) overnight at room temperature after which the concentration of formic acid was decreased to 70 % by the addition of deionized water. The solution was allowed to stir until gelation occurred (5 days) after which, the gelled material was washed with formic acid (70 %) and methanol, then resuspended in methanol (40 mL) and aged at -10 °C for two days. Powder samples were obtained by centrifugation of the aged suspension to isolate the solid, followed by drying in vacuo overnight at room temperature.
2.2.7.3 Preparation of Poly(AG)₅EG-I

The preparation of poly(AG)₅EG-I was carried out by stirring poly(AG)₅EG (30 mg/mL) in aqueous formic acid (90 %) overnight at room temperature after which the concentration of formic acid was decreased to 70 % by the addition of deionized water. The solution was allowed to stir until gelation occurred (3 days) after which, the gelled material was washed with formic acid (70 %) and methanol, then resuspended in methanol (40 mL) and aged at -10 °C for two days. Powder samples were obtained by centrifugation of the aged suspension to isolate the solid, followed by drying in vacuo overnight at room temperature.

2.2.7.4 Preparation of Poly(AG)₆EG-I

The preparation of poly(AG)₆EG-I was carried out by stirring poly(AG)₆EG (30 mg/mL) in aqueous formic acid (90 %). Gelation occurred upon dissolution after which, the gelled material was washed with formic acid (70 %) and methanol, then resuspended in methanol (40 mL) and aged at -10 °C for two days. Powder samples were obtained by centrifugation of the aged suspension to isolate the solid, followed by drying in vacuo overnight at room temperature.

2.2.7.5 Preparation of Type I Crystal Mats

Crystal mats were obtained by allowing an aged suspension of recombinant protein form I to sediment from methanol (100 mL) onto a teflon filter (10 µm) followed by removal of the methanol by slow filtration. This procedure resulted in a methanol swollen crystal mat that was placed between two teflon filters, then sandwiched between...
two pieces of Whatman filters and dried overnight at room temperature under compression between two glass plates.

2.2.7.6 Preparation of Poly(AG)3EG-II

The preparation of poly(AG)3EG-II was carried out using a standard procedure developed for the preparation of PLAG II (85). Powder samples of poly(AG)3EG-II were prepared by dissolving poly(AG)3EG-II (0.1 %) in aqueous lithium bromide and dialyzing this solution against progressively diluted solutions of lithium bromide. A precipitate developed at a lithium bromide concentration of 12 % but the dilution process was continued until all the lithium bromide was removed. The precipitate was collected by centrifugation, washed with methanol and dried overnight in vacuo at room temperature. Only 32 % of the original sample weight was recovered in this process, the remaining being precipitated by the addition of methanol. The sum total recovered using this procedure was approximately 87 %.

2.3 Measurements

DNA concentrations were measured on a Hitachi U-2000 double-beam UV/Vis spectrophotometer using quartz cuvettes with a path length of 1 cm. Comparison of the fluorescence intensities of ethidium bromide stained samples with DNA standards was used as an alternative method for estimating amounts of DNA less than 10 μg.

Optical scattering measurements at 600 nm were obtained on liquid cell cultures using a Hitachi U-2000 double-beam UV/Vis spectrophotometer. Sample volumes of 1 mL were placed in disposable polystyrene cuvettes with a path length of 1 cm.
Infrared and Raman spectra were obtained on IBM IR32 Fourier Transform infrared and Bruker FRA 106 Fourier Transform Raman spectrophotometers, respectively.

Solution $^1$H-NMR spectra were obtained on a Varian XL-300 spectrometer at a frequency of 300 MHz. Samples were dissolved in formic acid-d$_1$ at a concentration of 50 mg/mL.

Cross polarization/magic angle spinning$^{13}$C NMR spectra were obtained at 50 MHz on powder samples using a Bruker 200AC spectrometer equipped with a DOTY solids probe and an IBM solids rack. Measurements were collected at a spinning speed of ca. 4000 Hz with a 5 µs 90° pulse and a cross polarization time of 2 ms. A line broadening factor of 50 was used during data processing.

X-ray diffraction measurements on crystal mats were performed on a Rigaku Denki Statton x-ray camera using Ni filtered CuKα radiation. Wide angle measurements were recorded at a camera length of 30.27 mm using a single flake of matted material. Low angle measurements were recorded at a camera length of 170 mm using a stack of mats ca. 1 mm in thickness constructed from several flakes of matted material attached at the edges with epoxy.

Thermal gravimetric and differential scanning calorimetric measurements were performed on DuPont TGA 2950 and DSC 10 analyzers, respectively. In the case of DSC measurements an indium standard was used for calibration. Samples were dried overnight invacuo at 78 °C over P$_2$O$_5$ prior to analysis.

Elemental analyses were performed by the Microanalytical Laboratory, Office of Research Services, University of Massachusetts, Amherst, Massachusetts, 01003.

Amino acid compositional analyses were performed by the W.M. Keck Foundation Biotechnology Resource Laboratory at Boyer Center for Molecular Medicine, 295 Congress Ave., Yale University, New Haven, Connecticut, 06510.
CHAPTER 3

RESULTS AND DISCUSSION

3.1 Construction of the Bacterial Expression System

A detailed discussion of the strategy used to harness the synthetic machinery of the cell to prepare artificial proteins has been presented in section 1.2. Oligonucleotide sequence 3.1, which encodes two repeats of the octapeptide sequence 1.4 (x = 3), was synthesized using CED-phosphoramidite chemistry (16).

\[
\begin{array}{cccccccccccc}
\text{Eco RI} & & & \text{Ban I} & & & & & & & & \\
\text{AAT TCG TAA GGT GCC GGC GCT GGT GCG GGC GAA} & & & & & & & & & & & \\
\text{GC ATT CCA CGG CCG CGA CCA CGC CCG CTT} & & & & & & & & & & & \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\text{Ban I} & & & \text{Bam HI} & & & & & & & & \\
\text{GGT GCA GGC GCT GGT GCG GGC GAA GGT GCC G} & & & & & & & & & & & \\
\text{CCA CGT CCG CGA CCA CGC CCG CTT CCA CGG CCT AG} & & & & & & & & & & & \\
\end{array}
\]

3.1

The codons used to encode each amino acid reflect our desire to minimize the use of rare codons (84), avoid strict sequence periodicity and eliminate all Ban I restriction sites other than those flanking the target sequence. The single-stranded oligonucleotides were purified by electrophoresis, enzymatically phosphorylated at the 5’ termini and annealed to form a DNA duplex. These single stranded oligonucleotides were designed to contain 5’ and 3’ terminal sequences that upon annealing, are compatible with termini generated by Eco RI and Bam HI digestion respectively. After ligating the target duplex into the polylinker region of the cloning vector pUC18 (21), E. coli strain DH5αF’ cells
were transformed with recombinant plasmid and the transformants were identified by blue-white screening. The presence of the TAA stop codon immediately adjacent to the 5' Eco RI restriction site in sequence 3.1 ensures insertional inactivation of the β-galactosidase α fragment encoded by pUC18, providing for efficient screening for clones harboring the target gene by blue-white selection. Restriction analysis with Eco RI/ Bam HI and Ban I shows the 48 bp Ban I and 64 bp Eco RI/ Bam HI fragments in 4 of the 6 clones chosen, suggesting the presence of the target duplex in these clones (Figure 3.1).

Figure 3.1. Identification of clones (C1-C6) containing recombinant plasmid by restriction analysis with Ban I (even lanes) and Eco RI/Bam HI (odd lanes). Lanes M1 and M2 contain pBR322-msp I and λ-Hind III digested DNA as molecular weight standards. Lane X is uncut pUC-18 as a digestion control.
The sequences of 3 recombinant plasmids were determined by double strand DNA sequencing. Figure 3.2 demonstrates all of the plasmids contain the correct target DNA sequence however, in pUC1-M (lane 2) a single base deletion has resulted in the elimination of the Bam HI restriction site. Ban I digestion of the pUC3-M yielded a 48 base pair fragment encoding two repeats of sequence 1.4 (x = 3). The monomer DNA was purified by polyacrylamide gel electrophoresis and recovered by electroelution.

![Sequence analysis](image)

Figure 3.2. Analysis of coding sequence of four recombinant plasmids by double strand DNA sequencing. The lanes are identified by the nucleotide used to terminate the chain extension reaction. The sequence shown is the anticoding strand.
The non-palindromic cohesive ends generated by restriction with Ban I reduce the probability of inverted repeats upon polymerization of the monomer DNA, thus preserving the integrity of the coding sequence during this step. Polymerization of the monomer DNA by self-ligation yielded a population of multimers that were cloned into the unique Ban I restriction site of Ban I digested, dephosphorylated p937.51, a small high copy cloning vector that contains an origin of replication derived from pBR322 and a gene encoding chloramphenicol acetyltransferase (85). The resulting mixture was used to transform E. coli strain HB101 cells. The low efficiency of the transformation step ensures each resulting colony contains cells harboring recombinant plasmids that are homogeneous in molecular weight; thus providing an effective method for molecular weight selection of the target gene sequence. The construction of p937.51 does not allow screening for the presence of insert as in pUC18, consequently colonies were screened for recombinant plasmid by Bam HI restriction analysis. Dephosphorylation of Ban I digested p937.51 prevents recircularization of this cloning vector in the absence of insert. This prevents the transformation of clones that have not been successfully ligated with target multimers, thus simplifying the screening procedure. Figure 3.3 shows, of the 11 colonies screened, only one contained an insert of appreciable size. The other colonies had been transformed with recombinant plasmids containing insert sizes of dimer at best. Of the 160 transformants screened during this phase, 115 contained monomer inserts, 44 contained dimer to octamer inserts and only 1, p937.M18, contained an octadecamer insert. Other workers have experienced similarly low transformation efficiencies for p937.51 recombinant plasmids containing large inserts even after size fractionation (86). The reasons for this behavior are not clear, however close examination of Figure 3.3 suggests recombinant plasmids of p937.51 containing multimers of appreciable size are genetically unstable. This instability shifts the distribution of multimer sizes toward a lower degree of polymerization making it experimentally difficult to isolate transformants with large inserts.
Excision of the target DNA sequence from the recombinant plasmid p937.M18 was accomplished by *Bam* HI digestion. This DNA fragment, encoding a polypeptide containing 36 repeats of sequence 1.4 (x = 3), was purified by polyacrylamide gel electrophoresis, recovered by electroelution and inserted into *Bam* HI digested pET3-b.

Figure 3.3. Screening recombinant multimers of p937.51 for size identification by *Bam* HI digestion of clones 1-11. Lanes M1 and M2 contain the *Hind* III and *BstE* II digests of λ-DNA as molecular weight standards.
The resulting recombinant plasmids show no evidence of the instability observed in p937.M18. Clones containing the target sequence in the correct orientation were identified by restriction analysis. Under the conditions used, recombinant plasmids containing inserts with the correct orientation should yield two fragments, 3453 and 2114 base pairs in length. Inserts with the opposite orientation should yield fragments of 4247 and 1320 base pairs in length. Of the seven plasmids analyzed two contained inserts in the correct orientation. One of these, pET1-M18, was selected for protein expression.

3.2 Expression of Poly(AG)3EG-F

In constructing the expression system developed for the preparation of the target protein sequence, DNA sequences flanking the target sequence accumulate such that the target protein is expressed as a fusion containing nonrepetitive N- and C-terminal extensions of 23 and 32 residues respectively, as shown in sequence 3.2.

MASMTGGQQMGRDPMFKYSRDPMG-[GAGAGAGE]₃₆-
ARMHIRPGRYQLDPAANKARKEAAEAAATAEQ

The host used for expression of the fusion product poly(AG)₃EG-F, was *E. coli* strain BL21(DE3) (87). This strain contains a gene encoding T7 RNA polymerase incorporated into the bacterial chromosome under control of the *lacUV5* promoter. This configuration allows protein production to be induced by the addition of IPTG. This strain can be equipped with ancillary plasmids pLysE or pLysS. These plasmids confer Chloramphenicol resistance to their host and provide a low level source of T7 lysozyme that inhibits T7 RNA polymerase activity, suppressing basal level protein expression. The plasmids are constructed in two different configurations and provide different levels
of T7 lysozyme to the host organism. In pLysE, T7 lysozyme synthesis is directed by the tet promoter resulting in greater levels of accumulation than in pLysS where synthesis is directed by the ϕ8.8 promoter. The higher T7 lysozyme concentration present in a host organism transformed with pLysE results in more complete inhibition of T7 RNA polymerase activity and stronger suppression of basal level protein expression. Suppression of basal level protein expression is extremely important if the gene product of interest is toxic to the host organism. Protein expression experiments were carried out in BL21(DE3) pLysS transformed with pET1-M18. Control experiments using BL21(DE3) pLysS (no plasmid control), BL21(DE3) pET3-b (no insert control) and BL21(DE3) pLysS pET1-M18 (no induction control) were conducted in tandem. The expression of the target protein was monitored by the incorporation of 3H-glycine following induction during log phase growth. The rate of cell growth prior to induction was normal in all cases, but declined in BL21(DE3) pET3-b and BL21(DE3) pLysS pET1-M18 shortly after the addition of IPTG (Figure 3.4). The reduction in the growth rate was coincident with the accumulation of a new protein product in BL21(DE3) pLysS pET1-M18. This product was not present in BL21(DE3) pLysS, BL21(DE3) pLysS pET3-b, or uninduced BL21(DE3) pLysS pET1-M18. Figure 3.5 shows the electrophoretic analyses of crude cell lysates collected at various time points after the addition of IPTG.

The protein product migrates with an apparent molecular weight of slightly under 68,000 Da. Although this is higher than the expected molecular weight of 26,741 Da, similar proteins are observed to migrate anomalously on SDS polyacrylamide gels (75, 89). Initially, this behavior was attributed to low SDS binding resulting from the large number of acid residues present in the target protein sequence. However, anomalous migration behavior has also been observed for protein sequences constructed from neutral amino acids (89). This apparent anomalous migration can be attributed to the bias of the primary sequence of poly(AG)3EG toward amino acids of relatively low
molecular weight. In fact, one can entirely account for the apparent anomalous migration in poly(AG)3EG and other related sequences (90) by dividing the calculated molecular weight of the target protein and calibration standards by the average molecular weight of the amino acids contained within their primary sequences.

Figure 3.4. Cell growth curves for (1) BL21(DE3) pLysS pET1-M18, (2) BL21(DE3) pLysS, (3) BL21(DE3) pET3-b and (4) BL21(DE3) pLysS pET1-M18 (no induction). Cell growth is measured as a change in the optical density of the culture as a function of time. Time zero indicates the point of addition of IPTG.
Figure 3.5. Electrophoretic analysis of protein expression by *in vivo* labeling of proteins derived from; (1) BL21(DE3) pLysS, (2) BL21(DE3) pET3-b and (3) BL21(DE3) pLysS pET1-M18 (no induction), (4) BL21(DE3) pLysS pET1-M18 at various time points after the addition of IPTG.
The two bands migrating at approximately 40,000 Da have been attributed to conformational isomers of the band identified as the target protein (91). This assignment was confirmed by excising the target protein band from the gel and upon additional electrophoretic analysis of the excised band, the two bands migrating at approximately 40,000 Da were observed. It is interesting to note the dramatic reduction in background intensity observed in BL21(DE3) pLysS pET1-M18. This observation coupled with the decrease in cell growth rates upon induction demonstrates the efficiency with which the T7 expression system seizes the cellular resources for production of the poly(AG)3EG-F.

3.3 Purification of Poly(AG)3EG-F.

The initial procedure for the purification of poly(AG)3EG-F was developed on the small scale using radiolabeled material, as the target protein is poorly visualized using conventional staining techniques. Poly(AG)3EG-F was found in the soluble fraction of the crude cell lysate and a simple procedure developed to purify similar proteins was used as an initial purification step (88). This procedure involves a series of sequential pH adjustments of the supernatant resulting from cell lysis. Figure 3.6 shows the Coomassie Blue stained and radiolabeled polyacrylamide gels that result from the electrophoretic analysis of the crude lysate and the pH 5.0, pH 4.5 and pH 4.0 supernatants. Examination of Figure 3.6 demonstrates a substantial enrichment of poly(AG)3EG-F occurs in the supernatant as a result of this initial purification step. The pH 4 supernatant was adjusted to pH 7 and nucleases were added to digest any DNA and RNA remaining in the solution. Following incubation, the pH of the solution was readjusted to 4 and poly(AG)3EG-F was collected by ethanol precipitation. After several water washings, poly(AG)3EG-F was washed with acetone and dried under vacuum at 78°C. The average yield of poly(AG)3EG-F was approximately 40 mg per
liter of cell culture induced. The amino acid compositional and elemental analyses of poly(AG)3EG-F are listed in Tables 3.1 and 3.2 respectively. The results indicate that poly(AG)3EG-F contains approximately 3 % by weight waters of hydration, is relatively free from contaminating proteins and is composed of primarily alanine, glycine, and glutamic acid residues. As a result of the biased composition of poly(AG)3EG-F, the exact composition is experimentally inaccessible, as the limited dynamic range of the detector used to measure the concentrations of the derivatized amino acids present in the hydrolyzate precludes the simultaneous determination of the concentrations of trace and abundant amino acids present in the sample.

Figure 3.6. Electrophoretic analysis of poly(AG)3EG-F initial purification. Proteins were visualized by (a) Coomassie Blue staining and (b) radiolabeling. Lanes M, 1, 2, 3 and 4 are protein molecular weight standards, crude lysate, pH 5.0 supernatant, pH 4.5 supernatant, pH 4.0 supernatant.
Table 3.1

Amino acid composition of poly(AG)3EG-F.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Calculated (mole %)</th>
<th>Observed (mole %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>34.3</td>
<td>32.7</td>
</tr>
<tr>
<td>Arg</td>
<td>1.74</td>
<td>1.25</td>
</tr>
<tr>
<td>Asx</td>
<td>1.16</td>
<td>1.51</td>
</tr>
<tr>
<td>Cys</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Glx</td>
<td>12.5</td>
<td>12.6</td>
</tr>
<tr>
<td>Gly</td>
<td>43.3</td>
<td>45.4</td>
</tr>
<tr>
<td>His</td>
<td>0.291</td>
<td>0.000</td>
</tr>
<tr>
<td>Ile</td>
<td>0.291</td>
<td>0.188</td>
</tr>
<tr>
<td>Leu</td>
<td>0.581</td>
<td>0.209</td>
</tr>
<tr>
<td>Lys</td>
<td>0.872</td>
<td>0.570</td>
</tr>
<tr>
<td>Met</td>
<td>1.74</td>
<td>1.69</td>
</tr>
<tr>
<td>Phe</td>
<td>0.291</td>
<td>0.000</td>
</tr>
<tr>
<td>Pro</td>
<td>1.16</td>
<td>1.23</td>
</tr>
<tr>
<td>Ser</td>
<td>0.581</td>
<td>1.03</td>
</tr>
<tr>
<td>Thr</td>
<td>0.581</td>
<td>0.503</td>
</tr>
<tr>
<td>Trp</td>
<td>0.581</td>
<td>not determined</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.000</td>
<td>0.638</td>
</tr>
<tr>
<td>Val</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 3.2
Elemental analysis of poly(AG)3EG-F.

<table>
<thead>
<tr>
<th>Element</th>
<th>Calculated (wt. %)</th>
<th>Observed (wt. %)</th>
<th>3 % H2Oa</th>
<th>3.5 % H2Oa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>47.25</td>
<td>46.37</td>
<td>45.8</td>
<td>45.6</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>6.22</td>
<td>6.14</td>
<td>6.30</td>
<td>6.33</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>19.45</td>
<td>18.36</td>
<td>18.9</td>
<td>18.8</td>
</tr>
<tr>
<td>Ash</td>
<td>0.00</td>
<td>&lt;0.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a: Calculated composition for poly(AG)3EG-F containing 3 and 3.5 wt. % water.

3.4 Preparation of Poly(AG)3EG

The terminal extensions present in poly(AG)3EG-F can be removed by selective cleavage using CNBr following the procedure of Smith (82). The reaction shown in Figure 3.7 involves the nucleophilic displacement of bromide ion by the sulfide group in methionine to form the corresponding cyanoalkylsulfonium salt. Displacement of cyanomethylsulfide by acyl attack on the β-carbon results in formation of the corresponding cycloamidate. Subsequent hydrolysis of this cycloamidate results in cleavage of the peptide chain; converting the methionine residue into a C-terminal
homoserine lactone (92). This reaction has been shown to be a highly efficient method for removal of the fusion fragments from the repetitive polypeptide sequence of interest.

The amino acid compositional analysis of the product poly(AG)3EG (sequence 3.3) is listed in Table 3.3. This analysis indicates 98% of the amino acids present in the sample are alanine, glycine, and glutamic acid in comparison to the expected value of 99.4%. This value is well within the range of experimental error observed for this technique. The low concentrations of contaminating amino acids present in the sample indicated by this analysis, probably result from the formic acid used to prepare the sample for analysis, as has been shown for similar proteins (77). The elemental analysis of poly(AG)3EG is listed in Table 3.4. The values observed for nitrogen are lower than expected even after the addition of 2.5 weight % waters of hydration.
Figure 3.7. Selective peptide cleavage at methionine with cyanogen bromide.
Table 3.3
Amino acid composition of poly(AG)3EG.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Calculated (mole %)</th>
<th>Observed (mole %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>37.4</td>
<td>35.6</td>
</tr>
<tr>
<td>Arg</td>
<td>0.000</td>
<td>0.311</td>
</tr>
<tr>
<td>Asx</td>
<td>0.342</td>
<td>0.605</td>
</tr>
<tr>
<td>Cys</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Glx</td>
<td>12.3</td>
<td>13.1</td>
</tr>
<tr>
<td>Gly</td>
<td>49.7</td>
<td>49.3</td>
</tr>
<tr>
<td>His</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Ile</td>
<td>0.000</td>
<td>0.149</td>
</tr>
<tr>
<td>Leu</td>
<td>0.000</td>
<td>0.256</td>
</tr>
<tr>
<td>Lys</td>
<td>0.000</td>
<td>0.240</td>
</tr>
<tr>
<td>Met</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Phe</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Pro</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Ser</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Thr</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Trp</td>
<td>0.000</td>
<td>not determined</td>
</tr>
<tr>
<td>Tyr</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Val</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 3.4
Elemental analysis of poly(AG)3EG.

<table>
<thead>
<tr>
<th>Element</th>
<th>Calculated (wt. %)</th>
<th>Observed (wt. %)</th>
<th>+2.5 % H2O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>46.27</td>
<td>46.12</td>
<td>45.1</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>6.04</td>
<td>6.21</td>
<td>6.17</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>19.73</td>
<td>17.96</td>
<td>19.2</td>
</tr>
<tr>
<td>Ash</td>
<td>0.00</td>
<td>&lt;0.1</td>
<td></td>
</tr>
</tbody>
</table>

The 300 MHz $^1$H NMR spectrum of poly(AG)3EG is presented in Figure 3.8. Peak assignments have been included. The spectrum shows the protons on the β-carbon of alanine split into a doublet centered at 1.3 ppm and the proton on the α-carbon split into a quartet centered at 4.5 ppm. The complex multiplets centered at 2.0 and 2.15 ppm have been assigned to the magnetically inequivalent protons on the β-carbon in glutamic acid. The doublet of doublets centered at 2.45 ppm is assigned to the protons on the γ-carbon of glutamic acid. The singlet at 4.0 ppm has been assigned to the protons on the α-carbon in glycine. The complex multiplet appearing as a shoulder on the quartet at 4.5 ppm is assigned to the protons on the α-carbon of glutamic acid. The complex set of peaks over the range of 7.5 to 8 ppm have been assigned to the amide protons of alanine (upfield peak centered at 7.55 ppm) and glycine and glutamic acid (down field peak centered at 7.75 ppm). The observed integral ratios of the peak areas for the protons on the α-carbons in alanine, glycine and glutamic acid are 3: 4: 1, in excellent agreement with the expected values for these hydrogens in poly(AG)3EG.
Figure 3.8. 300 MHz $^1$H NMR spectrum of poly(AG)$_3$EG in formic acid-d$_1$.

Chemical shifts were referenced to the acyl proton (*) in the formic acid.
3.5 Thermal Analysis of Poly(AG)3EG

Thermogravimetric analysis performed in air on poly(AG)3EG (Figure 3.9) shows a two stage decomposition process with the onset of decomposition at 235 °C and the second stage decomposition beginning at 482 °C. The analysis also indicates that the sample contains less than 5 % water. The first stage of the decomposition process consumes approximately 60 % of the original mass with the second stage consuming the remainder. The onset of decomposition, level of bound water and decomposition profile observed in poly(AG)3EG are quite different from that observed for poly(AG)3PEG (86). These differences are most probably a consequence of the higher degree of crystallinity in poly(AG)3EG compared to that of poly(AG)3PEG. Combustion of poly(AG)3EG under a nitrogen atmosphere shifts the onset to decomposition to 245° C. Under these conditions, the second decomposition stage is not observed and 40 % of the original mass is still present after the sample has reached a temperature of 490 °C. The higher value of the onset to decomposition under nitrogen atmosphere suggests decomposition is oxidative in nature.

Differential scanning calorimetry performed on poly(AG)3EG (Figure 3.10) shows no observable transitions in the range of 45° C to decomposition at 235° C even after quenching the sample with liquid nitrogen or cooling slowly from 195° C to room temperature at a rate of 2° C/min. The apparent absence of a glass transition in poly(AG)3EG suggests a relatively high degree of crystallinity is present in the sample. By comparison, poly(AG)3PEG, shows a glass transition at 171° C (88). This observation is consistent with the higher degree of crystallinity present in poly(AG)3EG.
Figure 3.9. TGA thermogram of poly(AG)3EG.
Figure 3.10. DSC thermogram of poly(AG)3EG.
3.6 Solid-state Structural Characterization of Poly(AG)3EG-I

3.6.1 Sample Preparation

The preparation of poly(AG)3EG6-I was carried out by dissolving cyanogen bromide cleaved protein in 90 % formic acid overnight at room temperature. This solution was adjusted to 70 % formic acid by the addition of water. After stirring for 7 days, gelation occurred and the resulting gel was washed with formic acid (70 %) and methanol, then resuspended in methanol and aged at -10°C for two days. A similar gelation effect has been previously reported to occur in nylon-6 (93). It was necessary to age the methanol suspension at -10°C for at least two days to prevent flocculation of the suspension during the preparation of the sedimented crystal mat. Unoriented powders of this material were prepared by centrifugation of the aged suspension to isolate the solid, followed by drying *in vacuo* overnight at room temperature.

Crystal mats were prepared by allowing the aged suspension to sediment over the course of several hours onto a teflon filter. After sedimentation was complete, the supernatant was removed by vacuum filtration and the resulting swollen pellet was placed between two teflon filters. This sandwich was placed between two Whatman filters and compressed between two glass plates. No orientation was evident in samples allowed to dry prior to the compression step as determined by wide angle x-ray diffraction analysis. Increasing the compressive force increased the degree of orientation observed by wide angle x-ray analysis. Oriented samples were also produced by using a spatula to spread the swollen gel onto a teflon filter and allowing the resulting film to air dry.

Alternatively, a more traditional approach to preparing microcrystalline powders of PLAG-I was used to prepared unoriented samples (83). Protein solutions in aqueous formic acid were precipitated by vapor diffusion of methanol. The percentage of water
present in the formic acid is critical to the preparation of the type I crystal structure. Although a detailed analysis of the effect of water on crystallization by vapor diffusion was not carried out, it was observed that amounts exceeding 10% lead to the formation of the type II crystal structure.

3.6.2 Infrared Spectroscopy

The infrared spectrum of poly(AG)3EG-I is shown in Figure 3.11. Selected vibrational frequencies and their assignments are listed in Table 3.5. The amide I, II and III vibrational modes observed at 1623 cm\(^{-1}\), 1521 cm\(^{-1}\) and 1229 cm\(^{-1}\) in polyAG3EG36-I are characteristic of the \(\beta\)-sheet conformation (50, 55). The presence of the weak amide I component at 1698 cm\(^{-1}\) indicates the regular alternation of chain direction that is exhibited in the \(\alpha\beta\)-sheet architecture (54). The band at 1257 cm\(^{-1}\) results from splitting of the amide III band observed in \(\alpha\beta\)-sheet structures (59). Although not diagnostic for \(\alpha\beta\)-sheet structures, the vibrational frequencies of the amide A and B modes observed at 3286 cm\(^{-1}\) and 3066 cm\(^{-1}\) respectively, are consistent with this assignment. The bands observed at 996 cm\(^{-1}\) and 975 cm\(^{-1}\) have been assigned to skeletal modes associated with the regular alternation of glycyl and alanyl residues present in the sequence (94). These bands are also observed in PLAG (50). Figure 3.12 shows an expansion of the amide I and amide II regions in the infrared spectrum of poly(AG)3EG-I. Close examination the amide I region reveals the presence of four vibrational modes unrelated to an \(\alpha\beta\)-sheet type architecture. The two weak higher frequency vibrations centered at 1710 cm\(^{-1}\) and 1744 cm\(^{-1}\) can be attributed to carbonyl stretching of the side chain carboxyl groups present in the glutamic acid residues. The frequencies of these vibrations suggest the carboxyl groups exist in both monomeric and dimeric forms (95). Similar vibrational modes have been
observed in polyglutamic acid adopting the $\beta_1$ (1710 cm$^{-1}$) and $\beta_2$ (1732 cm$^{-1}$) type architectures (96). In $\beta$-polyglutamic acid, these vibrations have been attributed to carbonyl stretching of the side chain carboxyl groups adopting different conformations; hydrogen bonding in the $\beta_2$ conformation being substantially weaker than in the $\beta_1$ conformation. In the $\beta_2$ conformation, the more closely packed crystalline modification, a strong interaction between the hydrogen bonded peptide groups in one sheet and the carboxyl groups of the side chains in the adjacent sheet occurs, shifting the amide I and amide II vibrational frequencies to 1600 cm$^{-1}$ and 1554 cm$^{-1}$, respectively. This may provide explanation for the shoulder observed at 1550 cm$^{-1}$.

The amide I component observed at 1652 cm$^{-1}$ in poly(AG)3EG-I and the weaker shoulder at 1665 cm$^{-1}$ indicate that some fraction of the polypeptide chains have adopted a secondary structure different from the traditional $\alpha\beta$-sheet architecture. Calculations of the normal vibrational modes for various turn structures (64) suggest that the amide I components observed at 1660 cm$^{-1}$ and 1652 cm$^{-1}$, and the prominent shoulder at 1550 cm$^{-1}$ in the amide II region, may be of indicative reverse turn structures of either the $\beta$ or $\gamma$-type. The band at 1332 cm$^{-1}$ and the very weak broad band at 1300 cm$^{-1}$ in the amide III region (Figure 3.13) may also indicate the presence of these types of structures. The modes in the amide V (Figure 3.13) region are highly mixed and the bands are very broad making it difficult to assign distinct vibrational modes.
Figure 3.11. Infrared spectrum of poly(AG)$_3$EG-I.
Table 3.5

Frequencies and assignments for selected vibrations observed in the infrared spectrum of poly(AG)3EG-I.

<table>
<thead>
<tr>
<th>Frequency (cm⁻¹)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3550</td>
<td>Glutamic acid</td>
<td>O-H Stretch&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>3286</td>
<td>Amide A</td>
<td></td>
</tr>
<tr>
<td>3066</td>
<td>Amide B</td>
<td></td>
</tr>
<tr>
<td>1744</td>
<td>Glutamic acid</td>
<td>γ-C=O Stretch&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>1710</td>
<td>Glutamic acid</td>
<td>γ-C=O Stretch&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>1698</td>
<td>Amide I</td>
<td>ν (π,0) apβ-sheet&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>1652</td>
<td>Amide I</td>
<td>undetermined</td>
</tr>
<tr>
<td>1623</td>
<td>Amide I</td>
<td>ν (π,0) apβ-sheet&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>1550</td>
<td>Amide II</td>
<td>undetermined</td>
</tr>
<tr>
<td>1521</td>
<td>Amide II</td>
<td>ν (0,π) apβ-sheet&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>1257</td>
<td>Amide III</td>
<td>undetermined</td>
</tr>
<tr>
<td>1229</td>
<td>Amide III</td>
<td>ν (0,π) apβ-sheet&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>996</td>
<td>alanylglycine</td>
<td>skeletal modes&lt;sup&gt;c,d&lt;/sup&gt;</td>
</tr>
<tr>
<td>975</td>
<td>alanylglycine</td>
<td>skeletal modes&lt;sup&gt;c,d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> assignment based on reference 95, <sup>b</sup> assignment based on reference 96, <sup>c</sup> assignment based on reference 50, <sup>d</sup> assignment based on reference 94.
Figure 3.12. Expansion of the amide I and II region of the infrared spectrum of poly(AG)3EG-I.
Figure 3.13. Expansion of the amide III-V region of the infrared spectrum of poly(AG)₃EG-I.
3.6.3 Raman Spectroscopy

The Raman spectrum of poly(AG)3EG-I is shown in Figure 3.14. Selected vibrational frequencies and their assignments are listed in Table 3.6. The amide I band observed at 1664 cm$^{-1}$ is consistent with the calculated and observed values in PLAG (55, 59). The splitting of the amide III band into two components observed at 1260 cm$^{-1}$ and 1228 cm$^{-1}$ in poly(AG)3EG-I is characteristic of the apβ-sheet conformation (59). The broad, weak amide II band observed at 1535 cm$^{-1}$ is as expected for the apβ-sheet structure (97). The amide A mode observed at 3287 cm$^{-1}$ is consistent with this assignment. The bands observed at 1002 cm$^{-1}$ and 977 cm$^{-1}$ have been assigned to the skeletal modes previously identified in the infrared spectrum.

The strong, sharp band at 1083 cm$^{-1}$ is attributed to carbon-carbon stretching modes of the alanine residues. Similar bands are observed in poly-β-alanine and PLAG-I but not in polyglycine (59). The weak shoulder observed at 1063 cm$^{-1}$ is assigned to carbonyl-oxygen stretching of the side chain carboxyl groups present in the glutamic acid residues (98). As in the infrared spectrum, it is possible to associate the bands in the 1300-1330 cm$^{-1}$ region with turn structures (64). However discrimination between specific turn types is not possible. Identification of glutamic acid side chain conformationally sensitive bands in the Raman spectrum of poly(AG)3EG-I was not possible.
Figure 3.14. Raman spectrum of poly(AG)$_3$EG-I.
Table 3.6

Frequencies and assignments for selected vibrations observed in the Raman spectrum of poly(AG)3EG-I.

<table>
<thead>
<tr>
<th>Frequency (cm⁻¹)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3287</td>
<td>Amide A</td>
<td></td>
</tr>
<tr>
<td>1664</td>
<td>Amide I</td>
<td>v (0,0) of αβ-sheet⁸</td>
</tr>
<tr>
<td>1535</td>
<td>Amide II</td>
<td></td>
</tr>
<tr>
<td>1260</td>
<td>Amide III</td>
<td>undetermined</td>
</tr>
<tr>
<td>1228</td>
<td>Amide III</td>
<td>v (0,0) of αβ-sheet⁸</td>
</tr>
<tr>
<td>1083</td>
<td>alanine</td>
<td>C-C stretch⁸</td>
</tr>
<tr>
<td>1063</td>
<td>glutamic acid</td>
<td>C-O stretch⁹</td>
</tr>
<tr>
<td>1002</td>
<td>alanylglucose</td>
<td>skeletal modes</td>
</tr>
<tr>
<td>977</td>
<td>alanylglucose</td>
<td>skeletal modes</td>
</tr>
</tbody>
</table>

a; assignment based on reference 59, b; assignment based on reference 98.

3.6.4 CP/MAS Solid-state ¹³C NMR Spectroscopy

The CP/MAS ¹³C NMR spectrum of poly(AG)3EG-I is shown in Figure 3.15. The observed chemical shifts and their assignments for poly(AG)3EG-I and PLAG-I (71) are summarized in Table 3.7. The spectrum shows the α and β-carbons of alanine centered at 49.9 ppm and 20.7 ppm respectively. The shoulder at 43.6 ppm is assigned to the α-carbon of glycine. Although the spectral resolution does not allow specific identification of the β and γ-carbons in glutamic acid, the broad peak centered at approximately 30 ppm has been assigned to those carbons (99). The carbonyl, amide and α-carbons in glutamic acid could not be assigned. The peak centered at 171.4 ppm is assigned to the amide carbons of glycine and alanine. The chemical shifts for all of...
these carbons are consistent with those assigned to result from the peptide backbone adopting an apβ-sheet architecture (71).

Figure 3.15. CP/MAS $^{13}$C NMR spectrum of poly(AG)$_3$EG-I. Chemical shifts are reported downfield from 1,4-di-t-butylbenzene. ssb; spinning side band.

Close examination of the line shape of the β-carbon in alanine reveals the presence of a shoulder at 16.8 ppm. Deconvolution of this region into its two main components and comparison of their peak areas, suggests the ratio of alanine residues adopting a β-sheet type conformation to those adopting a different conformation is 2:1. The chemical shift of this shoulder is close to that observed for the β-carbon in alanine in the silk-I crystal structure (100), but the absence of the corresponding carbonyl peaks at 176 ppm suggests this peak is not associated with silk-I. In addition, the intensity of this peak was not changed by washing the sample with formic acid, a good solvent for
the silk-I modification, reinforcing the conclusion that a fraction of this material has not adopted the silk-I modification. The magnitude of the chemical shift difference between the main peak and its shoulder provides no clear discrimination between conformationally dependent chemical shift changes or chemical shift changes resulting from packing effects (101). However, it would be consistent with the infrared and Raman analysis to propose that this peak is a consequence of turn structures.

Table 3.7

Chemical shifts and assignments for selected peaks observed in the CP/MAS $^{13}$C NMR spectra of poly(AG)3EG-I, PLAG-I (70) and silk-II (70).

<table>
<thead>
<tr>
<th>Chemical Shift (ppm)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>polyAG3EG36-I</td>
<td>silk-II</td>
<td>PLAG-I</td>
</tr>
<tr>
<td>49.9</td>
<td>48.8</td>
<td>48.5</td>
</tr>
<tr>
<td>20.7</td>
<td>19.2</td>
<td>20.0</td>
</tr>
<tr>
<td>171.4</td>
<td>171.5</td>
<td>171.8</td>
</tr>
<tr>
<td>43.6</td>
<td>43.1</td>
<td>43.3</td>
</tr>
<tr>
<td>171.4</td>
<td>169.8</td>
<td>168.4</td>
</tr>
</tbody>
</table>

3.6.5 X-ray Diffraction

Wide angle x-ray diffraction patterns of poly(AG)3EG-I were obtained with the x-ray beam parallel to the plane of the mat. The diffraction patterns exhibit discrete Bragg diffraction signals consistent with crystalline material. A representative example is shown in Figure 3.16. The wide angle diffraction signals index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.060$ nm and $c$ (chain axis) = 0.695 nm (errors $\pm 0.002$ nm). The measured and calculated interplanar spacings and indexing details are tabulated in Table 3.8. The prominent diffraction signals are indexed in the schematic pattern shown in Figure 3.16c.
Figure 3.16. X-ray diffraction patterns from oriented samples of poly(AG)3EG-I: (a) Wide angle x-ray pattern taken using a Statton camera with the beam parallel to the plane of the mat. (b) Wide angle x-ray pattern taken using a Debye-Scherrer camera (Ni filter removed) while rotating the sample through 360° continuously. (c) Schematic illustration of the prominent diffraction signals observed with indexing details. The symbol LX refers to the low angle signal. (d) Low angle diffraction pattern taken using a Statton camera with the beam parallel to the plane of a stack of mats.

These unit cell parameters are commensurate with previously published x-ray diffraction results from various silk fibroins (49) and synthetic polypeptides (47, 50)
that exhibit similar crystalline structures. The value of the 0.948 nm spacing was assigned on the basis of its second diffraction order observed at 0.474 nm. This value is close to the characteristic hydrogen-bond distance observed in manynyls (102-106) and is attributed to the distance between hydrogen-bonded chains in the αβ-sheet protein structure. Therefore, it is possible to associate $a$ with the hydrogen-bond direction. The value of 1.060 nm for $b$, or more specifically its second diffraction order at 0.53 nm represents the average intersheet stacking periodicity. Since the amino acid side chains within the αβ-sheets decorate the sheet surfaces, the precise value for this spacing is a function of the amino acid composition (107). Values for this spacing have been reported as low as 0.34 nm for polyglycine (118), and as high as 0.79 nm for $N$. senegalensis (49) silk fibroin, a silk that contains a high percentage of amino acid residues with bulky side chains. The values reported for PLAG-I (50) and poly-β-alanine (47) are 0.44 nm and 0.54 nm respectively thus, the spacing observed for poly(AG)3EG is broadly in line with the expected.

An interesting characteristic of the oriented x-ray diffraction pattern from poly(AG)3EG-I is the fact that the orientation direction is along $a$, i.e. the hydrogen bond direction. This suggests that the orienting forces are acting on crystalline aggregates, not on individual polymer chains. The geometry of these crystalline aggregates must be such that they have a substantially greater dimension in the $a$ direction compared with the other two mutually orthogonal directions $b$ and $c$. Fraser et al. (50) obtained a similar texture by stroking viscous solutions of PLAG and the same orientation occurs in the thin fibrous egg-stalks drawn by the green lace-wing fly Chryosopa (43). In the texture observed in poly(AG)3EG-I, the directionally coincident $a$ and $a^*$ axes lie along the meridian line and the 200 and 400 diffraction signals appear as arcs on that meridian (Figure 3.16a, c). All the 0k0 and 00l diffraction signals appear on the equator, suggesting cylindrical symmetry about the $a$, $a^*$ axes (see Figure 3.16a, c). Previous analyses of the αβ-sheet structures suggest that of the 00l diffraction signals, only the
002 and 006 diffraction signals have significant intensity. In poly(AG)3EG-I, the observation of the 002 diffraction signal is obstructed because it coincides with 030 (the interplanar spacings differ by only 1.4 %), however the 006 is observed at 0.116 nm using a Debye-Scherrer x-ray diffraction camera (Figure 3.16b).

Table 3.8

Comparison of observed diffraction signal spacings \( (d_o) \) in poly(AG)3EG-I with those calculated \( (d_c) \) for an orthorhombic unit cell with dimensions

\[ a = 0.948 \text{ nm}, \ b = 1.060 \text{ nm}, \ c = 0.695 \text{ nm} \ (\text{errors} \pm 0.002 \text{ nm}). \]

<table>
<thead>
<tr>
<th>( k )</th>
<th>( l )</th>
<th>( d_o )</th>
<th>( d_c )</th>
<th>( k )</th>
<th>( l )</th>
<th>( d_o )</th>
<th>( d_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>h = 0</td>
<td>10</td>
<td>1.060</td>
<td>1.060</td>
<td>h = 2</td>
<td>00</td>
<td>0.474</td>
<td>0.474</td>
</tr>
<tr>
<td>20</td>
<td>0.530</td>
<td>0.530</td>
<td>10</td>
<td>0.437</td>
<td>0.433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.424</td>
<td>0.421</td>
<td>11</td>
<td>0.368</td>
<td>0.367</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.353</td>
<td>0.353</td>
<td>h = 3</td>
<td>01</td>
<td>0.285</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>0.348</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.330</td>
<td>h = 4</td>
<td>00</td>
<td>0.241</td>
<td>0.238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0.215</td>
<td>10</td>
<td>0.230</td>
<td>0.231</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.298</td>
<td>20</td>
<td>0.212</td>
<td>0.216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>0.116</td>
<td>0.116</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.151</td>
<td>0.151</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differences in line broadening of the various diffraction signals which relate in a consistent way to their Miller indices \( hkl \) are observed in poly(AG)3EG-I. The 200 and the 400 arcs are particularly sharp. This demands long range correlation order along \( a \), supporting the conclusion the crystallites are of substantial length along the \( a \)
diffraction. Diffraction signals with indices of the type hk1 are considerably broader than those with indices hk0. In particular, we estimate that the 211 diffraction signal is approximately 10 times broader than its 210 near neighbor (Figure 3.16a, b). These two diffraction signals are sufficiently close to each other in reciprocal space to make a realistic comparison and yet not compromised by other diffraction signals generated by the lattice. This comparison gives an unmistakable indication that the coherent scattering length in the c direction is small (2-4 nm) compared with the other crystallographic directions. The relative broadness of the 010 and 020 diffraction signals allows us to establish that the coherent scattering length in the b direction, or the intersheet direction, is intermediate between those in the a and c directions.

A low angle x-ray diffraction pattern from poly(AG)3EG-I taken with the beam parallel to the plane of a stack of mats is shown in Figure 3.16d. This pattern is composed of a prominent arced diffraction signal focused on the equator corresponding to spacing of 3.6 ± 0.1 nm, together with a weaker and diffuse order. This low angle pattern resembles an interparticle interference function and is similar to that observed in other chain folded systems (109). A spacing of this magnitude oriented along the mat normal is indicative of stacks of lamellae, as are usually found in crystal mats (110), suggesting the chains are folded and the chain axis is oriented perpendicular to the lamellar surface. This diffraction signal disappears after swelling the stacked mats in glycerol. In contrast, with the exception of a small change in the degree of orientation, the character and spacings observed for all other diffraction signals remain unchanged (Figure 3.17). These results indicate that this long period is associated with the dimensions of some chain folded inter-crystalline repeat. Although the spacing associated with this long period is not a direct measure of lamellar thickness (111), a long period spacing of 3.6 nm is commensurate with the calculated fold-to-fold distance of 2.8 nm expected for a lamellar structure such as the one illustrated in Figure 1.9.
Figure 3.17. X-ray diffraction patterns from oriented samples of poly(AG)3EG-I swollen in glycerol: (a) Wide angle and (b) low angle x-ray patterns taken using a Statton camera with the beam parallel to the plane of the stacked mats.
3.6.6 Thermal Treatment

Wide and low x-ray diffraction patterns from a crystal mat of poly(AG)3EG-I after thermal treatment are shown in Figure 3.18. These patterns were recorded with the beam parallel to the plane of a stack of mats. No change in spacing of the wide angle diffraction signals was observed. Examination of the wide angle region, specifically the signals indexed 200, 020, and 211, indicates that under the heating conditions used, the correlation lengths normal to the h00, 0k0 and 00l planes, as measured by the widths of the signals associated with these crystallographic planes, do not change. This observation is consistent with that inferred from an analysis of the infrared spectrum of this material after heat treatment. Although a detailed analysis has not been performed, no significant changes in the vibrational frequencies, line shapes and relative intensities occur upon heat treatment (Figure 3.19), indicating that no changes in the degree or quality of the hydrogen bonding network have occurred as a result of thermal treatment.

Although there are no observable changes in the wide angle pattern upon heat treatment, examination of the low angle region reveals that substantial changes have taken place (Figure 3.18b). In addition to the low angle signal associated with the long period in the sample prior to heat treatment, a new signal together with weaker orders has appeared at an angle corresponding to 6.6±0.1 nm in the heat treated sample. This signal is quite different in character from the original signal observed. It is sharper and more intense, indicating that the boundary between the crystalline and amorphous phases is more regular and that the density difference between these phases has increased. Although the origin of this low angle signal is not understood, a similar phenomenon has been reported in nyons 66, 610 and 612 (42).
Figure 3.18. X-ray diffraction patterns from thermally treated oriented samples of poly(AG)3EG-I: (a) Wide angle and (b) low angle x-ray patterns taken using a Statton camera with the beam parallel to the plane of the stacked mats.
Figure 3.19. Overlay of the infrared spectra of (1) thermally treated poly(AG)$_3$EG-I and (2) as matted poly(AG)$_3$EG-I.
3.7 Solid-state Structural Characterization of Proteins of the Sequence Poly(AG)$_x$EG-I

The expression systems developed for the preparation of these proteins, designated here as poly(AG)$_4$EG, poly(AG)$_5$EG, and poly(AG)$_6$EG, were obtained by the author from Yoshikuni Deguchi. The expression systems produce cleaved proteins with molecular weights of 19,114 daltons, 16,138 daltons and 13,038 daltons, respectively. The synthesis, cloning, expression and purification of these materials will be presented elsewhere (80).

3.7.1 Sample Preparation

Type I crystals were prepared by a method similar to that used to prepare poly(AG)$_3$EG-I. In the case of poly(AG)$_4$EG and poly(AG)$_5$EG, cyanogen bromide cleaved protein was dissolved in formic acid overnight at room temperature after which the solution was adjusted to 70% formic acid by the addition of water. Poly(AG)$_6$EG gelled upon dissolution in formic acid. The rate of gelation increased as the length of the alanylglucose segment increased, with gelation occurring in 5 days, 3 days and 1 hour, respectively. After gelation occurred, the resulting gels were washed with formic acid (70%) and methanol, then resuspended in methanol (40 mL) and aged at -10 °C for two days. In all cases, it was necessary to age the methanol suspension at -10 °C for at least two days to prevent flocculation of the suspension during the preparation of the sedimented crystal mat. Unoriented powders of these materials were prepared by centrifugation of the aged suspension to isolate the solid, followed by drying in vacuo overnight at room temperature. Crystal mats were prepared by a method analogous to that described for poly(AG)$_3$EG-I.
3.7.2 Infrared Spectroscopy

An overlay of the infrared spectra obtained from poly(AG)3EG-I, poly(AG)4EG-I, poly(AG)5EG, poly(AG)6EG-I is shown in Figure 3.20. Selected vibrational frequencies and their assignments are listed in Table 3.9. As in the analysis of poly(AG)3EG-I, the frequencies of the characteristic amide vibrations observed in all of the samples are those associated with the apβ-sheet conformation. The overall sharpening of the amide I vibrational mode as one progresses through the series indicates that the structure is becoming more well defined as the AG length is increased. The bands observed at 996 cm⁻¹ and 975 cm⁻¹ in poly(AG)3EG-I, assigned to skeletal modes associated with the regular alternation of glycyl and alanyl residues, sharpen upon progression through the series. These changes are consistent with what is expected as the length of the AG repeating segment is increased. Figure 3.21 shows an expansion of the 1800 to 900 cm⁻¹ region in the overlay. Examination of this figure shows that significant changes occur in the intensities and frequencies of the bands in this region upon progressing through the series. Close examination of the amide I region shows that the two weak higher frequency vibrations centered at 1710 cm⁻¹ and 1744 cm⁻¹ in poly(AG)3EG-I weaken further in AG4EG28-I and merge into one band centered at 1727 cm⁻¹ in poly(AG)5EG-I, shifting to 1723 cm⁻¹ in poly(AG)6EG-I. Although the decrease in the relative intensity of these bands is consistent with the decrease in the proportion of glutamic acid residues in the chemical sequence, the coalescence of the two bands observed in poly(AG)3EG and poly(AG)4EG to one band in poly(AG)5EG and poly(AG)6EG suggests that the side chain carboxyl groups in the latter have achieved a conformation where the strength of the hydrogen bond is intermediate to those observed in the former.
Figure 3.20. Infrared spectra of the protein series: (1) poly(AG)$_3$EG-I, (2) poly(AG)$_4$EG-I, (3) poly(AG)$_5$EG-I and (4) poly(AG)$_6$EG-I.
Table 3.9

Frequencies and assignments for selected vibrations observed in the infrared spectra of poly(AG)4EG-I, poly(AG)5EG-I, and poly(AG)6EG-I.

<table>
<thead>
<tr>
<th>Frequency (cm⁻¹)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG4</td>
<td>AG5</td>
<td>AG6</td>
</tr>
<tr>
<td>3566</td>
<td>3550</td>
<td>3560</td>
</tr>
<tr>
<td>3281</td>
<td>3279</td>
<td>3280</td>
</tr>
<tr>
<td>3064</td>
<td>3070</td>
<td>3068</td>
</tr>
<tr>
<td>1744</td>
<td>1727</td>
<td>1723</td>
</tr>
<tr>
<td>1711</td>
<td>1698</td>
<td>1696</td>
</tr>
<tr>
<td>1652</td>
<td>1650</td>
<td>1651</td>
</tr>
<tr>
<td>1621</td>
<td>1620</td>
<td>1621</td>
</tr>
<tr>
<td>1552</td>
<td>1556</td>
<td>1554</td>
</tr>
<tr>
<td>1519</td>
<td>1531</td>
<td>1525</td>
</tr>
<tr>
<td>1256</td>
<td>1263</td>
<td>1261</td>
</tr>
<tr>
<td>1229</td>
<td>1230</td>
<td>1229</td>
</tr>
<tr>
<td>995</td>
<td>998</td>
<td>998</td>
</tr>
<tr>
<td>973</td>
<td>973</td>
<td>973</td>
</tr>
</tbody>
</table>

a; assignment based on reference 95, b; assignment based on reference 96, c; assignment based on reference 50, d; assignment based on reference 94.
Figure 3.21. Expansion of the amide I, II and III regions in the infrared spectra of the protein series: (1) poly(AG)$_3$EG-I, (2) poly(AG)$_4$EG-I, (3) poly(AG)$_5$EG-I and (4) poly(AG)$_6$EG-I.
The stronger amide I component observed at 1652 cm\(^{-1}\) in poly(AG)\(_3\)EG-I and the weaker shoulder at 1665 cm\(^{-1}\) become more pronounced in poly(AG)\(_4\)EG-I, then weaken again in poly(AG)\(_5\)EG-I and poly(AG)\(_6\)EG-I. The continual presence of the band at 1652 cm\(^{-1}\) indicates that some fraction of the polypeptide chains have adopted a secondary structure different from the traditional \(\alpha\beta\)-sheet architecture in all samples. Although small changes in the frequencies of these bands occur upon progression through the series, they are still consistent with reverse turn structures of either the \(\beta\) or \(\gamma\)-type.

3.7.3 X-ray Diffraction

Wide angle x-ray diffraction patterns obtained from sedimented crystal mats of the protein series are shown in Figure 3.22. These patterns were recorded with the beam parallel to the plane of the mat. The wide angle diffraction signals observed in these samples index on orthorhombic unit cells. The measured and calculated interplanar spacings and the indexing details for poly(AG)\(_4\)EG-I, poly(AG)\(_5\)EG-I and poly(AG)\(_6\)EG-I are listed in Tables 3.10, 3.11 and 3.12. Although the level of orientation in these diffraction patterns is, in some cases, significantly less than in poly(AG)\(_3\)EG-I, the orientation direction still remains along \(a\) as observed in poly(AG)\(_3\)EG-I.

A detailed examination of these x-ray patterns indicates, as the length of the alanyl glycine segment increases, signals with indices of the type \(hk1\), sharpen considerably. Specifically, the signal indexed 211 is approximately twice as sharp in poly(AG)\(_4\)EG-I as in poly(AG)\(_3\)EG-I and the sharpness increases another factor of 1.5 from poly(AG)\(_4\)EG-I to poly(AG)\(_5\)EG-I. Upon increasing the stem length an additional AG dyad to poly(AG)\(_6\)EG-I, no increase in the sharpness of the 211 signal is observed. This observation indicates the correlation length along the \(c\) axis increases in
a monotonic fashion as the length of the segment is increased. The apparent increase in
the sharpness of the diffraction signal indexed 210, as one proceeds through the series,
is believed to be the result of better resolution between the 200 and 210 and not a
consequence of an increase in the correlation length along \( b \).

In addition to changes in the line width of signals indexed \( hkl \), changes in
interplanar spacing and relative intensity of signals indexed \( 0k0 \) occur upon progression
through the series. A decrease in the relative intensity 010 signal is observed as the
number of AG dyads is increased. This decrease in intensity is accompanied by a
overall reduction in the intersheet spacing of 10.2 %, with the greatest percentage
change (9.4 %) occurring between poly(AG)\(_3\)EG-I and poly(AG)\(_4\)EG-I. This result
correlates with the decreasing proportion of glutamic acid residues in the chemical
sequence and is in agreement with previously published work relating the effect of
amino acid composition on the intersheet spacing (49). These changes in the character
and spacing of the 010 signal are accompanied by an increase in the relative intensity
and a corresponding decrease in the interplanar spacing of the 020 signal.

Low angle x-ray diffraction patterns from the series are shown in Figure 3.22.
These patterns were taken with the beam parallel to the plane of a stack of mats and, in
the case of poly(AG)\(_4\)EG-I and poly(AG)\(_5\)EG-I, are composed of prominent arced
diffraction signals corresponding to spacings of 6.2 ± 0.1 nm and 5.5 ± 0.1 nm
respectively. The signals are focused on the equator together with a weaker and more
diffuse order. These low angle patterns resemble interparticle interference functions
similar to that observed in the low angle pattern of poly(AG)\(_3\)EG-I. The position and
character of these signals are indicative of stacks of lamellae, as observed in
poly(AG)\(_3\)EG-I, suggesting that these chains are also folded and that the chain axis is
oriented perpendicular to the lamellar surface.
Figure 3.22. Wide (upper case) and low angle (lower case) x-ray diffraction patterns from the protein series: (a) poly(AG)3EG-I, (b) poly(AG)4EG-I, (c) poly(AG)5EG-I and (d) poly(AG)6EG-I.
Table 3.10

Comparison of observed diffraction signal spacings ($d_0$) in poly(AG)$_4$EG-I with those calculated ($d_c$) for an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.028$ nm, $c = 0.695$ nm (errors ± 0.002 nm).

<table>
<thead>
<tr>
<th>$kl$</th>
<th>$d_0$</th>
<th>$d_c$</th>
<th>$kl$</th>
<th>$d_0$</th>
<th>$d_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0$</td>
<td></td>
<td></td>
<td>$h = 2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.028</td>
<td>1.028</td>
<td>00</td>
<td>0.474</td>
<td>0.474</td>
</tr>
<tr>
<td>20</td>
<td>0.514</td>
<td>0.514</td>
<td>10</td>
<td>0.439</td>
<td>0.430</td>
</tr>
<tr>
<td>21</td>
<td>0.445</td>
<td>0.413</td>
<td>11</td>
<td>0.370</td>
<td>0.366</td>
</tr>
<tr>
<td>30</td>
<td>0.360</td>
<td>0.363</td>
<td>$h = 3$</td>
<td>01</td>
<td>0.286</td>
</tr>
<tr>
<td>02</td>
<td>0.348</td>
<td></td>
<td>11</td>
<td></td>
<td>0.277</td>
</tr>
<tr>
<td>12</td>
<td>0.329</td>
<td></td>
<td>$h = 4$</td>
<td>00</td>
<td>0.238</td>
</tr>
<tr>
<td>31</td>
<td>0.304</td>
<td>0.307</td>
<td>10</td>
<td>0.230</td>
<td>0.231</td>
</tr>
<tr>
<td>22</td>
<td>0.288</td>
<td></td>
<td>20</td>
<td>0.211</td>
<td>0.215</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>0.257</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.11

Comparison of observed diffraction signal spacings ($d_0$) in poly(AG)$_5$EG-I with those calculated ($d_c$) for an orthorhombic unit cell with dimensions $a = 0.957$ nm, $b = 0.970$ nm, $c = 0.695$ nm (errors ± 0.002 nm).

<table>
<thead>
<tr>
<th>$kl$</th>
<th>$d_0$</th>
<th>$d_c$</th>
<th>$kl$</th>
<th>$d_0$</th>
<th>$d_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0$</td>
<td></td>
<td></td>
<td>$h = 2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.928</td>
<td>0.970</td>
<td></td>
<td>00</td>
<td>0.479</td>
</tr>
<tr>
<td>20</td>
<td>0.485</td>
<td>0.485</td>
<td></td>
<td>10</td>
<td>0.429</td>
</tr>
<tr>
<td>21</td>
<td>0.445</td>
<td>0.398</td>
<td></td>
<td>11</td>
<td>0.368</td>
</tr>
<tr>
<td>30</td>
<td>0.323</td>
<td></td>
<td>$h = 3$</td>
<td>01</td>
<td>0.286</td>
</tr>
<tr>
<td>02</td>
<td>0.348</td>
<td></td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.327</td>
<td></td>
<td>$h = 4$</td>
<td>00</td>
<td>0.241</td>
</tr>
<tr>
<td>31</td>
<td>0.291</td>
<td>0.293</td>
<td></td>
<td>10</td>
<td>0.229</td>
</tr>
<tr>
<td>22</td>
<td>0.282</td>
<td></td>
<td></td>
<td>20</td>
<td>0.211</td>
</tr>
<tr>
<td>40</td>
<td>0.242</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

113
Table 3.12

Comparison of observed diffraction signal spacings ($d_o$) in poly(AG)$_6$EG$_4$-I with those calculated ($d_c$) for an orthorhombic unit cell with dimensions

$a = 0.964$ nm, $b = 0.962$ nm, $c = 0.695$ nm (errors $\pm 0.002$ nm).

<table>
<thead>
<tr>
<th>kl</th>
<th>$d_o$</th>
<th>$d_c$</th>
<th>kl</th>
<th>$d_o$</th>
<th>$d_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0$</td>
<td></td>
<td></td>
<td>$h = 2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.943</td>
<td>0.962</td>
<td>00</td>
<td>0.482</td>
<td>0.482</td>
</tr>
<tr>
<td>20</td>
<td>0.481</td>
<td>0.481</td>
<td>10</td>
<td>0.436</td>
<td>0.431</td>
</tr>
<tr>
<td>21</td>
<td>0.445</td>
<td>0.396</td>
<td>11</td>
<td>0.373</td>
<td>0.366</td>
</tr>
<tr>
<td>30</td>
<td>0.321</td>
<td></td>
<td>01</td>
<td></td>
<td>0.292</td>
</tr>
<tr>
<td>02</td>
<td>0.348</td>
<td></td>
<td>11</td>
<td></td>
<td>0.279</td>
</tr>
<tr>
<td>12</td>
<td>0.327</td>
<td></td>
<td>00</td>
<td>0.241</td>
<td>0.241</td>
</tr>
<tr>
<td>31</td>
<td>0.301</td>
<td>0.291</td>
<td>10</td>
<td>0.229</td>
<td>0.234</td>
</tr>
<tr>
<td>22</td>
<td>0.282</td>
<td></td>
<td>20</td>
<td>0.211</td>
<td>0.215</td>
</tr>
<tr>
<td>40</td>
<td>0.243</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The values of these spacings, particularly in the case of poly(AG)₄EG-I, are considerably larger than the fold-to-fold distances (3.5 nm and 4.2 nm respectively) calculated for lamellae resulting from chemical sequence directed chain folding. However, it has been demonstrated that substantial discrepancies can exist between lamellar thicknesses as estimated by small angle x-ray scattering and other more direct techniques (112-114). As a consequence, the values determined by these measurements can only indicate the upper limit of the lamellar thickness, the actual value being smaller than measured by this technique.

The low angle signal observed in poly(AG)₆EG-I is quite different in character from the low angle signals observed in the other members of the series. This weak diffuse signal centered at 3.2 ± 0.1 nm is attributed to a Fraunhofer particle scattering profile of an electron dense protein slab. The spacing responsible for the signal observed can be calculated from equation 3.1 where d is the distance of the observed signal (115).

\[ D = \frac{3d}{2} \quad 3.1 \]

The maximum at 3.2 nm would be expected from a protein slab 4.8 nm thick in excellent agreement with the calculated fold-to-fold distance of 4.8 nm expected for chemical sequence directed chain folding.

3.8 Model Proposed for the Crystal Structures of Proteins of the Sequence Poly(AG)ₓEG-I

All of the experimental evidence collected on the solid state structure of poly(AG)₃EG-I and other members of the series strongly supports a crystalline αβ-sheet architecture involving a chain-folded lamellar structure as the basic crystalline unit.
This structure, designated model-I, is illustrated in Figure 3.23 and is similar to the cross-β structure proposed for Chrysopa silk (43) (Figure 1.4). The lamellar crystals illustrated in Figure 3.23 are constructed by the lateral stacking of apβ-sheets. The plane of the apβ-sheets is parallel to the longer dimension of the crystalline lamellae and perpendicular to the lamellar surface. The hydrogen bond direction of the sheet is parallel to the long dimension of the crystalline lamellae and the chain axis is normal to the lamellar surface.

The long period spacings of 3.6 nm, 6.2 nm, and 5.5 nm observed for poly(AG)3EG-I, poly(AG)4EG-I and poly(AG)5EG-I respectively, in addition to the slab thickness of 4.8 nm in the case of poly(AG)6EG-I, indicates for all the members of the series, that the lamellar thickness is shorter than the molecular length of the chains. Therefore, the chains must fold back at the lamellar surfaces and re-enter the crystalline lamellae, as generally accepted for flexible polymer chains (116). A direct consequence of the chains reversing their trajectory in this manner, is the antiparallel arrangement of adjacent polypeptide chains within the lamellar crystal as demanded by the spectroscopic and x-ray analysis of these materials. Although no direct measurements of the lamellar thickness have been carried out, the long period spacings observed in the low angle x-ray patterns, in combination with the selective line broadening observed in the wide angle x-ray patterns suggests, in all cases, that chain folding is predominantly in register with the chemical sequence periodicity as dictated by the necessity to position the glutamic acid residue at the fold surface. Relatively infrequent deviations from this periodicity to form interlamellar tie molecules are envisaged to occur. The folding is considered to be of the regular adjacent reentry type, with the occasional occurrence of irregular folding. These folds are sharp with a β-turn conformation (vida infra) as opposed to the other types of chain reversals observed in globular proteins (117). This fold geometry results in apβ-sheet surfaces equally decorated with alanine and glycine.
Figure 3.23. A schematic illustration of the model-I crystal structure.
side chain residues. Signal intensity is believed to be directed to the 010 planes through the pairing of adjacent sheets, perhaps as a consequence of hydrogen bonding between carboxyl groups in glutamic acid residues at the fold surface. As the length of the AG segment is increased the effect of this interaction on the pairing of sheets (essentially an end group effect) decreases, resulting in an overall decrease in the spacing and intensity of the 010 signal.

Such a model is compatible with the spectroscopic analysis of these materials. It also is consistent with the selective line broadening of the diffraction signals (the 211 signal in particular) in the wide angle pattern, and provides a rationale for the continuous sharpening of these diffraction signals with increasing AG segment length. This model also offers a straightforward explanation for the character and spacing of the low angle signals observed within the series and explains why a axis orientation is produced. A model such as the one proposed, has been determined to be stereochemically feasible from previous detailed analyses (43) and is consistent with current views on polymer crystal morphology (116).

3.9 Alternative Models

3.9.1 Variations in Folding in the Model-I Structure

In model-I, folding is considered to be periodic and predominantly in register with the chemical sequence periodicity. However, a model constructed from polypeptide chains that fold in a periodic fashion but out of register with the chemical sequence periodicity is certainly conceivable. If this occurred, glutamic acid residues would systematically decorate both pleated sheet surfaces. Stacking of these sheets would provide an opportunity for the side chain carboxyl groups of glutamic acid to form intersheet hydrogen bonds or interact with the main chain peptide groups as suggested
by the infrared analysis and offer an explanation for the observed decrease in the intersheet spacing as the relative AG content is increased within the series. However, the closest distance observed between two contiguous sheets in β2-polyglutamic acid is 0.78 nm (96); this is 0.25 nm larger than the value observed in poly(AG)3EG-I. One might anticipate the stacking of these sheets would lead to a substantial disruption of the sheet structure centered around glutamic acid. A local disruption of this type could explain the presence of the 1652 cm\(^{-1}\) band observed in the infrared analysis and the intensity and chemical shift of the 16.8 ppm shoulder observed in the CP/MAS \(^{13}\)C NMR analysis. However, an intersheet spacing larger than that observed and a greater reduction in this value upon progression through the series would be expected, since the concentration of glutamic acid residues on the sheet surface decreases by 50 % from poly(AG)3EG-I to poly(AG)6EG-I. In addition, the model does not offer an explanation for the occurrence of the 010 diffraction signal or the systematic increase in the long period spacings or the successive sharpening of the 211 diffraction signal as the AG content is increased.

A variant of this model is one that considers aperiodic folding, for example, perhaps the folding periodicity fluctuates between 6 and 10 residues. This would also produce apβ-sheets with both surfaces decorated with the side chains of glutamic acid residues. This model would offer no advantages over the parent model and again would not explain the successive sharpening of the 211 diffraction signal with increasing AG content since there would still be a broadening contribution from the short straight stems. It would also be expected to reduce the relative intensity of the observed low angle interference function.
3.9.2 Variations in Intersheet Stacking in the Model-I Structure

Crystals constructed in model-I are characterized by apβ-sheet surfaces equally decorated with alanine and glycine side chain residues. The stacking of these sheets in the $b$ direction would eliminate the 010 diffraction signal, directing all the intensity into the 020 second order at 0.53 nm. Therefore, the intensity and spacing of the 010 diffraction signal in poly(AG)$_3$EG-I and its subsequent decline upon progression through the series must be accounted for in any model proposed to describe the crystal structure of these materials. To account for the intensity of the 010 observed, model-I incorporates sheet pairing though intersheet hydrogen bond interactions between carboxyl groups present at the fold surface. The basis of this proposal is founded in the carbonyl stretching frequency observed for the side-chain carboxyl group in the glutamic acid residues. However, this is not the only route to introduce intensity into the 010 signal and maintain the essential characteristics of model-I. The scattering of the glutamic acid side group is determined by its 39 electrons, which may be compared to a sum of 34 electrons for the other 7 side groups in the repeating sequence. Thus, selective scattering by this group alone is capable of generating the observed intensity of the 010 diffraction signal. Since the shortest distance between glutamic acid residues on the surface is 0.53 nm between contiguous sheets, steric crowding of the glutamic acid residues decorating the surface is of some concern. One mechanism of relieving any possible congestion is to slip successive sheets by one corrugation, an integral distance of 0.695 nm in the $c$ direction, in an alternate fashion thus directing some intensity to the 010 planes. Interactions between the carboxyl groups and the main chain amide groups would provide an explanation for the 1600 cm$^{-1}$ band observed in the infrared analysis of these materials. The reduced influence of the bulky glutamic acid residues on sheet stacking, as the AG content is increased, could explain the decrease in the relative intensity and spacing of the 010 signal upon progression through the series. The
current experimental evidence does not allow the discrimination between this model and model-I; in fact, a combination of these two situations may exist.

3.9.3 Variations in Turn Conformation in the Model-I Structure.

So far, consideration has been given only to models in which chain folding occurs by utilizing turns constructed from an even number of amino acids. It is possible that the turns consist of an odd number of amino acids. The only turn of this type that can be seriously considered is a 3 amino acid turn, i.e. a γ-turn (118). In this model, folding would be regular and in register with the chemical sequence periodicity, again as directed by the necessity to position the glutamic acid residues at the apex of the turn. Incorporating such a turn into the crystal, regardless of the fold periodicity, shifts the relative amino acid registry between adjacent stems in the hydrogen bonded sheets by one amino acid. The result is that one surface of a hydrogen-bonded sheet is entirely decorated with glycine side chains and the other with alanine side chains. If successive sheets stack with like surfaces in contact, the crystallographic subcell (the orthorhombic unit cell) would be capable of generating intensity in the 010 planes, since alternate sheets would be shifted off the mean distance between two sheets. This model is similar in concept to that proposed for PLAG-I and shown in Figure 1.5. Hydrogen bonding interactions would be possible between carboxyl groups at the fold surfaces in a manner similar to that suggested in model-I, directing additional intensity into the 010 planes and providing a mechanism for the decline of 010 intensity upon progression through the series.

The primary difference between this model and model-I is the αβ-surface composition. As in model-I, this model is compatible with the spectroscopic analysis of these materials and provides a reasonable explanation for the observed intensity of the 010 signal. It also is consistent with the selective line broadening of the diffraction
signals in the wide angle pattern and provides a rationale for the sharpening of these
diffraction signals with increasing AG segment length. This model also offers a
straightforward explanation for the character and spacing of the low angle signals
observed within the series and explains why the a axis orientation is produced.

Unfortunately, the experimental data available on this series of polymers does
not allow discrimination between this model and model-I. However, studies on related
polypeptides favor chain folding through β-turns. In particular, sequence 1.1 (ie.
poly(AG)3PEG) would be capable of regularly folding by excluding proline from
within the stem segment though incorporation into the tripeptide fold. This would allow
the stem sequence to crystallize, forming the core of the chain-folded lamella. All
ttempts to crystallize this polymer in a chain folded structure, including methods similar
to the techniques used for poly(AG)3EG-I, have failed (75). Thus, if γ-bends were a
mechanism for regular folding of polypeptide chains we would have expected sequence
1.1 to have adopted a chain-folded structure, contrary to experiment.

3.10 Solid-state Structural Characterization of Poly(AG)3EG-II

3.10.1 Sample Preparation

The preparation of poly(AG)3EG-II was carried out using a standard procedure
developed for the preparation of PLAG-II (85). Powder samples of poly(AG)3EG-II
were prepared by dissolving poly(AG)3EG in aqueous lithium bromide and dialyzing
this solution against progressively diluted solutions of lithium bromide. A precipitate
developed at a lithium bromide concentration of 12 % but the dilution process was
continued until all the lithium bromide was removed. The precipitate was collected by
centrifugation, washed with methanol and dried overnight in vacuo at room temperature.
Only 32 % of the original sample weight was recovered in this process, the remaining
being precipitated by the addition of methanol. The sum total recovered using this procedure was approximately 87%. CP/MAS $^{13}\text{C}$ NMR and Raman spectroscopic analysis (not shown) indicated both precipitates were identical in structure; however, only the fraction recovered by dialysis against progressively diluted solutions of lithium bromide was characterized in detail.

3.10.2 Infrared Spectroscopy

An overlay of the infrared spectra obtained from poly(AG)$_3$EG-I and poly(AG)$_3$EG-II is shown in Figure 3.24. Selected vibrational frequencies and their assignments are tabulated in Table 3.13. Comparison of these spectra demonstrate that there are substantial differences in the frequencies and intensities of the bands observed in poly(AG)$_3$EG-I and poly(AG)$_3$EG-II. The frequencies of the amide A, amide I and amide II vibrational modes observed at 3289 cm$^{-1}$, 1652 cm$^{-1}$ and 1538 cm$^{-1}$ in poly(AG)$_3$EG-II indicate the hydrogen bonding present in this structure is weaker than that present in poly(AG)$_3$EG-I.

Figure 3.25 shows an expansion of the amide I and II region in the spectrum of poly(AG)$_3$EG-II. Close examination of the amide I region reveals the presence of five components. The pronounced shoulder at 1715 cm$^{-1}$ is assigned to carbonyl stretching of the carboxyl group in glutamic acid (95). The frequency of this vibration indicates that these carboxyl groups exist primarily in the dimerized form. The weak shoulder present at 1620 cm$^{-1}$ is indicative of $\beta$-sheet structures (54). The two bands at 1642 cm$^{-1}$ and 1696 cm$^{-1}$ suggest that the chain arrangement within these sheets is a mixture of parallel and antiparallel (54). The weak splitting of the amide II band into two bands at 1538 cm$^{-1}$ and 1550 cm$^{-1}$ confirms the assignment of the 1642 cm$^{-1}$ band and the weak shoulder at 1521 cm$^{-1}$ confirms the assignment of the 1696 cm$^{-1}$ band (54). Alternatively, the band at 1646 cm$^{-1}$ can be attributed to a helical structure; consistent
with this assignment the band at 1550 cm\(^{-1}\) would be attributed to the amide II component of that helical structure (97). However, an additional amide II band characteristic of helical chains is not observed at 1515 cm\(^{-1}\), suggesting that the former assignment is more reasonable. It is interesting to note that the main component of the amide I vibrational mode in poly(AG)\(_3\)EG-II is observed at 1652 cm\(^{-1}\). The frequency of this peak is lower than the amide I vibrational mode observed in the infrared spectrum of PLAG-II or silk-I (119). A peak of identical frequency but lower relative intensity is

![Figure 3.24. Overlay of the infrared spectrum of (1) poly(AG)\(_3\)EG-I and (2) poly(AG)\(_3\)EG-II.](image-url)
Table 3.13

Frequencies and assignments for selected vibrations observed in the infrared spectrum of poly(AG)3EG-II.

<table>
<thead>
<tr>
<th>Frequency (cm(^{-1}))</th>
<th>Assignment</th>
<th>Comments(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3550</td>
<td>Glutamic acid</td>
<td>O-H Stretch</td>
</tr>
<tr>
<td>3289</td>
<td>Amide A</td>
<td></td>
</tr>
<tr>
<td>3071</td>
<td>Amide B</td>
<td></td>
</tr>
<tr>
<td>1715</td>
<td>Glutamic acid</td>
<td>C=O Stretch</td>
</tr>
<tr>
<td>1652</td>
<td>Amide I</td>
<td></td>
</tr>
<tr>
<td>1538</td>
<td>Amide II</td>
<td></td>
</tr>
<tr>
<td>1238</td>
<td>Amide III</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) assignment based on previous analysis of poly(AG)3EG-I.
Figure 3.25. Overlay of the amide I and II region of the infrared spectrum of (1) poly(AG)3EG-I and (2) poly(AG)3EG-II.
observed in poly(AG)3EG-I. In the infrared analysis of poly(AG)3EG-I, it was suggested that this peak was a consequence of turn structures. If that assignment is correct, this crystalline modification is highly folded in comparison to the modification observed in poly(AG)3EG-I.

3.10.3 Raman Spectroscopy

An overlay of the Raman spectra obtained from poly(AG)3EG-I and poly(AG)3EG-II is shown in Figure 3.26. Selected vibrational frequencies and their assignments are listed in Table 3.14. As in the infrared analysis, comparison of these spectra demonstrate that there are substantial differences in the frequencies and intensities of the bands observed in poly(AG)3EG-I and poly(AG)3EG-II. Figure 3.27 shows an expansion of the amide I and II region in the spectrum of poly(AG)3EG-II. Close examination of the amide I region reveals the presence of five components. The weak shoulder at 1700 cm\(^{-1}\) is assigned to carbonyl stretching of the carboxyl group in glutamic acid (98). The amide I component at 1670 cm\(^{-1}\) is generally in the range assigned to proteins adopting an \(\alpha\beta\)-sheet architecture (120). A peak of similar frequency is observed in polyglycine-I (121). The amide III component observed at 1245 cm\(^{-1}\) is consistent with this assignment (120). The frequency of the strong band at 1275 cm\(^{-1}\) is too high to result from the expected splitting observed for \(\beta\)-sheet structures. The frequency of this band suggests that it results from the amide III vibrational mode of a helical conformation or a turn structure (64). The absence of a strong band at 900 cm\(^{-1}\), characteristic of helical conformers, indicates that the 1275 cm\(^{-1}\) vibration is more reasonably assigned to turn structures. The weak shoulder at 1688 cm\(^{-1}\) reinforces this assignment. The main amide I band at 1660 cm\(^{-1}\) suggests
Figure 3.26. Overlay of the Raman spectrum of (1) poly(AG)3EG-II and (2) poly(AG)3EG-I.
Table 3.14

Frequencies and assignments for selected vibrations observed in the Raman spectrum of poly(AG)3EG-II.

<table>
<thead>
<tr>
<th>Frequency (cm(^{-1}))</th>
<th>Assignment(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3285</td>
<td>Amide A</td>
</tr>
<tr>
<td>1660</td>
<td>Amide I</td>
</tr>
<tr>
<td>1275</td>
<td>Amide III</td>
</tr>
<tr>
<td>1245</td>
<td>Amide III</td>
</tr>
</tbody>
</table>

\(^a\) assignment based on previous analysis of poly(AG)3EG-I.
Figure 3.27. Expansion of the amide I and II region of the Raman spectrum of poly(AG)3EG-II.
that a substantial fraction of this material is disordered (120), and it is possible that the amide III band at 1275 cm\(^{-1}\) results from this fraction. The band observed as a weak shoulder at 1637 cm\(^{-1}\) could not be assigned.

3.10.4 CP/MAS Solid-state \(^{13}\)C NMR Spectroscopy

The CP/MAS \(^{13}\)C NMR spectrum of poly(AG)\(_3\)EG-II is shown in Figure 3.28. The observed chemical shifts and their assignments for poly(AG)\(_3\)EG-II and PLAG II (70) are summarized in Table 3.15.

![CP/MAS NMR spectrum of poly(AG)\(_3\)EG-II](image)

**Figure 3.28.** CP/MAS \(^{13}\)C NMR spectrum of poly(AG)\(_3\)EG-II. ssb; spinning side band.
The spectrum shows the \( \alpha \) and \( \beta \)-carbons of alanine centered at 50.9 ppm and 16.8 ppm respectively. The peak assigned to the \( \beta \)-carbon shows no evidence of a shoulder at 20.7 ppm indicating, within the resolution of this experiment, that poly(AG)\(_3\)EG-II is not a mixture of type I and II crystal structures. The peak at 43.5 ppm is assigned to \( \alpha \)-carbon of glycine. As in the case of poly(AG)\(_3\)EG-I, the spectral resolution does not allow for the specific identification of the \( \beta \) and \( \gamma \)-carbons in glutamic acid; the broad peak centered at ca. 30 ppm has been assigned to those carbons (99). The carbonyl, amide and \( \alpha \)-carbons in glutamic acid could not be assigned. The peaks centered at 171.1 ppm and 176.0 ppm are assigned to the amide carbons of glycine and alanine, respectively. The chemical shifts for all of these carbons are consistent with a conformation isomorphous with the PLAG-II crystalline modification (70).

Table 3.15

<table>
<thead>
<tr>
<th>Chemical Shift (ppm)</th>
<th>Assignment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>poly(AG)(_3)EG-II</td>
<td>silk-I</td>
<td>PLAG-II</td>
</tr>
<tr>
<td>51.0</td>
<td>50.3</td>
<td>50.5</td>
</tr>
<tr>
<td>16.8</td>
<td>16.6</td>
<td>16.6</td>
</tr>
<tr>
<td>176.2</td>
<td>177.5</td>
<td>177.1</td>
</tr>
<tr>
<td>43.5</td>
<td>43.9</td>
<td>43.7</td>
</tr>
<tr>
<td>171.1</td>
<td>171.1</td>
<td>171.9</td>
</tr>
</tbody>
</table>

Chemical shifts and assignments for peaks observed in the CP/MAS \(^{13}\)C NMR spectra of poly(AG)\(_3\)EG-II, PLAG-II (70) and silk-I (70).
3.10.5 X-ray Diffraction

Wide angle diffraction patterns of poly(AG)3EG-II were obtained as powders. The preparation of oriented samples of poly(AG)3EG-II was never attempted. The powder diffraction patterns of poly(AG)3EG-II exhibit discrete Bragg diffraction signals consistent with crystalline material. A representative example is shown in Figure 3.29. A wide angle diffraction pattern obtained from a powder sample of poly(AG)3EG-I is included for comparison. Qualitative examination of Figure 3.32a reveals that there are differences between the powder patterns of poly(AG)3EG-I and poly(AG)3EG-II indicating the structure of poly(AG)3EG-II is different from poly(AG)3EG-I. However, a comparison of the diffraction signals and their measured spacings suggests that these differences may be rationalized on the basis of a chain axis contraction, similar to that proposed by Lotz and coworkers for PLAG-II (67). On the basis of this argument, the diffraction pattern of poly(AG)3EG-II was indexed on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.734$ nm and $c$ (chain axis) = 0.940 nm (errors $\pm 0.002$ nm). The measured and calculated interplanar spacings and indexing details are listed in Table 3.16. The calculated and measured spacings show excellent agreement. The value of the 0.948 nm spacing was assigned on the basis of its second diffraction order at 0.474 nm. This diffraction signal is also observed in poly(AG)3EG-I and, in both cases, is attributed to the distance between hydrogen-bonded chains. The value of 1.734 nm for $b$, or more specifically its second diffraction order observed at 0.87 nm, represents the average intersheet stacking periodicity. This value is 21 % greater than the value reported for PLAG II (0.72 nm) (67). The same percent difference in the spacing is observed for the diffraction signal indexed 020 in poly(AG)3EG-I and PLAG-I.
Figure 3.29. Wide angle x-ray diffraction patterns from powders of (a) poly(AG)3EG-II and (b) poly(AG)3EG-I taken using a Statton camera.
Table 3.16

Comparison of the observed diffraction signal spacings ($d_0$) in poly(AG)$_3$EG-II with those calculated ($d_c$) for an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.734$ nm, $c = 0.940$ nm (errors $\pm 0.002$ nm).

<table>
<thead>
<tr>
<th>kl</th>
<th>$d_0$</th>
<th>$d_c$</th>
<th>kl</th>
<th>$d_0$</th>
<th>$d_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0$</td>
<td>10</td>
<td>1.734</td>
<td>1.734</td>
<td>$h = 3$</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.867</td>
<td>0.867</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>$h = 2$</td>
<td>00</td>
<td>0.474</td>
<td>0.474</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.461</td>
<td>0.457</td>
<td>$h = 4$</td>
<td>00</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.411</td>
<td>0.411</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.226</td>
<td>0.227</td>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>
Therefore, based on the analysis of poly(AG)3EG-I, the spacing observed for poly(AG)3EG-II is reasonable. The direct determination of the 001 value was not possible. This diffraction signal and all of its higher orders are obstructed by other hkl diffraction signals observed in the powder patterns of poly(AG)3EG-II. As a result, the value of 0.940 nm was determined from the diffraction signal indexed 211. This diffraction signal, although shifted to lower angle, is similar in character to the diffraction signal indexed equivalently in poly(AG)3EG-I, suggesting that the coherence length in the c dimension is small compared to the other crystallographic dimensions. Although the positions of the observed diffraction signals in poly(AG)3EG-II make it difficult to establish a hierarchy of coherent scattering lengths in the a, b and c directions, it is assumed that they parallel the order observed in poly(AG)3EG-I. No low angle diffraction signals were observed in any poly(AG)3EG-II samples prepared.

3.11 Model Proposed for the Crystal Structure of Poly(AG)3EG-II

Based on the previous detailed analysis of the solid state structure of poly(AG)3EG-I, the experimental evidence collected on the solid state structure of poly(AG)3EG-II supports a crystalline sheet architecture involving a chain folded lamellar structure as the basic crystalline unit. The structure envisaged is constructed by contracting the traditional ß-sheet structure as proposed for poly(AG)3EG-I and other members of that series in a manner similar to that proposed by Lotz and coworkers for the structure of PLAG-II (67). This contraction is accomplished by alternative amino acid residues adopting conformations such that successive peptide groups are aligned approximately parallel and perpendicular to the chain axis. Lotz and coworkers refer to this chain arrangement as a "crankshaft" conformation. Such an arrangement results in a peptide backbone that is highly contracted such that the rise per amino acid residue is 0.24 nm in poly(AG)3EG-II compared to 0.35 nm in poly(AG)3EG-I. This c-axis
contraction (identical to that reported by Lotz and coworkers for PLAG-II) leads to an expansion in the $b$ dimension, resulting in the increased intersheet spacing observed in poly(AG)$_3$EG-II. Lamellar crystals that form the basis of this crystalline entity are constructed from the lateral stacking of sheets as in poly(AG)$_3$EG-I. As in the model proposed by Lotz and coworkers, there are four amino acid residues per chain axis repeat related by a 2-fold screw axis. The alanine residues adopt dihedral angles that approximate an extended $\beta$-type conformation and glycine residues are configured as left handed $\alpha$-helices. As a result of the absence of any low angle signals in the x-ray diffraction patterns of poly(AG)$_3$EG-II, no statements regarding the specific details of chain folding can be made.

Although, as mention previously, this construction results in a net helical twisting of the peptide backbone that seemingly precludes any possible transitions between the poly(AG)$_3$EG-I and poly(AG)$_3$EG-II structures without complete disruption of the hydrogen bonding sheets, the similarity between the unit cell dimensions of PLAG-II (67) and poly(AG)$_3$EG-II suggests that these structures must be closely related.

3.12 Effect of Chemical Sequence on the Relative Stability of the Poly(AG)$_3$EG-II Crystal Structure

Regardless of the model one uses to describe the specific details of the crystal structures of poly(AG)$_3$EG-II, PLAG-II and silk-I, the similarities between the x-ray diffraction patterns and infrared spectra obtained from these materials indicate their structures are very similar. It is interesting to note the different physical properties that each of these materials exhibit. As mentioned earlier, the crystalline component of $B.\;mori$ silk fibroin can be characterized primarily by the repeat Gly-X where X is Ala or Ser in a ratio of two to one, respectively. Silk-I is obtained when the contents of the
silk gland are dried without mechanical disturbance or when the chymotrypsin digested soluble fraction is dialyzed from an aqueous solution of LiBr (45). This material is water soluble. The presence of alcohols in, or mechanical shearing of, aqueous solutions of silk-I induces the formation of silk-II (45). PLAG-II can also be obtained by dialysis from aqueous solutions of LiBr. However, poly(AG)3EG-II can be isolated by dialysis from aqueous solutions of LiBr only under acidic conditions and the presence of alcohols in the solution does not induce the formation of poly(AG)3EG-I. PLAG-II is completely insoluble in water (45) and poly(AG)3EG-II can only be dissolved in water under basic conditions. Only two routes have been successfully used to induce the formation of poly(AG)3EG-I from poly(AG)3EG-II. These are mechanical shearing of solutions of poly(AG)3EG-I in aqueous formic acid and vapor diffusion of methanol into solutions of poly(AG)3EG-I in aqueous formic acid (water content less than 10%).

These differences in physical properties would have to be attributed to the chemical sequence of the respective polypeptides. The primary sequence of poly(AG)3EG is strongly biased toward acidic residues, glutamic acid composing 22 weight percent of the amino acid composition. In the case of B. mori silk, glycine, serine and alanine constitute approximately 85 weight percent of the amino acid composition, with the remaining 15 weight percent being composed of the other 17 amino acids (107). Therefore, B. mori silk contains a relatively small number of ionizable amino acids and PLAG contains none. The pKa of glutamic acid is 4.6; above this value a substantial portion of the glutamic acid residues are ionized in aqueous solution and thus highly solvated. The pH sensitivity and increased relative stability of poly(AG)3EG-II compared to that of silk-I and PLAG-II may be a consequence of this high degree of solvation, indicating that solvation plays a critical role in the formation and stabilization of these crystal structures.
4.1 Conclusions

The application of molecular genetics as a synthetic tool is a powerful new strategy for the preparation of polymeric materials. As discussed earlier, previous workers have shown that this approach provides a level of control over molecular structure that results in materials with a homogeneity of chemical structure unparalleled by the traditional methods for the preparation of polymeric materials (30, 75, 89). The specific objective of this research was to explore the feasibility of this approach to exercise chemical sequence control over solid state structure and prepare chain folded lamellar crystals of predetermined thickness and surface functionality. This objective was pursued by utilizing biomolecular synthesis to achieve absolute control over the selection and placement of individual monomers within polymer sequence 1.4.

Solid state structural characterization of poly(AG)3EG demonstrates that this material adopts two crystal modifications that are isomorphic with those observed in PLAG and B. mori silk fibroin. Infrared, Raman, and CP/MAS $^{13}$C NMR spectroscopic analyses of poly(AG)3EG-I show characteristic features of apβ-sheets. Wide angle x-ray diffraction experiments demonstrate the crystalline nature of this material, support the conclusion of the antiparallel arrangement of chains within the β-sheet structure, and indicate that the crystals are oriented with cylindrical symmetry about the $a$ axis. These crystals index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.060$ nm, and $c$ (chain axis) = 0.695 nm (errors ±0.002 nm). Selective line broadening of the wide angle diffraction signals with $l = 1$ argues for a
coherence length of approximately 2-4 nm in the c direction. Low angle x-ray diffraction measurements show that regular fluctuations in electron density, perpendicular to a, occur at approximately 3 nm which are consistent with chainfolded lamellae. The results are commensurate with a model constructed from stacked, regularly chain-folded crystalline lamellae composed of polypeptide chains that reverse polarity in register with the sequence periodicity of poly(AG)3EG-I.

Solid state structural analysis on other members in the series show similar crystalline architectures and support a model in which chain folding is directed by the chemical sequence periodicity. Although experimental evidence collected on poly(AG)3EG-I and other members in the series does not provide detailed information regarding the turn geometry, evidence collected on similar systems favors a model that exploits β-bends to reverse chain polarity (75).

Comparison of the infrared and Raman spectra of poly(AG)3EG-II and poly(AG)3EG-I shows that substantial differences exist in the frequencies and intensities of the observed bands, suggesting these materials have different crystal structures. The CP/MAS $^{13}$C NMR spectrum of poly(AG)3EG-I reinforces this conclusion and provides evidence that the solid state structure of poly(AG)3EG-II is isomorphic with PLAG-II and Silk-I. Wide angle x-ray diffraction patterns from poly(AG)3EG-II exhibit discrete Bragg diffraction signals that index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.734$ nm, and $c$ (chain axis) = 0.940 nm (errors ±0.002 nm). These unit cell dimensions are consistent with those observed in similar systems (67). Based on the detailed structural analysis of poly(AG)3EG-I, the experimental evidence collected on poly(AG)3EG-II supports a crystalline structure composed of chain folded lamellae constructed from the lateral stacking of hydrogen bonded sheets, in which, the peptide backbone is highly contracted. In this model, the proposed contraction is a result of the glycine residues adopting a left handed α-helical conformations. As a result of the absence of any low
angle signals in the x-ray diffraction patterns of poly(AG)3EG-II, no statements regarding the specific details of chain folding can be made.

In conclusion, the integration of molecular genetics into the field of material science is a novel approach for the preparation of new materials. Exploitation of this approach provides a level of control over molecular structure that results in materials with a homogeneity of chemical structure unparalleled by the traditional methods for the preparation of polymeric materials. By utilizing this synthetic method, one can achieve absolute control over the selection and placement of individual monomers within a polymer sequence and begin to exert chemical sequence control over solid state molecular structure. At present, genetic engineering is the only approach for the preparation of polymeric materials available that allows such a high degree of control. Further exploitation of this synthetic method for the preparation of novel materials and the study of their properties will most certainly lead to new insights regarding macromolecular behavior eventually permitting molecular design of new materials. In addition, these new materials will allow one to abandon the bulk processing technologies presently used for a new technology which exploits molecular self-assembly to build complex structures.

4.2 Future Work

It is the opinion of the author that future work on these materials should proceed along in two paths: (i) additional characterization of the specific structures prepared and discussed in this work and (ii) chemical modification of these polymers to prepare new materials.

Additional experiments to solidify the foundation of the spectroscopic analysis presented in this work include, quantification of the amide vibrational modes observed by infrared and Raman spectroscopy, specifically the amide III bands. In particular,
gelation of poly(AG)₃EG in deuterated aqueous formic acid and infrared and Raman spectroscopic analysis of the resulting material would provide a means for decoupling the pure amide modes from the mixed vibrational bands observed experimentally. Specific identification and quantification of these modes would be useful in confirming the assignments made in this work.

Detailed molecular modeling utilizing the Silicon Graphics work station equipped with Biodesign and Cerius software can be used to construct the proposed crystal structures and simulate their x-ray diffraction patterns, providing additional information regarding the feasibility of the structure proposed for poly(AG)₃EG-II and possibly allowing discrimination between the possible turn geometries suggested in the discussion of the structure of poly(AG)₃EG-I.

Transmission electron microscopy and diffraction of both crystalline modifications observed in poly(AG)₃EG can provide additional information regarding crystal morphology and single crystal order within the microcrystalline solid. Samples of the suspensions used to prepare the crystal mats can be deposited onto carbon coated microscope grids and viewed in the electron microscope under conditions of transmission and diffraction. The application of microspheres with known dimensions followed by shadowing of the samples with a Pt/Pd alloy will allow for an independent estimation of the crystal thickness that can be compared with that determined by x-ray diffraction.

As mentioned earlier the structure proposed for poly(AG)₃EG-II results in a net helical twisting of the peptide backbone that seemingly precludes any possible transitions between the poly(AG)₃EG-I and poly(AG)₃EG-II structures without complete disruption of the hydrogen bonding sheets. An alternative model that merits consideration is one in which successive glycine residues alternate between left and right handed helical conformations. This idea is essentially a composite of the models presented by Lotz and coworkers (67) and Fossey and coworkers (71). This
construction eliminates any net helical twisting of the peptide backbone. Conversion of this structure to the traditional αβ-sheet structure would occur by the gentle disruption of hydrogen bonds within the sheet, in such a way that chain extension can occur without requiring catastrophic disruption of the sheet.

The materials prepared in conjunction with this work can be derivatized by reaction at the carboxylic acid functional group present in the glutamic acid residues. These modifications can be used to incorporate groups that will provide additional driving forces for crystallization to occur. For example, mesogenic groups that favor the formation of intrachain liquid crystalline phases can be used to stabilize the chain folded structure under conditions that would normally disrupt it. Alternatively, one can exploit the chemical sequence periodicity of these materials and incorporate species with the ability for molecular recognition and design self-assembling systems. In the future, it is the chemical modification these materials, where the polypeptide backbone serves only to organize and create the appropriate spatial arrangements between the sites of interest, that will lead to the most interesting and useful materials.
ENDNOTES


44. Fraser, R. D. B.; MacRae, T. P. *Conformations in Fibrous Proteins*; Academic: 1973.


45. Lotz, B.; Colonna-Cesari, F. Biochimie 1979, 61, 205.


60. Frushour, B. G.; Koenig, J. L. Biopolymers 1974, 13, 455.


90. Auroy, P. CEA, Laboratoire Leon Brillouin, CEN Saclay, France: information provided by experiments performed as a visiting scientist, 1993.
91. Deguchi, Y. Central Research Laboratories, Kaneka Corporation, Kobe
Japan: information provided by experiments performed as a visiting scientist,

1015.


1958, 59, 322.

95. Conley, R. T. Infrared Spectroscopy, 2 ed.; Allyn and Bacon: Boston,
1966, p 144.


1729.


100. Ishida, M.; Asakura, T.; Yokoi, M.; Saito, H. Macromolecules 1990, 23,
88.


BIBLIOGRAPHY


Auropy, P. CEA, Laboratoire Leon Brillouin, CEN Saclay, France: information provided by experiments performed as a visiting scientist, 1993.


Brack, A.; Spach, G. Biopolymers 1972, 11, 563.


Deguchi, Y. Central Research Laboratories, Kaneka Corporation, Kobe Japan: information provided by experiments performed as a visiting scientist, 1992

Deguchi, Y.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. manuscript in preparation.


Fraser, R. D. B.; MacRae, T. P. *Conformations in Fibrous Proteins*; Academic: 1973.


Goldberg, I; Salerno, A. J.; Paterson, T.; Williams, J. I. Gene 1989, 80, 305.


Lotz, B.; Colonna-Cesari, F. Biochimie 1979, 61, 205.


