# Poly(L-alanylglycine): Multigram-Scale Biosynthesis, Crystallization, and Structural Analysis of Chain-Folded Lamellae

**Alyssa Panitch,[†] Kunio Matsuki,[†] Eric J. Cantor,[‡] Sharon J. Cooper,[§] Edward D. T. Atkins,[§] Maurille J. Fournier,[‡] Thomas L. Mason,[‡] and David A. Tirrell*,[†]**

*Department of Polymer Science and Engineering and Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Massachusetts 01003, and H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, U.K.*

*Received July 18, 1996; Revised Manuscript Received October 17, 1996*[⊗]

ABSTRACT: The biosynthesis of poly(L-alanylglycine) (poly(AG)) was performed in high cell density cultures of recombinant *Escherichia coli*. The purity of the material was determined by amino acid analysis, elemental analysis, and $^1$H NMR spectroscopy. Fed batch fermentation increased the yield of recombinant protein from levels of tens of milligrams per liter (typical of batch fermentation in rich media) to hundreds of milligrams per liter. Poly(AG) comprising 64 diads [$(AG)_{64}$] was recrystallized from dichloroacetic acid solutions in the form of texture-oriented chain-folded lamellae with a lamellar stack periodicity of 3.2 nm. The crystal structure within the lamellar core is similar in general, but different in detail, to the antiparallel $\beta$-sheet structure previously reported for oriented films of poly(AG) and fibers of *Bombyx mori* silk fibroin (silk II). The structure consists of polar antiparallel (ap) $\beta$-sheets, with repetitive folding through $\gamma$-turns every eighth amino acid (including the fold), stacking with like surfaces together. The wide-angle X-ray diffraction signals index on an orthorhombic unit cell with *a* (hydrogen bond direction) = 0.948 nm, *b* (sheet stacking direction) = 0.922 nm, and *c* (chain direction) = 0.695 nm. The stacking distance (*b*-value) is increased by about 3% in comparison with the previously reported structure of poly(AG), owing, we believe, to steric interaction at the lamellar fold surfaces. Random shears of approximately $\pm a/4$ and shears of $\pm c/2$ in the *ac* plane are required to obtain a good fit between the calculated and measured X-ray structure factors.

## Introduction

Recombinant DNA technology is a powerful tool for synthesis of novel proteins for pharmaceutical and biomedical applications[1,2] as well as for research in structural biology and protein folding. We and others have previously reported[3−13] the biosynthesis of highly repetitive, monodisperse, silk-like proteins, including derivatives containing non-proteinogenic amino acids[11] or elastomeric domains.[12,13] We describe herein the preparation, on a multigram scale, of two monodisperse variants of the "parent" polymer in this family of materials, poly(L-alanylglycine) (poly(AG)). These polymers can be crystallized to form silk II-like $\beta$-sheet structures as shown by Fourier transform infrared spectroscopy (FTIR), small-angle X-ray scattering (SAXS), and wide-angle X-ray scattering (WAXS).

Poly(AG) (as prepared by chemical synthesis) has been studied previously.[14a,b,15] Fraser et al. prepared poly(AG) by polymerization of an L-alanylglycine *p*-nitrophenyl ester monomer, resulting in a degree of polymerization of about 40.[14a] Use of the pentachlorophenyl ester has been reported to yield molecular weights in the range of 15 000.[15] Infrared and X-ray scattering data indicate that such polymers form either the silk I (crankshaft geometry) or the silk II ($\beta$-sheet geometry) structures depending on sample preparation conditions.[15]

We have reported previously the use of genetic engineering to produce artificial proteins that adopt predictable, chain-folded antiparallel (ap) $\beta$-sheet structures in the solid state.[3−7] The chain-folded lamellar architecture[16−19] is a kinetic trap in the crystallization of all linear polymers sufficiently flexible to form hairpin-like folds, and it is known that protein chains in the $\beta$-conformation can make tight turns, with either one or two amide groups (two or three amino acid $C_\alpha$-atoms) in the fold. The egg-stalk silk protein of the green-wing fly *Chrysopa flava* folds with a periodicity corresponding to eight amino acids to form lamellae of thickness 2−3 nm,[20] and nylons 4[21] and 4,6[22] crystallize with amide units, rather than polymethylene segments, in the turn. The biosynthesis of multigram quantities of poly(AG) with known chain lengths well in excess of the expected fold periodicity allowed us to search for chain folding in poly(AG) and to compare the folded-chain structure with that previously reported by Fraser et al.[14a,b]

The yield of a recombinant protein in bacterial fermentations is determined in part by the total number of cells in the culture. Traditional batch reactors yield approximately 1.5 g of dry cell weight per liter. This low ratio of biomass to culture volume leads to low efficiency in terms of space, time, and cost. Fed batch processes yield significantly greater cell densities; values up to approximately 100 g of dry cell weight per liter have been reported.[23]

We describe in this paper simple fed batch techniques with the addition of oxygen that have increased the yield of poly(AG) from tens of milligrams per liter (using batch techniques) to hundreds of milligrams per liter. We also discuss the structural analysis of poly(AG) by infrared and X-ray diffraction methods.

[†] Department of Polymer Science and Engineering, University of Massachusetts.
[‡] Department of Biochemistry and Molecular Biology, University of Massachusetts.
[§] H. H. Wills Physics Laboratory, University of Bristol.
[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1996.

S0024-9297(96)01059-5 CCC: $14.00 © 1997 American Chemical Society

## Experimental Section

**Materials.** T4 DNA ligase, T4 polynucleotide kinase, calf intestinal phosphatase, and all restriction enzymes were purchased from New England Biolabs. Ribonuclease A, deoxyribonuclease I, Organic Antifoam, ampicillin, and chloramphenicol were purchased from Sigma Chemical Co. *Escherichia coli* strains DH5αF′ and HB101 were purchased from Bethesda Research Laboratories. Plasmid p937.51 was a gift of Protein Polymer Technologies, Inc. The expression plasmid pET-3b and *E. coli* strain BL21(DE3)pLysS[24] were obtained from Novagen.

**General Methods.** Bacterial growth in rich medium, DNA manipulations, and transformation conditions were performed as described as described in Sambrook, Frisch, and Maniatis.[25] Fourier transform infrared spectra were obtained from KBr pellets on a Perkin-Elmer 1600 series FTIR spectrophotometer. $^1$H NMR spectra were run on a Bruker AMX 500 MHz spectrometer. Ultraviolet spectra were obtained in distilled water on a Hitachi U2000 spectrophotometer in quartz cuvettes with a path length of 1 cm, while cell density measurements were done in plastic cuvettes at $\lambda = 600$ nm. High cell density medium for fermentations was adapted from Riesenberg et al.[26] with the addition of 100% $O_2$ once an optical density ($OD_{600}$) of 10 was reached. A New Brunswick 80L Mobile Pilot Plant (MPP-80) was used for high cell density fermentation. Centrifugation was performed using a Cepa Z41 continuous centrifuge. Amino acid analyses were obtained from an Applied Biosystems 420/130A derivatizer−analyzer following hydrolysis and derivatization with phenyl isothiocyanate; data were analyzed with the Applied Biosystems 920A data module.

**Preparation of Synthetic DNA.** Oligonucleotides (**1**) were synthesized on a 1 μmol scale on a Biosearch Model 8700 DNA synthesizer using phosphoramidite chemistry.[27] Crude oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis and eluted using the "crush and soak" method.[25] Purified oligonucleotides were annealed at 80 °C and allowed to cool to room temperature over several hours. The annealed duplex (**1**) was phosphorylated with T4 polynucleotide kinase, ethanol precipitated, and dried *in vacuo*.

```
      Stop Gly  Ala  Gly  Ala  Gly  Ala  Gly  Ala  Gly
AAT TCG TAA GGT GCC GGT GCG GGT GCA GGT GCG GGT GCC G        1
    GC ATT CCA CGG CCA CGC CCA CGT CCA CGC CCA CGG CCTAG
Eco RI      Ban I                          Ban I    Bam HI
```

**Cloning and Amplification of Synthetic DNA.** The purified synthetic duplex was ligated into *Eco*RI/*Bam*HI digested pUC18[28] and used to transform *E. coli* strain DH5αF′. Cells were grown at 37 °C overnight on 2xYT solid medium containing 200 μg/mL ampicillin, 25 μg/mL β-isopropyl thiogalactoside (IPTG), and 40 μg/mL of the chromogenic substrate 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) for blue−white screening. Plasmid DNA from white transformants was sequenced using Sequenase 2.0 (Amersham Life Sciences) to verify the identity of the insert. After isolation of the recombinant plasmid from a 1 L 2xYT culture, the DNA was digested with *Ban*I and the fragments were separated by nondenaturing polyacrylamide gel electrophoresis. The DNA monomer was eluted from the polyacrylamide slice using the crush and soak method.[25]

**Polymerization of the DNA Monomer and Cloning of DNA Multimers.** Purified monomeric DNA was self-ligated with T4 DNA ligase to form a population of multimers. Multimers were resolved by electrophoresis and ligated into *Ban*I digested, dephosphorylated p937.51, a small, high copy number cloning vector that carries an origin of replication from pBR322 and a gene encoding chloramphenicol acetyltransferase.[29] The recombinant plasmids were used to transform *E. coli* strain HB101. Transformants were screened by restriction enzyme digest analysis and clones encoding 60 and 16 repeats of the monomer were selected. The recombinant plasmids were designated p937.51-$(AG)_{240}$ or p937.51-$(AG)_{64}$, respectively.

**Construction of Bacterial Expression Vector.** Recombinant p937.51-$(AG)_{240}$ or p937.51-$(AG)_{64}$ was digested with *Bam*HI to release the multimer. The fragments were separated by 1% low melting point agarose gel electrophoresis and recovered by extraction with phenol and with phenol/chloroform, and finally by ethanol precipitation. The recovered DNA was ligated into *Bam*HI digested, dephosphorylated pET3b and used to transform *E. coli* strain HB101. Transformants were screened for the presence and orientation of each multimer by *Ava*I digestion. Recombinant plasmids, designated pET3b-$(AG)_{240}$ or pET3b-$(AG)_{64}$, respectively, containing inserts in the correct orientation were used to transform the expression host, *E. coli* strain BL21(DE3)pLysS.

**Protein Expression: Batch Culture, Rich Medium.** Single colonies of BL21(DE3)pLysS containing recombinant pET3b-$(AG)_{240}$ or pET3b-$(AG)_{64}$ were used to inoculate 5 mL of LB medium containing chloramphenicol (34 μg/mL) and ampicillin (200 μg/mL). The cultures were grown to saturation and used to inoculate 750 mL volumes of LB medium containing chloramphenicol (34 μg/mL) and ampicillin (200 μg/mL). The cultures were grown to $OD_{600} = 1$ at 37 °C, and protein expression was induced by the addition of IPTG to a final concentration of 0.4 mM. Cells were harvested after 2 h by centrifugation (2520*g*, 15 min at 4 °C). The cell pellets were resuspended in 10 mL of water and stored at −20 °C.

**Protein Purification.** Frozen cells were thawed and phenylmethylsulfonyl fluoride (PMSF) (1 mM), $MgCl_2$ (5 mM), RNAse (20 μg/mL), and DNAse (20 μg/mL) were added. The solutions were incubated at 37 °C for 30 min, and the insoluble portion was recovered by centrifugation (10000*g*, 4 °C, 15 min). The pellets were resuspended in 10 mL of TENT solution [50 mM Tris-Cl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 0.5−0.7% Triton X-100] and incubated with shaking at room temperature for 30 min. The insoluble portion was recovered by centrifugation (10000*g*, 4 °C) for 15 min. The pellets were resuspended in 10 mL of TENS [50 mM Tris-Cl (pH 8.0), 10 mM EDTA, 100 mM NaCl, 0.5% sodium dodecyl sulfate (SDS)] and incubated at room temperature with shaking for 60 min. The insoluble portion was recovered by centrifugation (10000*g*, 4 °C, 15 min) and the TENS treatment was repeated. The resulting pellet was resuspended in TEN plus porcine pancreatic lipase (20 μg/mL) and incubated at 37 °C for 30 min. The insoluble portion was recovered by centrifugation (10000*g*, 4 °C 15 min). The pellet was then washed sequentially with 10 mL of TENT, 10 mL of water, 10 mL of chloroform: methanol:acetone 40:30:30, 10 mL of acetone, 10 mL of water, and 10 mL of water. The insoluble portion was recovered by centrifugation (10000*g*, 4 °C, 15 min) after each wash. The recovered fusion protein was then dried and weighed.

**Cleavage of Fusion Protein.** Cyanogen bromide cleavage of the fusion proteins was accomplished by the method of Smith.[30] Approximately 35 mg of fusion protein was dissolved in 5.5 mL of 98% formic acid and diluted to 70% formic acid with distilled water. The solution was deoxygenated with nitrogen, and 50 mg of CNBr crystals was added. The solution was stirred at room temperature for 24 h and the solvent was removed by rotary evaporation. The residue was washed with distilled water several times, and the insoluble portion was recovered by centrifugation and dried by lyophilization.

**High Cell Density Fermentation and Protein Expression.** A 5 mL volume of 2xYT medium containing chloramphenicol (34 μg/mL) and ampicillin (200 μg/mL) was inoculated with a single colony of BL21(DE3)pLysS pET3b-$(AG)_{240}$. The culture was incubated in a rotary shaker at 37 °C until visible turbidity was reached (approximately 6 h) and then used to inoculate 100 mL of 2xYT medium containing chloramphenicol (34 μg/mL) and ampicillin (200 μg/mL). This was grown to visual turbidity at 37 °C and then used to inoculate 2L of 2xYT medium supplemented with chloramphenicol (34 μg/mL) and ampicillin (200 μg/mL). The culture was grown to saturation overnight and the cells were harvested by centrifugation at 4000 rpm for 10 min. The resulting pellet was used to inoculate 35 L of the minimal medium[26] specified in Table 1. Ampicillin and chloramphenicol were added to final concentrations of 200 μg/mL and 34 μg/mL, respectively. Ammonium hydroxide (25%) was used as the base (which was fed through the base pump of the pH controller) both to control pH and to provide a nitrogen source. The culture was grown in an 80 L

**Table 1. Minimal Medium Composition for High Cell Density Fermentation**

| component | initial concn |
|---|---|
| $KH_2PO_4$ | 13.3 g/L |
| $(NH_4)HPO_4$ | 4.0 g/L |
| citric acid | 1.7 g/L |
| $MgSO_4 \cdot 7H_2O$ | 1.2 g/L |
| trace metal solution[a] | 10 mL/L |
| thiamine hydrochloride | 4.5 mg/L |
| glucose$\cdot H_2O$ | 27.5 g/L |
| antifoam | 0.1 mL/L |
| feed solution | |
| glucose | 500 g/L |
| $NH_2OH$ | 25% |

[a] Trace metal solution: 6 g/L iron(III) citrate, 1.5 g/L $MnCl_2 \cdot 4H_2O$, 0.8 g/L $Zn(CH_3COO)_2 \cdot H_2O$, 0.3 g/L $H_3BO_3$, 0.25 g/L $Na_2MoO_4 \cdot 2H_2O$, 0.25 g/L $CoCl_2 \cdot 6H_2O$, 0.15 g/L $CuCl_2 \cdot 2H_2O$.

New Brunswick fermentor until the glucose was depleted as determined by a rise in pH above the set point, pH 6.8. At this time a glucose feed (50% w/v glucose/water) was established. Glucose was fed through the acid pump of the pH controller; pH was maintained at 6.8.[26] Pure $O_2$ was used to ensure aerobic conditions when air was insufficient to maintain the dissolved oxygen concentration above 20%. Induction was achieved by the addition of 1.0 mM IPTG when the $OD_{600}$ reached 25−30. Cultures were grown for 4 h prior to harvesting by centrifugation with a Cepa Z41 continuous centrifuge. Cells were weighed and resuspended in 4 mL of water per gram of cells (wet weight). Protein purification was performed as described above.

**Infrared Spectroscopy.** Poly(AG) comprising 64 diads ($(AG)_{64}$) was prepared for crystallization by dissolving 30 mg of protein in 1 mL of 98% formic acid. The solution was diluted to 70% formic acid by addition of distilled water and stirred until a gel formed (about 10 h). The gels was washed with methanol, dried, and crushed to form a KBr pellet for infrared analysis.

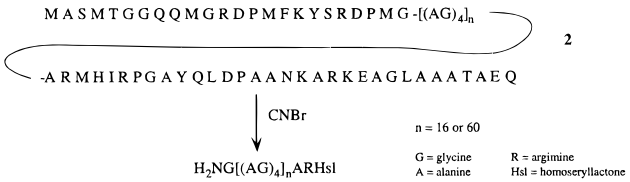
**X-ray Diffraction.** Poly(AG), in the form of a white powder, was continuously added to a mixture to dichloroacetic acid and 10% aqueous lithium hydroxide at 60 °C and stirred until a gel formed. The gel was compressed between two Teflon filters at 50 °C until dry (approximately 4 h). X-ray diffraction photographs from these compressed crystalline mats were obtained using a Statton-type X-ray camera, evacuated to backing pump pressure to reduce air scatter. The nickel-filtered Cu Kα sealed beam source was collimated with a system of 200 $\mu$m pinholes. The diffraction photographs were recorded on X-ray film with specimen-to-film distances of 50-100 mm. X-ray diffraction photographs of the mats were obtained with the X-ray beam directed either parallel or perpendicular to the surface of the mat.

**Computer Modeling and Simulated X-ray Diffraction Patterns.** Molecular modeling was performed on a Silicon Graphics Iris 4D20 workstation using the following software packages: Discover 2.9 (Biosym Technologies) and Cerius 2, Release 1.6 (Molecular Simulations, Inc.). The simulated X-ray diffraction patterns were created on a Silicon Graphics Indigo R4000 using the Diffraction 1 module of the Cerius 2, Release 1.6, program. The appropriate Lorentz and polarization correction factors are incorporated in the displayed intensities. The degree of arcing and the temperature factor were chosen to match the observed X-ray diffraction photographs. The final structure was checked for close contacts using the Geometry module of the Cerius 2 program.

## Results and Discussion

**Gene Construction.** The strategy used in our laboratory to prepare artificial proteins has been described previously.[7] In this work, DNAs encoding two chain-length variants of poly(AG) were ligated into the *Bam*HI site of the expression plasmid pET3b. The resulting recombinant plasmids encode either 16 or 60

**Scheme 1**




repeats of the alanylglycine tetramer, flanked by N- and C-terminal extensions of 23 and 33 amino acids, respectively (**2**). These terminal regions are derived from the transfer and expression vectors and can be removed by CNBr cleavage at flanking methionine residues (Scheme 1).

**Protein Expression.** The host used for protein expression was *E. coli* strain BL21(DE3)pLysS.[24] In this strain, a gene encoding T7 RNA polymerase is incorporated into the bacterial chromosome under *lacUV5* control, and protein production is induced by the addition of IPTG. The pLysS plasmid provides low levels of T7 lysozyme, which inhibits T7 RNA polymerase and suppresses the basal level of protein expression.

Protein expression in rich medium was performed as described in the Experimental Section. Aliquots of cultures representing approximately equal numbers of cells (as determined by the optical density at 600 nm) were removed periodically after induction of protein synthesis ($t$ = 0, 30, 60, 120, and 180 min). Samples were loaded onto a 12.5% polyacrylamide gel and separated by electrophoresis. Protein bands were stained with Coomassie Brilliant Blue R-250. The recombinant protein was insoluble in SDS and did not migrate during SDS-PAGE; however, increased stain intensity following induction could be seen in the wells of the stacking gel (data not shown), indicating inducible protein production.

Protein production in minimal medium was carried out in a New Brunswick 80L Mobile Pilot Plant as described above. The growth medium (Table 1) was adapted from that reported by Riesenburg et al.[26] Oxygenation of the culture must be maintained in order to ensure aerobic growth during the induction period; otherwise no recombinant protein is produced. Oxygen was bubbled through the medium at a low rate which maintained the dissolved oxygen content approximately 20% above the $O_2$ level obtained when the medium was deoxygenated with $N_2$. After the termination of batch phase growth ($OD_{600}$ = 21), a 50% (w/v) glucose solution was fed on demand through the pH pump. Upon depletion of glucose as the carbon source, the culture begins to metabolize acetic acid previously produced as a result of glucose metabolism. Depletion of acetic acid results in an increase in pH; the acid pump is turned on and glucose is fed to the culture. Acetic acid is then produced and again lowers the pH. This allows for the addition of both the carbon source (glucose) and the nitrogen source (ammonium hydroxide) to be fed on demand through the pH control mechanism of the system. The addition of glucose by pH control is similar to the method used by Frude et al.,[31] except that in the present work, the base source is also the nitrogen source.

**Protein Purification and Analysis.** Table 2 shows the weight of recombinant protein recovered per gram wet cell weight for cultures of BL21(DE3)pLysS pET3-b-$(AG)_{240}$ grown to optical densities (600 nm) of 1, 10, or 25 prior to protein induction with IPTG. An optical density of 25 corresponds to approximately 10 cell doublings of the inoculum. There is no loss of protein

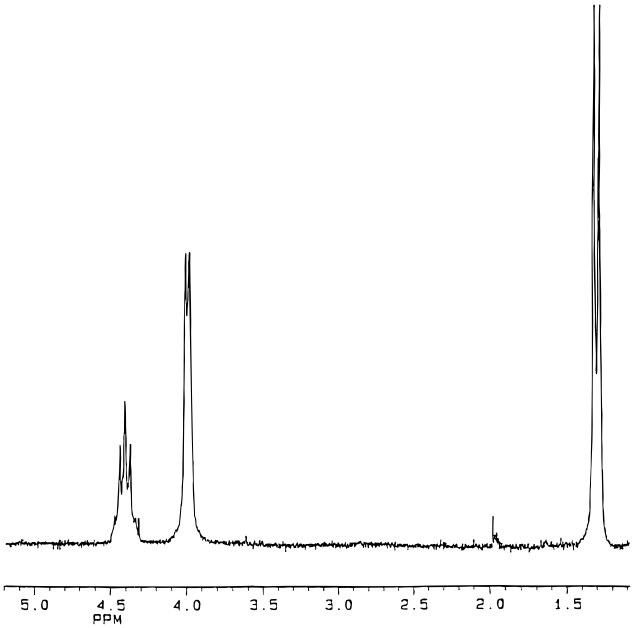



**Figure 1.** 500 MHz $^1$H NMR spectrum of $(AG)_{64}$ in [formyl-*d*]formic acid. The alanine α-CH signals appear at 4.45 ppm (α-CH) and 1.36 ppm ($CH_3$). The glycine α-$CH_2$ protons appear at 4.05 ppm.

**Table 2. Recombinant Protein [$(AG)_{240}$ Fusion] Recovered from Cultures of Three Different Cell Densities**

| $OD_{600}$ | wet cell weight (g/L) | recombinant protein/wet cell weight (mg/g) | product yield (g/L) | total product yield (g) |
|---|---|---|---|---|
| 1 | 3 | 16 | 0.048 | 0.048 |
| 10 | 26 | 20 | 0.52 | 5.2 |
| 25 | 64 | 15 | 0.96 | 25 |

**Table 3. Theoretical and Experimental Amino Acid Analysis for $(AG)_{64}$ and $(AG)_{240}$**

| | mol % $(AG)_{64}$ | | mol % $(AG)_{240}$ | |
|---|---|---|---|---|
| amino acid | theor | exptl | theor | exptl |
| aspartic acid | 0.00 | 0.62 | 0.00 | 0.00 |
| alanine | 49.23 | 52.59 | 49.79 | 49.00 |
| glycine | 49.23 | 46.28 | 49.79 | 46.54 |
| arginine | 0.77 | 0.51 | 0.21 | 0.00 |
| homoserine | 0.77 | — | 0.21 | 0.00 |
| valine | 0.00 | 0.00 | 0.00 | 4.45 |

produced per gram of wet cell weight (WCW) as the cell density is increased from 3−4 g WCW/L to 64 g WCW/L prior to protein induction. This result indicates that we are not at the limit of the method and suggests that these cultures might be grown to still higher densities with further increase in the amount of recombinant protein produced per liter of culture.

Amino acid analyses for both $(AG)_{64}$ and $(AG)_{240}$ are shown in Table 3. The $(AG)_{240}$ product was found to contain approximately 5% valine. Two-dimensional NOSY NMR spectra of the recombinant protein confirmed the presence of valine in the polymer (data not shown), suggesting mutation of the coding sequence subsequent to sequencing of the DNA monomer. Conversion of alanine to valine was observed previously in our work on the related polypeptide $[(AG)_4PEG]_{14}$.[32] The shorter recombinant molecule does not appear to contain these changes as determined by amino acid analysis (Table 3).

The $^1$H NMR spectrum of $(AG)_{64}$ in formic acid (Figure 1) shows the alanine α-proton multiplet at 4.45 ppm and the $CH_3$ doublet at 1.36 ppm. The glycine α-$CH_2$ signal occurs at 4.05 ppm. The absence of other signals indicates sample purity consistent with the results of amino acid analysis reported in Table 3.

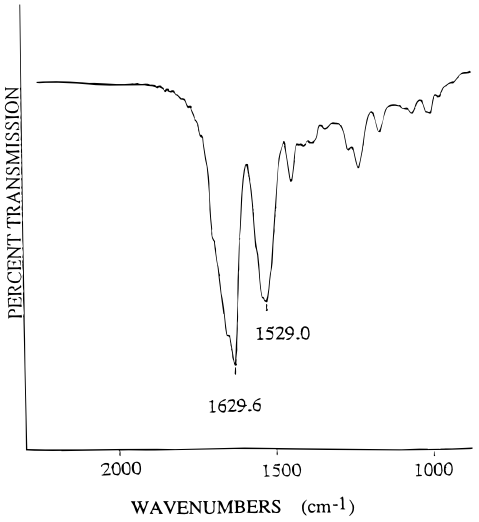



**Figure 2.** FTIR spectrum of silk II form of $(AG)_{64}$. Amide vibrations at 1529 and 1629 $cm^{-1}$ indicate the *β*-sheet architecture.

**Table 4. Elemental Analysis of $(AG)_{64}$ Including both Experimental and Theoretical Calculations**

| element | theor | theor ± 3% $H_2O$ | exptl |
|---|---|---|---|
| carbon | 46.85 | 45.45 | 45.39 |
| hydrogen | 6.33 | 6.48 | 6.35 |
| nitrogen | 21.86 | 21.20 | 19.78 |

The results of elemental analysis are shown in Table 4 and are compared with the theoretical analysis for the product containing 3% water. Addition of 3% water gives good agreement with the experimental results for carbon and hydrogen content; we find nitrogen analysis consistently lower than theoretical for these kinds of artificial proteins.[4,7]

**Solid-State Properties.** Crystalline samples of $(AG)_{64}$ were prepared as described in the Experimental Section. The infrared spectrum of the crystalline polymer, shown in Figure 2, exhibits amide I and amide II vibrational modes at 1629 and 1529 $cm^{-1}$, respectively, characteristic of the *β*-sheet conformation.

The wide-angle X-ray diffraction photograph, taken with the X-ray beam directed parallel to the plane of the compressed mat and with the mat normal horizontal, is shown in Figure 3a. The diffraction photograph exhibits discrete Bragg diffraction signals consistent with an oriented crystalline polymer, and all observed signals index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 0.922$ nm, and $c = 0.695$ nm (errors ± 0.002 nm). The measured and calculated interplanar spacings are listed for comparison in Table 5 together with an estimate of the observed intensities. In addition, a set of low-angle equatorial diffraction arcs, orders of 3.2 nm, was seen. This set of diffraction signals is noticeably different in character as compared to the wide-angle diffraction signals; the degree of arcing is about a factor of 2 less and they vary in intensity relative to the wide-angle signals in different samples; indeed, in some cases they are so weak that they are barely detectable, except for a diffuse first order at a spacing 3.6 nm. Their appearance or nonappearance did not change the nature of spacing of the other diffraction signals. The orientation direction is along

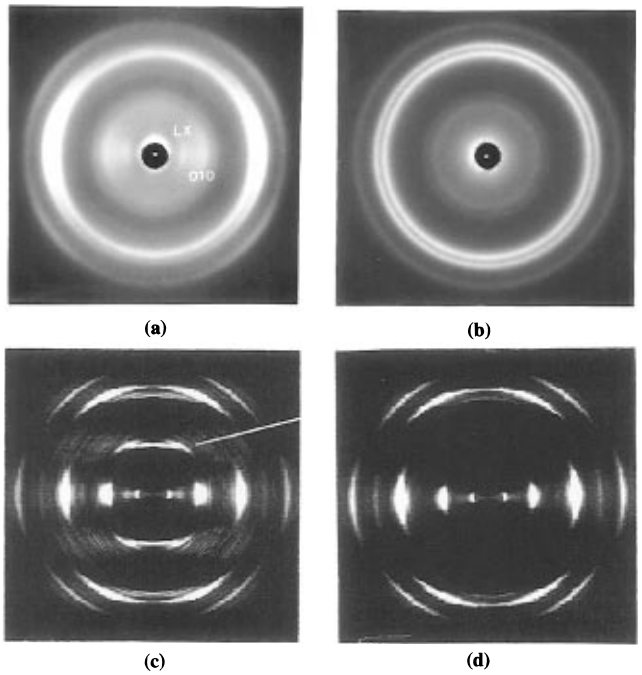



**Figure 3.** X-ray diffraction photographs from $(AG)_{64}$. (a) With the X-ray beam directed parallel to the plane of the compressed mat and with the mat normal horizontal. The diffraction signals on this texture-oriented pattern index on an orthorhombic unit cell with dimensions *a* (hydrogen bond direction) = 0.948 nm, *b* (inter-ap$\beta$-sheet direction) = 0.922 nm, and *c* (chain direction) = 0.695 nm. Only *h*00 diffraction signals occur on the meridian (vertical bisector) and 0*kl* on the equator (horizontal bisector). The low-angle diffraction signals are indicated by LX. (b) With the X-ray beam directed orthogonal to the plane of the compressed mat. In this case, the diffraction signals are a set of concentric rings which index on the same orthorhombic unit cell as in (a). (c) SXDP of the poly(AG) model proposed by Fraser et al.,[14b] with alternating $\pm a/4$ shears in the *ac* plane. Notice the unwanted diffraction intensity on the first layer line ($h = 1$) as indicated. (d) SXDP of the chain-folded model shown in Figure 7 which incorporates random $\approx\pm a/4$ shears and random $\pm c/2$ shears in the *ac* plane.

**Table 5. Comparison of Observed Diffraction Signal Spacings ($d_0$) with Those Calculated ($d_c$), in Nanometers (Errors ± 0.002 nm), for the Orthorhombic Unit Cell of the $(AG)_{64}$, Together with an Estimate of Observed Intensities ($I_0$)[a]**

$(AG)_{64}$: $a = 0.948$ nm, $b = 0.922$ nm, $c = 0.695$ nm

| *hkl* | $d_o$ | $d_c$ | $I_0$ |
|---|---|---|---|
| 010 | 0.922 | 0.922 | S |
| 020 | 0.461 | 0.461 | S |
| 021 | 0.375 | 0.384 | M |
| 030 | 0.310 | 0.307 | M |
| 031 | 0.282 | 0.281 | M |
| 022 | | 0.277 | |
| 040 | 0.231 | 0.231 | W |
| 200 | 0.474 | 0.474 | M (sharp) |
| 210 | 0.420 | 0.422 | M |
| 211 | 0.356 | 0.360 | M (broad) |
| 301 | 0.282 | 0.288 | M |
| 311 | | 0.275 | |
| 400 | 0.237 | 0.237 | W (sharp) |
| 410 | 0.230 | 0.230 | W |
| 006 | 0.116 | 0.116 | W |

*a*. Thus the directionally coincident *a* and *a** axes lie along the meridian line (vertical bisector), and therefore, the *h*00 diffraction signals, the easily seen 200 and 400 in particular, appear as arcs, centered on that meridian. The families of 0*k*0 and 00*l* diffraction signals appear on the equator (horizontal bisector). This type of texture oriented X-ray diffraction pattern is similar to that reported previously for poly$(AG)_3EG$.[4−6] A detailed explanation is given in ref 6.

Differences in the line broadening of the various diffraction signals, which relate in a consistent way to their assigned Miller indices, are observed in Figure 3a. The 200 and 400 diffraction arcs are particularly sharp and demand long-range correlational order along the *a*-axis. Diffraction signals with indices of the general *hkl* type are considerably broader than those with indices *hk*0; this comparison leads us to the conclusion that the coherent scattering length in the *c* direction is small. We estimate a length of 2−4 nm (allowing for instrumental line broadening); this is ca. 16 times less than the length of the protein chain. The coherent scattering length in the *b* direction is intermediate between those in the *a* and *c* directions.[7]

The wide-angle x-ray photograph taken with the X-ray beam directed orthogonal to the compressed mat is shown in Figure 3b. The diffraction signals appear as a set of concentric rings, which index on the orthorhombic lattice deduced from analysis of the oriented pattern.

The texture-oriented wide-angle X-ray diffraction photograph shown in Figure 3a, obtained with the X-ray beam parallel to the surface of the mat, is similar to that previously reported for oriented films of poly(AG) by Fraser et al.[14a,b]; however, in that paper, no X-ray diffraction photographs taken with the X-ray beam orthogonal to the films were reported. The films were stroked during drying and it was implied that the *a*-axis is a *unique* direction;[14a] as a consequence, their structure analysis assumed cylindrical averaging (fiber symmetry) about the *a*-axis. In our samples, prepared by compression into mats, the *a*-axis is merely confined to the plane of the mat, with random rotation about the mat normal.[11] It is vital to recognize that there is a texture and to establish its nature if the X-ray structure analysis is to be performed correctly.[11] The overall features of our X-ray diffraction pattern are sufficiently similar to those of Fraser et al.[14] that we are confident that we have the same basic crystalline structure, even if our texture is different.

Fraser et al.[14b] proposed the following model. The structure consists of polar ap$\beta$-sheets, stacked in pairs with alanyl faces together and glycyl surfaces together, respectively. The stacking direction is parallel to the *b* axis; the unequal packing of the ap$\beta$-sheets is responsible for the appearance of the 010 diffraction signal. Successive sheets shear $\pm a/4$ parallel to the *a* axis (hydrogen bond direction) in the *ac* plane. The *c* axis is parallel to the chain axis. The only difference between our unit cell and that reported by Fraser et al.[14b] is the value of *b*, the sheet stacking parameter; their reported value[14b] of 0.896 nm is 2.9% less than our 0.922 nm value. A perspective view of the structure and the unit cell is shown in Figure 4, and a simulated X-ray diffraction pattern (SXDP) of this structure generated for comparison with the observed X-ray data is shown in Figure 3c. Comparison of Figures 3a and 3c shows that although the match between the observed and calculated diffraction patterns is reasonably good overall, there are noticeable discrepancies, the most obvious of which is the unwanted 1*kl* diffraction signals on the first layer line.[33]

The low-angle equatorial diffraction pattern (which consists of orders of 3.2 ± 0.1 nm) and the relative broadening of selected wide-angle diffraction signals favor a chain-folded lamellar structure for the poly(AG)

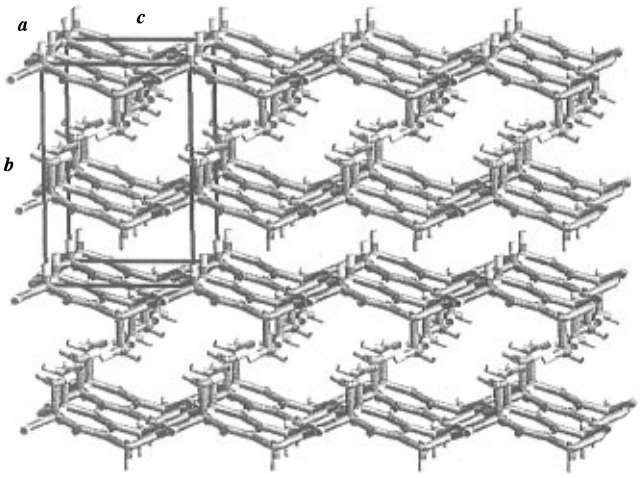



**Figure 4.** Three-dimensional structure generated by a stack of *polar* ap$\beta$-sheets of poly(AG) arranged so that like surfaces are in contact; thus the hydrophobic methyl groups are layered between a pair of ap$\beta$-sheets, followed by a layer of hydrogen atoms. The unit cell is shown; note the sheets are alternately sheared by $\pm a/4$ in the *ac* plane. The alternating deviation from the *b*/2 stacking periodicity is evident and is, in part, responsible for the occurrence of the 010 diffraction signal. The *b*-value is 2.9% greater than that reported for oriented films of poly(AG) by Fraser et al.[14b] We believe this is due to steric interactions at the regular (every eighth amino acid) $\gamma$-turns in our structure (see Figure 7).

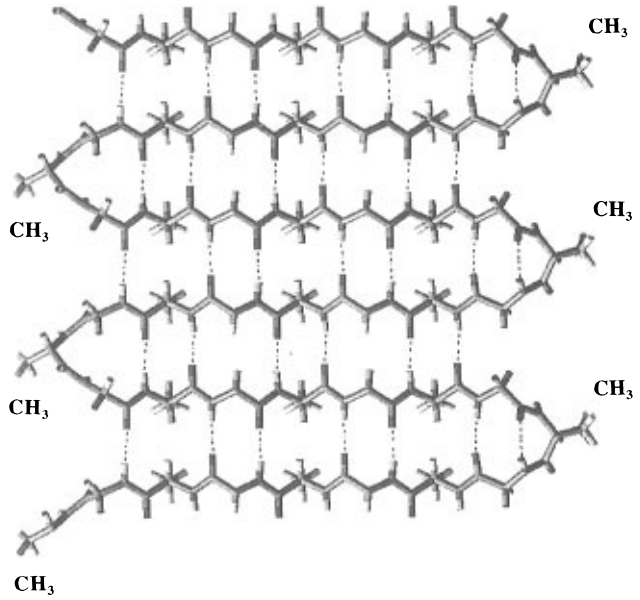



**Figure 5.** View orthogonal (parallel to *b*-axis) to a single ap$\beta$-sheet of poly(AG), folding in phase with the octapeptide periodicity and forming $\gamma$-turns. The chains (*c*-direction) run horizontally and the linear hydrogen bonds (shown as dashed lines) are directed vertically (*a*-direction). Note that the conformations of the $\gamma$-turns are different at the left hand and right-hand edges (fold surfaces) of the ap$\beta$-sheet. The peptides connect through *inward-pointing* bonds on the left and *outward-pointing* bonds on the right.[22]

crystals[34,35] prepared in this work. Figure 5 shows an orthogonal view (parallel to *b*-axis) of a chain-folded ap$\beta$-sheet that is consistent with the X-ray data. The $\gamma$-turns are necessary to generate the *polar* ap$\beta$-sheets.[4,5] Alanine residues are placed at the apex of each $\gamma$-turn.[36] It should be noted that the folds on the right-hand edge of the sheet are conformationally different from those on the left. The structure is constructed as follows. The second chain in the ap$\beta$-sheet is generated from the first (and starting chain conformation) by a rotation $\pi$ about the *b*-axis so that it occurs at a distance of *a*/2 away in the *ac* plane. Energy-minimized $\gamma$-folds were added to complete the chain-folded ap$\beta$-sheet structure. The second ap$\beta$-sheet was generated by rotation of the first ap$\beta$-sheet by a rotation $\pi$ about the *a*-axis, such that the alanine decorated surfaces were adjacent to each other and at an intersheet distance of greater than *b*/2 so that there were no short contacts. Thus, in the chain-folded lamellar crystal both surfaces are equally populated with the two types of fold conformation.[37] In the first instance the two ap$\beta$-sheets were displaced by *a*/4 and *c*/2 relative to each other. The intersheet Ala−Ala and Gly−Gly distances were refined so that no close contacts occurred and a good fit was obtained between the relative intensities of the 010 and 020 diffraction signals in the simulated diffraction patterns, as compared with the experimentally obtained X-ray diffraction pattern. Our intersheet distances differ from the Fraser et al.[14b] model for poly(AG), who gave values of 0.517 and 0.379 nm for the Ala−Ala and Gly−Gly intersheet distances, respectively. Our values are 0.554 and 0.368 nm, respectively. The presence of the regular chain folding in our proposed model (see Figure 7) creates slight overcrowding at the fold surfaces for the Ala−Ala intersheet region only. Thus our Ala−Ala intersheet distance is larger. Our Gly−Gly intersheet distance is less than that reported by Fraser et al.[14b] but contains no short contacts. A more detailed perspective view of one of the $\gamma$-turns is shown in Figure 6.

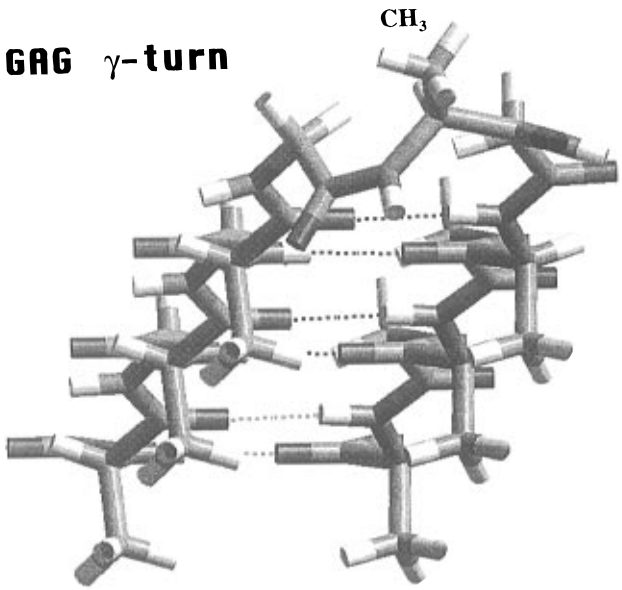



**Figure 6.** Perspective view of a GAG $\gamma$-turn (from left-hand side folds of Figure 5). The methyl groups of the alanine residues on the underside of the ap$\beta$-sheet surface can be clearly seen. The dashed lines represent the hydrogen bonds. Note the $CH_3$ group at the apex of the $\gamma$-turn.

Various structures for poly(AG) were computer generated and the stereochemical feasibility of each was examined using modeling procedures. SXDPs were generated for comparison with the observed X-ray data. We discovered that all structures that incorporate regular intersheet shears of $\pm a/4$ parallel to the *a* axis generate unwanted intensity for the 1*kl* diffraction signals. This can be seen in the SXPD shown in Figure 3c. The only way to remove this intensity is by incorporating random $\pm a/4$ shears in the *ac* plane. In addition, there are unwanted diffraction signals on the equator which are removed by random $\pm c/2$ shears in

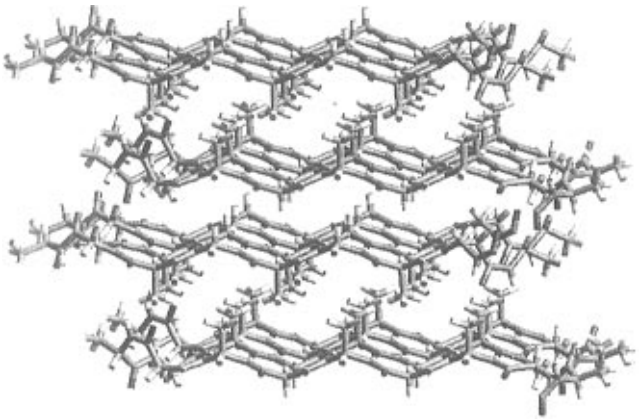


**Figure 7.** Computer-drawn view of refined structure of the chain-folded lamellar structure of poly(AG). The SXDP of this structure is shown in Figure 3d and can be compared with the experimental X-ray photograph shown in Figure 3a.

the *ac* plane. We have also found this effect in the structure of poly$(AG)_3$ EG.[4] A better fit still can be found if the random shears are only approximately $\pm a/4$. The closest overall fit with the observed X-ray data is illustrated in Figure 3d and the final structure is shown in Figure 7.

Fraser et al.[14] did not raise the issue of chain folding in the ap$\beta$-sheets of poly(AG). In our recent studies of related periodic polypeptides,[4,5,39] crystallized in very similar circumstances, we have found that the chains fold regularly through $\gamma$-turns to form crystalline lamellae approximately 3−4 nm thick. Chain folding, of course, is an obvious geometric contender in structures involving antiparallel chains. We suspect that our slightly increased value for the average intersheet stacking parameter $b/2$ is a result of steric interactions at fold surfaces.[38]

## Conclusions

Poly(AG) has been prepared by biological synthesis in multigram quantities; 25 g of the recombinant fusion protein was recovered from 35 L of culture. This required only a simple fed batch fermentation without elaborate feedback controls. The observation that there is no decrease in recombinant protein production with increased cell density suggests that limiting yields have not yet been reached. Little or no decrease in recombinant protein production with increased cell doubling also indicates good plasmid stability in the host strain.

The X-ray diffraction results show that poly(AG) can crystallize in the form of chain-folded lamellae approximately 3.2 nm thick. The structure consists of polar ap$\beta$-sheets folding through $\gamma$-turns. The sheets stack in pairs with alanyl faces together and glycyl faces together, respectively, with a stacking periodicity about 3% greater than that reported previously by Fraser et al.,[14] where chain folding was not evident. To obtain good agreement between the calculated and observed X-ray structure factors, successive sheets must shear randomly $\approx\pm a/4$ and $\pm c/2$ in the *ac* plane. The calculated diffraction data are presented in the form of look-alike patterns which constitute a convenient, stringent test for the proposed structure.

**Acknowledgment.** This work was supported by grants from the Polymers and Genetics Programs of the National Science Foundation. NMR spectra were recorded in the University of Massachusetts NMR Facility, which is supported in part by the NSF Materials Research Science and Engineering Center at the University.

## References and Notes

(1) Bailey, F. J.; Blankenship, J.; Condra, R. Z.; Ellis, R. W. *J. Ind. Microbiol.* **1987**, *2*, 47.
(2) Bauer, K. A.; Ben-Bassat, A.; Dawson, M.; De La Puente, V. T.; Neway, J. O. *Appl. Environ. Microbiol.* **1990**, *56*, 1296.
(3) Parkhe, A.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1993**, *26*, 6691.
(4) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427.
(5) Krejchi, M. T.; Deguchi, Y.; Atkins, E. D. T.; Mason, T. L.; Fournier, M. J.; Tirrell, D. A., submitted for publication.
(6) Atkins, E. D. T., Appendix to ref 5.
(7) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.
(8) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213.
(9) Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779.
(10) Deguchi, Y.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Macromol. Sci., Pure Appl. Chem.* **1994**, *A31*, 1691.
(11) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471.
(12) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198.
(13) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Mater. Res. Soc. Symp. Proc.* **1990**, *174*, 267.
(14) (a) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706. (b) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C. *J. Mol. Biol.* **1966**, *19*, 580.
(15) Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 201.
(16) Keller, A. *Philos. Mag.* **1957**, *2*, 1171.
(17) Fisher, E. *Z. Naturforsh., A* **1957**, *12*, 753.
(18) Geil, P. H. *Polymer Single Crystals*: Interscience: New York, 1963.
(19) Keller, A. *Rep. Prog. Phys.* **1968**, *31*, 624.
(20) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343.
(21) Bellinger, M. A.; Waddon, A. J.; Atkins, E. D. T.; MacKnight, W. J. *Macromolecules* **1994**, *27*, 2130.
(22) Atkins, E. D. T.; Hill, M. J.; Hong, S. K.; Keller, A.; Organ, S. J. *Macromolecules* **1992**, *25*, 917.
(23) Yee, L.; Blanch, H. W. *Biotechnol. Bioeng.* **1993**, *41*, 781.
(24) Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. *Methods Enzymol.* **1990**, *185*, 60.
(25) Sambrook, J.; Frisch, E. F.; Maniatis, T. *Molecular Cloning, A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Press: Cold Spring Harbor, NY, 1989.
(26) Riesenburg, D.; Schultz, V.; Knorre, W. A.; Pohl, H.-D.; Korz, D.; Sanders, E. A.; Rob, A.; Deckwer, W.-D. *J. Biotechnol.* **1991**, *20*, 17.
(27) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
(28) Yanisch-Perron, C.; Vierira, J.; Messing, J. *Gene* **1985**, *33*, 103.
(29) Ferrari, F. A.; Richardson, C.; Chambers, J.; Causey, S. C.; Pollock, T. J. U.S. Patent Appl. 927,258, 1986.
(30) Smith, B. J. *Methods in Biology, New Protein Techniques*; Humana: Clifton, NJ, 1988; p 70.
(31) Frude, M. J.; Read, A.; Kennedy, L.; *Biotechnol. Lett.* **1993**, *15*, 797.
(32) Beavis, R. C.; Chait, B. T.; Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 7584.
(33) Unwanted first layer line diffraction occurs even if fiber symmetry is assumed (although, of course, the relative intensities will be different). It can be removed by random shears of $\pm a/4$ in the *ac* plane; thus, the model proposed by Fraser et al.[14b] is not correct in detail. In addition, there are other discrepancies between the calculated and experimental X-ray diffraction patterns; there are unwanted 0*kl* equatorial diffraction signals in the region between the 010 and 020, and there is poor agreement with the observed *h00* intensities.
(34) Our interpretation of the poly(AG) X-ray diffraction results has benefited from the detailed analysis of poly$(AG)_3$EG.[4,5]
(35) The appearance of the first few orders of the 3.2 nm on the equator indicates that the lamellae are stacking reasonably well at the stacking periodicity of 3.2 nm. In a number of

crystal preparations, only a diffuse first order at spacing 3.6 nm (12% greater) was observed. This is consistent with poorly stacked lamellae.

(36) There are two possible sequences for the $\gamma$-turns, either GAG or AGA. Since we know *a priori* that the resolution of the experimental data will not allow us to delineate the fine details of the fold geometry, we chose the GAG $\gamma$-turn as the basis of our model.

(37) With regard to the stacking of the *polar* ap$\beta$-sheets, there are two possibilities if like surfaces are to be in contact, either a rotation, $\pi$, about the *a*-axis, or a rotation, $\pi$, about the *c*-axis. The former structure was favored since it results in a chain-folded lamellar crystal with both surfaces equally populated with the two types of fold conformations. These considerations are discussed in more detail in ref 5.

(38) It was noticed that the *b*-value decreased as the fold period increased in the series poly$(AG)_xEG$[4,5] as the integer *x* changed successively from 3 to 6, which would be consistent with our suggestion. Further support was found in the systematic study of the series poly$(AG)_3XG$, where X varied from serine to tyrosine; a linear relationship between increasing volume and increasing *b*-value was established.[39] Given the low degree of polymerization of the poly(AG) used by Fraser et al.,[14a] it may be that little, if any, chain folding occurred. No low-angle diffraction features were reported.

(39) Cantor, E. J.; Atkins, E. D. T.; Cooper, S. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A., submitted for publication.

MA961059M