Determination of the Chain-Folding Pattern in the Crystalline Domains of the Repetitive Polypeptide {(AlaGly)$_3$GluGly(GlyAla)$_3$GluGly}$_{10}$ by FTIR Studies of Its Blends with a ^{13}C-Enriched Analogue

A. D. Parkhe,† M. J. Fournier,‡ T. L. Mason,‡ and D. A. Tirrell*,†

Departments of Polymer Science and Engineering and of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Massachusetts 01003

Received August 23, 1993
Revised Manuscript Received September 27, 1993

Introduction. We are interested in the design and characterization of repetitive polypeptides that assemble into lamellae of uniform, predetermined thickness (Figure 1). In the initial stages of this work, the design of the polypeptide repeating unit (which comprises a sequence of amino acids traversing the thickness of the lamella and a turn sequence) has been based in part on work on the structure of silk and its analogues[1-4] and in part on the literature on reverse turns in globular proteins.[5-7] In particular, we have reported the synthesis and characterization of polypeptides comprising repeating unit sequences 1–3.[8-12]

$$-(AG)_3EG(GA)_3EG- \quad (1)$$

$$-(AG)_nEG- \quad (n = 3\text{–}6) \quad (2)$$

$$-(AG)_nPEG- \quad (n = 3\text{–}6) \quad (3)$$

In these designs, it was anticipated that the repetitive alanyl–glycyl (AG) diads would adopt β-sheet structures and that stacks of such sheets would constitute the lamella and define its thickness; the glutamyl–glycyl (EG) diad (in 1 or 2) or the prolyl–glutamyl–glycyl (PEG) triad (in 3) would force the chain to fold and reenter the lamella.

Wide-angle X-ray diffraction showed that all of these polypeptides, except 3 with $n = 3$, crystallize into domains comprising antiparallel β-sheets when precipitated from formic acid. Furthermore, small-angle X-ray diffraction results on crystallized 2 were consistent with the formation of lamellae of the anticipated thickness.[13] However, none of our diffraction analyses directly addresses the nature of chain folding, and the success of the assembly process (to yield structures such as that shown in Figure 1) depends on the ability of the chains to fold and crystallize in an adjacent reentry fashion.

In the 1950s characterization of polyethylene single crystals precipitated from dilute solutions showed the formation of folded-chain lamellae with thicknesses on the order of 100 Å.[14] The preference for adjacent reentry of the chains was demonstrated by Krimm and co-workers,[15] who carried out infrared studies on blends of polyethylene with deuterated polyethylene. A comparison of their spectroscopic results with those of a normal-mode analysis of unit cells containing varying fractions of deuterated polyethylene chains led to the conclusion that the chains prefer to fold with adjacent reentry along certain crystallographic planes.[15,16] The elegance of these deductions is underscored by the observation that the preferred fracture planes of the crystals are coincident with the fold planes inferred from the infrared data.

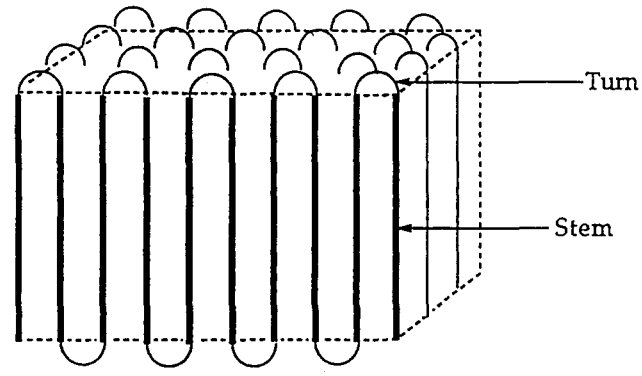


Figure 1. Schematic of the self-assembled lamellar morphology. The figure shows the crystallization of the repetitive polypeptide with adjacent reentry. In this model, the "stem" of the polypeptide consists of three alanyl–glycyl or glycyl–alanyl diads and the "turn" region consists of a glutamyl–glycyl diad. The thicker line represents the ^{13}C-enriched stem; the thinner line represents the natural-abundance polypeptide.

Single crystals have also been obtained and characterized for various nylons and polypeptides, by gradual precipitation of these polymers from dilute solutions.[17-21] The folding patterns for several nylons have been deduced by Atkins and co-workers from x-ray diffraction results.[21,22] We are, however, unaware of any work on nylons or polypeptides that probes for the folding habit directly.

Since the morphology desired in our work requires the chains to fold in adjacent reentry fashion, we are especially interested in the folding habit of the polypeptide chains. Furthermore, most work on single crystals has been based on crystallization from dilute solutions. At higher concentrations the folding analysis by diffraction methods is complicated by the formation of sheaflike or spherulitic aggregates of the single crystals.[14]

We have prepared a polymer (4) comprising 10 repeats of sequence 1, via a biosynthetic strategy described in an earlier paper.[9] This method of synthesis allows us to incorporate isotopically enriched amino acid residues in the sequence by supplementing the growth medium with the corresponding enriched amino acids. In this paper we present infrared data for blends of the natural-abundance (**4a**) and ^{13}C-enriched (**4b**) forms of the polypeptide **4** and an analysis of the folding habit during crystallization from solutions at concentrations of ca. 4 wt %. The results indicate that the chains fold predominantly in an adjacent reentry fashion even at the high polymer concentrations studied.

Experimental Section. The methods used for synthesis and expression of the artificial gene encoding **4** were similar to those described by McGrath and co-workers.[9] The sequence of DNA (**5**) corresponding to the repeating unit of **4** was prepared and polymerized as described therein and then was ligated into the unique *Ban*I cloning site of the cloning vector, pMD-3a.[23] The vector pMD-3a

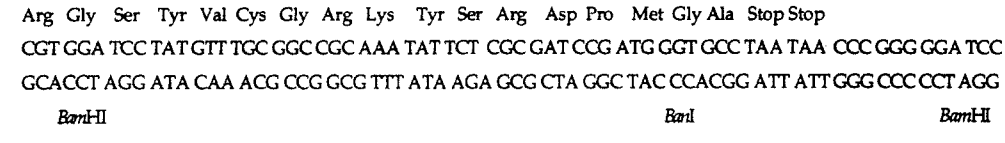
```
Gly Ala Gly Ala Gly Ala Gly Glu Gly Gly Ala Gly Ala Gly Ala Glu
GT GCC GGC GCT GGT GCG GGC GAA GGT GGC GCT GGT GCG GGC GCT GAA G
     G CCG CGA CCA CGC CCG CTT CCA CCG CGA CCA CGCCCG CGA CTT CCA CG
BanI                                                          BanI
```

5

contains, in the polylinker region (sequence **6**), two *Bam*HI sites flanking the *Ban*I site, a methionine codon preceding the *Ban*I site (to facilitate cleavage of the polypeptide by cyanogen bromide), and two termination codons. Sub-

† Department of Polymer Science and Engineering.
‡ Department of Biochemistry and Molecular Biology.

0024-9297/93/2226-6691$04.00/0



sequent to the insertion in pMD-3a, the population of DNA multimers was cloned in *Escherichia coli* strain HB101. A clone harboring DNA encoding **4** was isolated, and the *Bam*HI fragment was excised and transferred to the expression plasmid pET-3a.[9,24]

Fermentation on a 12*L*-scale was carried out as described by McGrath and co-workers[9] with the exception that the cells were grown in M9 minimal medium[25] supplemented with L-amino acids (20 μg/mL), glucose (2 mg/mL), and vitamin B1 (1 μg/mL). The presence of amino acids of natural abundance in the medium yielded a precursor of **4a** bearing a short (23-residue) N-terminal residue sequence; large-scale fermentation with ^{13}C-enriched glycine and alanine in the growth medium (99% ^{13}C, Cambridge Isotope Laboratories, Cambridge, MA) yielded a precursor of **4b**, containing the heavier isotope at the carbonyl positions of alanine and glycine. The protein was purified by a four-step reduction in the pH of the cell extract, leading to precipitation of impurities at every step. Ethanol was added to the supernatant at pH = 4.0 to a final concentration of 40%, and the mixture was incubated at 4 °C for 1 h. After centrifugation, ethanol was added to the supernatant to a final concentration of 80% and the mixture was incubated overnight at −20 °C. The purified protein was collected by centrifugation of the 80% ethanol suspension at 8000 rpm for 30 min at 4 °C and dried overnight in a vacuum oven at room temperature. The polypeptide of interest was obtained from the precursor by cleavage with cyanogen bromide. The polypeptide was dissolved at ca. 2.5 mg/mL in 70% formic acid, and cyanogen bromide was added to a concentration of ca. 2 mg/mL. The system was purged with nitrogen, and the flask was allowed to stand at room temperature in the dark for 24 h. Solvent was removed on a rotary evaporator, and the dried protein was suspended in distilled water at a concentration of ca. 6 mg/mL. The suspension was stirred for 24 h to extract the soluble proteins, and insoluble material was removed by centrifugation at 8000 rpm for 30 min. The clear supernatant contained **4**, which was precipitated by addition of ethanol to a final concentration of 80%. The precipitate was dried *in vacuo* at room temperature and then over P_2O_5 in a drying pistol with refluxing ethanol. The purity of the sample was confirmed by amino acid compositional analysis, combustion analysis, and ^1H NMR spectrometry (*vide infra*).

Polypeptide blends were crystallized from formic acid as follows. The mixture of polypeptides was dissolved in 90% formic acid, and the concentration of the formic acid was reduced to 70% by the addition of distilled water to obtain a 40-mg/mL protein solution. The solution was sheared for ca. 24 h by continuous stirring at ca. 700 rpm on a Thermix stirrer Model 220T (Fisher Scientific). The sample formed a gel which was precipitated by adding excess methanol. The gel was sedimented by centrifugation, and the precipitate was washed several times, first with water and then with methanol, and finally dried over P_2O_5 in a drying pistol with refluxing ethanol.

FTIR spectra were obtained on a Nicolet IR44 bench driven by PC/IR Version 3.00 software at a resolution of 1 cm^{-1} with 50 scans per spectrum. Sample pellets were made with IR-grade KBr on a "Macro–micro KBr die" (Aldrich Chemical Co.) on a 12-ton two-column laboratory press.

The yield of **4** was estimated at 40 mg/L. The amino acid analysis was consistent with the expected structure. Calcd: glycine, 50 mol %; alanine, 37.5 mol %; glutamic acid, 12.5 mol %. Found: glycine, 47.4 mol %; alanine, 31.3 mol %; glutamic acid, 12.3 mol %. The 300-MHz ^1H NMR spectrum and the combustion analysis were fully consistent with pure **4**. Anal. Calcd for **4** plus 3% water: C, 44.7; H, 6.1; N, 19.1. Found: C, 45.1; H, 6.1; N, 17.6. The incorporation of ^{13}C at the carbonyl positions of alanine and glycine was confirmed by ^{13}C NMR spectroscopy, and the labeling efficiency was estimated from the ^{13}C NMR spectrum to be approximately 90%.

Results and Discussion. The amide I vibration in polypeptides has been shown to contain contributions due to CO stretch (83%), CN stretch (15%), and CCN deformation (11%).[26] This region in the infrared spectrum (the unperturbed amide I absorption occurs at ca. 1655 cm^{-1}) has been shown to be sensitive to the conformation of the polypeptide chain and is characteristically split into two absorptions at ca. 1628 and 1701 cm^{-1} for antiparallel β-sheets.[26,27] Since this mode of vibration has a large contribution from the CO stretch, its frequency is dependent on the mass of the carbonyl carbon atom. The unperturbed frequency for the $^{13}C{=}O$ amide I vibration is expected to occur at ca. 1620 cm^{-1}, and the antiparallel β-sheet conformation should give rise to two characteristic vibrations, with the more prominent one at ca. 1591 cm^{-1}.[28] In both cases, the vibration at lower frequency is the dominant one and will be used herein in the analysis of the spectra. It has been suggested that the splitting of the unperturbed vibration is due to transition dipole coupling (TDC) between adjacent carbonyls within the hydrogen-bonded β-sheet.[27,29,30] Isotopic substitution of the carbon in some of the carbonyls surrounding an amide should result in a reduction of the frequency shift due to TDC of that amide. This phenomenon was first utilized by Lansbury and co-workers to determine locations of β-sheet regions in a model amyloid protein.[28]

We have studied the amide I region of the infrared spectra of blends containing **4a** and **4b** in varying ratios. The spectrum corresponding to 100% of **4a** shows the characteristic absorption at 1629 cm^{-1} due to the TDC of ^{12}C carbonyls in the antiparallel β-sheet structure (Figure 2). The TDC of ^{13}C carbonyls in the antiparallel β-sheet structure leads to an absorption at 1596 cm^{-1} in **4b** as shown in the sample containing 100% of **4b**. All intermediate blend compositions show the presence of both bands in the amide I region, and the relative absorption intensities scale qualitatively with the composition of the blend.[31]

These data lead us to propose that the majority of the chains are folding in adjacent reentry fashion. Any other structure, e.g., a fringed micellar morphology or a "switch board" reentry model,[14] would reduce the magnitude of the frequency shift due to TDC and move the amide I absorption toward the unperturbed frequency. Adjacent reentry of chains places chain segments containing ^{12}C carbonyls adjacent to one another and chain segments

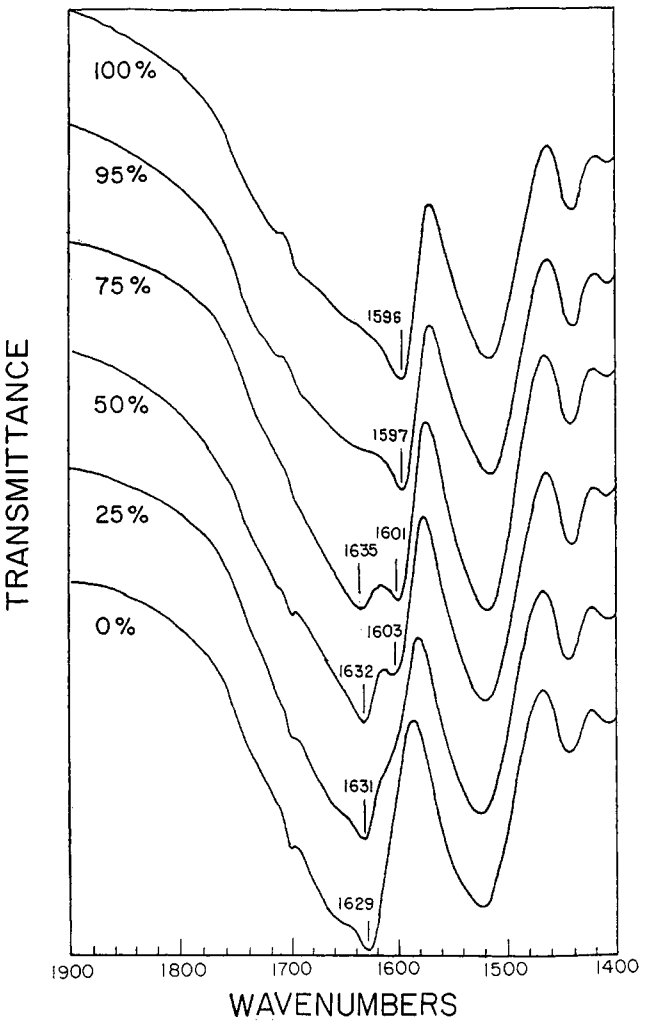


Figure 2. Infrared spectra of blends of **4a** and **4b**. The weight percentage of ^{13}C-enriched polypeptide **4b** in the blend is reported. Note the presence of two absorption maxima in spectra of all blend compositions.

containing ^{13}C carbonyls adjacent to one another. The frequency shift due to intrasheet interactions is therefore, not diminished, and the characteristic splitting is observed for amides containing either of the isotopes.

Acknowledgment. This work was supported by grants from Polymers and Genetics Programs of the National Science Foundation and by the NSF Materials Research Laboratory at the University of Massachusetts. We thank P. T. Lansbury (Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA) and S. L. Hsu (Department of Polymer Science and Engineering, University of Massachusetts, Amherst, MA) for discussions regarding FTIR spectroscopy.

References and Notes

(1) Lotz, B.; Brack, A.; Spach, G. *J. Mol. Biol.* **1974**, *87*, 193–203.
(2) Lotz, B.; Cesari, F. C. *Biochimie* **1979**, *61*, 205–214.
(3) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343–358.
(4) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706–712.
(5) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135–175.
(6) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759–777.
(7) Wilmot, C. W.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221–232.
(8) Krejchi, M. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1991**, *32* (1), 411–412.
(9) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727–733.
(10) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213–1214.
(11) Parkhe, A. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1993**, *34* (1), 150–151.
(12) The single-letter nomenclature is used for amino acids: A = alanine; G = glycine; E = glutamic acid.
(13) Krejchi, M. T.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Manuscript in preparation.
(14) Keller, A. *Rep. Progr. Phys., Part 2* **1968**, *31*, 623–704.
(15) Bank, M. I.; Krimm, S. *J. Polym. Sci., Polym. Phys. Ed.* **1969**, *7*, 1785–1809.
(16) Tasumi, M.; Krimm, S. *J. Polym. Sci., Polym. Phys. Ed.* **1968**, *6*, 995–1010.
(17) Dreyfuss, P. *J. Polym. Sci., Polym. Phys. Ed.* **1973**, *11*, 201–216.
(18) Geil, P. H. *J. Polym. Sci.* **1960**, *44*, 449–458.
(19) Padden, F. J.; Keith, H. D.; Giannoni, G. *Biopolymers* **1969**, *7*, 793–804.
(20) Keith, H. D.; Giannoni, G.; Padden, F. J. *Biopolymers* **1969**, *7*, 775–792.
(21) Atkins, E. D. T.; Hill, M.; Hong, S. K.; Keller, A.; Organ, S. *Macromolecules* **1992**, *25*, 917–924.
(22) Atkins, E. D. T.; Keller, A.; Sadler, D. M. *J. Polym. Sci., Polym. Phys. Ed.* **1972**, *10* (5), 863–875.
(23) Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779–1781.
(24) Rosenburg, A. H.; Lade, B. N.; Chui, D.; Lin, S.; Dunn, J. J.; Studier, F. W. *Gene* **1987**, *56*, 125–135.
(25) Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory: New York, 1972.
(26) Bandekar, J.; Krimm, S. In *Advances in Protein Chemistry*; Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eds.; Academic Press: Orlando, FL, 1986; Vol. 38; pp 183–364.
(27) Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69* (10), 2788–2792.
(28) Halverson, K. J.; Sucholeiki, I.; Ashburn, T. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1991**, *113*, 6701–6703.
(29) Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712–719.
(30) Cheam, T. C.; Krimm, S. *Chem. Phys. Lett.* **1984**, *107* (6), 613–616.
(31) The positions of the two main absorptions are not fixed at 1629 and 1596 cm^{-1} in all of the blend compositions. The slight changes in the frequencies of absorption as a function of blend composition may be due to small contributions to the TDC from nonadjacent reentry or from intersheet interactions between carbonyls.