Crystal Structures of Chain-Folded Antiparallel β-Sheet Assemblies from Sequence-Designed Periodic Polypeptides

Mark T. Krejchi,[†] Sharon J. Cooper,[§] Yoshikuni Deguchi,[†] Edward D. T. Atkins,[§] Maurille J. Fournier,[‡] Thomas L. Mason,[‡] and David A. Tirrell*,[†]

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

Received September 23, 1996; Revised Manuscript Received May 8, 1997[⊗]

ABSTRACT: The crystal structures and textures of a family of sequence-designed periodic polypeptides were investigated and analyzed using X-ray diffraction, vibrational spectroscopy, and cross-polarization magic angle spinning ^{13}C nuclear magnetic resonance. The repetitive amino acid sequences are described by −[(AG)$_x$EG]−, with integer x from 3 to 6. These macromolecules were prepared via bacterial expression of artificial genes and are monodisperse. Crystalline samples were obtained, and the interpretation of the X-ray diffraction results was aided by the generation of computer-simulated X-ray diffraction patterns. This allowed direct comparisons to be made with the observed texture-oriented X-ray diffraction photographs. All diffraction and spectroscopic evidence supports an antiparallel (ap) β-sheet structure, and all structures index on orthorhombic sublattices similar to those reported for *Bombyx mori* silk fibroin and poly(L-alanylglycine). The unit cell parameters for poly(AG)$_3$EG, for example, are $a = 0.948$ nm (hydrogen-bond direction), $b = 1.060$ nm (ap β-sheet stacking direction), and $c = 0.695$ nm (chain direction). Selective line broadening is observed for wide-angle diffraction signals with $l \neq 0$ (for the 211 in particular) and gives an estimated crystal size of <4 nm in the chain direction. This, coupled with the appearance of a low-angle particle interference peak at 3.6 nm, indicates a crystal size over an order of magnitude less than the chain length and suggests an adjacent reentry chain-folded lamellar structure incorporating the ap β-sheet architecture. A structure with polar ap β-sheets and γ-turns, stacking with the hydrophobic methyl groups of the alanyl residues in contact, is selected by X-ray structure refinement to give the best match with the experimental data. The pattern of crystallization behavior of the poly(AG)$_x$EG family is consistent with the folding periodicity being in-phase with the amino acid sequence so that the glutamic acid residues are confined to the lamellar surfaces.

Introduction

The controlled biosynthetic production of sequence-designed artificial proteins is an emerging area of polymer science and technology, with important ramifications for materials science, biomedical engineering, and structural biology. The factors governing the relationship between amino acid sequence and spatial architecture of protein molecules have been sought for some time. Significant progress has been reported in the design of elementary proteins that adopt predictable conformations, and preliminary efforts concerned with inducing higher-order protein folding have been successful in several laboratories.[1]

We have undertaken an enterprise in the controlled creation of crystalline protein structures based on repetitive amino acid sequences of fixed chain length.[2] A commonly occurring crystalline entity in polymers is the adjacent reentry chain-folded lamella.[3] It is generally accepted that chain-folded lamellae are common in linear polymers that are sufficiently flexible to form hairpinlike folds. It is well-known, from crystallography of globular proteins, that polypeptide chains can make compact turns, with either one or two amide groups (two or three amino acid C_α atoms) in the turn (β- or γ-turns, respectively). Indeed, the pioneering work of Keith, Lotz, Padden, and others in the late 1960s and early 1970s, showed that sequential polypeptides could crystallize in the form of chain-folded lamellae with thicknesses ≈5 nm;[4−7] futhermore, they crystallized in the β-form.[4−8] We were also aware in undertaking this work, that the aliphatic polyamides—the nylons—readily form chain-folded crystals.[9a,b] It is of particular interest to note that nylon 46[9a] folds with an amide group in the turn, rather than through the flexible polymethylene segments, as in polyethylene. Thus, we decided to direct a portion of our macromolecular bioengineering effort toward the design of chain-folded lamellar crystals. The basic structural elements we wish to control are as follows: (i) chain conformation, (ii) correspondence between sequence repeat and chain-folding periodicity, (iii) lamellar surface chemistry, and, possibly, (iv) interlamellar packing.

The choice of amino acid sequences was a result of a distillation of concepts and information drawn from polymer materials science and from protein structural biology. The family of polymers described herein can be represented by the generalized sequence[10]

$$-[(AG)_xEG]_n-$$

where $x = 3$, 4, 5, and 6. The alanylglycyl (AG) diads were selected because they are common in many β-silks and form extended β-strands, which usually assemble into antiparallel (ap) β-sheets that can stack to form stable crystals[11,8]. Adjacent chains in an ap β-sheet, as the notation implies, run in opposite directions—a fundamental ingredient in our crystal design. Thus, in our envisioned chain-folded lamellae, the AG sequences would create the straight stems and the additional EG

* Author to whom correspondence may be addressed.
[†] Polymer Science and Engineering Department, University of Massachusetts.
[‡] Biochemistry and Molecular Biology Department, University of Massachusetts.
[§] University of Bristol.
[⊗] Abstract published in *Advance ACS Abstracts,* July 15, 1997.

S0024-9297(96)01405-2 CCC: $14.00

diad would lie in the fold. The range of three to six alanylglycyl diads (8- to 14-mer repeating sequences) was chosen after consideration of the anticipated thickness of the chain-folded lamellae. There were two specific examples known to us of ap β-sheet proteins occurring naturally as chain-folded crystalline lamellae: (i) the silk egg-stalk of the green lace-wing fly *Chrysopa*[12] which has a pseudooctapeptide repeating sequence and lamellar thickness of approximately 3 nm, and (ii) the cell wall protein of the gas vesicles in blue-green algae, which has a lamellar thickness of about 4 nm.[13]

Glutamic acid (E) was incorporated into the repetitive sequences because of its bulk, polar nature, and chemical reactivity. It is also the weakest β-sheet former of the 20 natural amino acids, according to the Chou−Fasman[14] predictive scheme for protein conformation. The large size of E relative to A and G should deter its inclusion into the crystalline lamellar interior on steric grounds: the A-to-A intersheet spacing for poly(L-alanylglycine) (polyAG)[15] is 0.498 nm, whereas the smallest intersheet distance reported for the β-form of polyE is 0.78 nm,[4,16] nearly 60% larger. Keith et al.[4] also showed that even larger values, typically 0.88 nm, for the intersheet spacing were common for polyE when crystallized by solid state conversion from its calcium, strontium or barium salts. During crystal growth, interactions with the solvent should also act to segregate the glutamic acid residues to the lamellar surface. In terms of future studies, the ability to chemically modify the glutamic acid side group, either on the isolated chain prior to crystallization or subsequently at the lamellar crystal surfaces, offers attractive possibilities.

We describe herein the solid-state structures of four polymers of sequence $-(AG)_xEG-$, with $x = 3$, 4, 5, and 6. The genetic engineering methodology and the purification and biochemical characterization of the polymers have been described previously, and a preliminary discussion of the structural results has been published.[2b,e]

Experimental Section

Biochemical Synthesis and Molecular Characterization. The synthesis, sequence analysis, purification and characterization of the $[(AG)_xEG]_n$ family, for the four integer sets of x and n given by: $x = 3$, $n = 36$; $x = 4$, $n = 28$; $x = 5$, $n = 20$; and $x = 6$, $n = 14$ have been described in detail.[2b,e]

Crystalline Powders and Mats. Samples were prepared by stirring cyanogen bromide-cleaved protein[2b] (30 mg/mL) in formic acid overnight at room temperature. The solution was adjusted to 70% formic acid by addition of water and a gel formed. The rate of gelation increased with the alanylglycine segment length; from minutes in the case of poly$(AG)_6$EG to up to 7 days for poly$(AG)_3$EG. The gels were washed with a mixture of 70% formic acid and methanol before resuspension in methanol (40 mL) and incubation at −10 °C for 2 days to extract residual formic acid. Powders were prepared by centrifugation to isolate the solid component, followed by drying under vacuum overnight at room temperature. A crystalline mat was obtained by allowing the protein suspension to sediment from methanol (100 mL) onto a Teflon filter (10 mm) followed by removal of the methanol by slow filtration. The resulting methanol-swollen crystal mat was placed between two Teflon filters then sandwiched between two Whatman filters, and dried overnight at room temperature under compression between two glass plates.

X-ray Diffraction. X-ray diffraction photographs from powders and 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 μm pinholes. The diffraction photographs were recorded on X-ray film with specimen-to-film distances of about 4 cm (wide-angle) and 17 cm (low-angle), respectively. This enabled us to record diffraction data in the range of 7−0.2 nm. The X-ray powder diffraction photograph was obtained by packing the crystalline powder sample into a thin-walled glass tube. X-ray diffraction photographs of the crystalline mats were obtained with the X-ray beam directed either parallel or perpendicular to the surface. A cylindrical Debye-Scherrer powder-type X-ray camera was used to explore the ultrawide-angle diffraction region (<0.2 nm).

Computer Modeling and Simulated X-ray Diffraction Patterns. The model building and computer graphics were performed on a Silicon Graphics Iris 4D20 machine using the following software packages: Discover 2.9 (Biosym Technologies) and Cerius 2, Release 1.6 (Molecular Simulations Inc.). The X-ray diffraction patterns were simulated on the Silicon Graphics Indigo R4000 using the Diffraction 1 module of the Cerius 2, release 1.6, program. Parameters were adjusted to allow for the effect of texture on the relative intensities (see Appendix). The appropriate Lorentz and polarization correction factors were incorporated in the displayed intensities. The degree of arcing and the temperature factor were chosen to match the observed X-ray diffraction photographs.

Vibrational and CP/MAS Solid State ^{13}C NMR Spectroscopy. Fourier transform infrared spectra were obtained using an IBM IR32 Fourier transform spectrophotometer. KBr pellets were prepared from powder samples with a polymer weight concentration of 0.2%. Raman spectra were recorded using a Bruker FRA 106 Fourier transform Raman spectrophotometer. Cross-polarization magic angle spinning nuclear magnetic resonance (CP/MAS ^{13}C NMR) spectra were obtained at a frequency of 50.6 MHz from powder samples using a Bruker 200AC spectrometer equipped with a DOTY solids probe and IBM solids rack. Measurements were made 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 applied in data processing.

Results

X-ray Diffraction from Poly$(AG)_3$EG. The wide angle X-ray diffraction photograph of poly$(AG)_3$EG, 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 1A. The diffraction photograph exhibits discrete Bragg diffraction signals consistent with an oriented crystalline polymer, and all the observed signals index on an orthorhombic unit cell with dimensions $a = 0.948$ nm, $b = 1.060$ nm, and $c = 0.695$ nm. The measured and calculated interplanar spacings are listed for comparison in Table 1 together with an estimate of the observed intensities. A schematic diagram showing the indexing of the diffraction signals is shown in Figure 1B. A noticeable feature of the diffraction patterns shown in Figure 1a,b is that the orientation direction is along a. Thus the directionally coincident a and a^* axes lie along the meridian line (vertical bisector) and therefore the $h00$ diffraction signals, the easily seen 200 and 400 in particular, appear as arcs, centered on that meridian. Furthermore the families of $0k0$ and $00l$ diffraction signals appear on the equator (horizontal bisector). The equatorial region was explored further using a Debye−Scherrer X-ray diffraction camera which enabled the 006 diffraction signal to be observed at a spacing of 0.116 nm.

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 1A. The 200 and 400 diffraction arcs are particularly sharp. This demands long-range correlational order along the a axis. Diffraction signals with indices of the general hkl type are considerably broader than those with indices $hk0$. In particular the 211 diffraction signal is

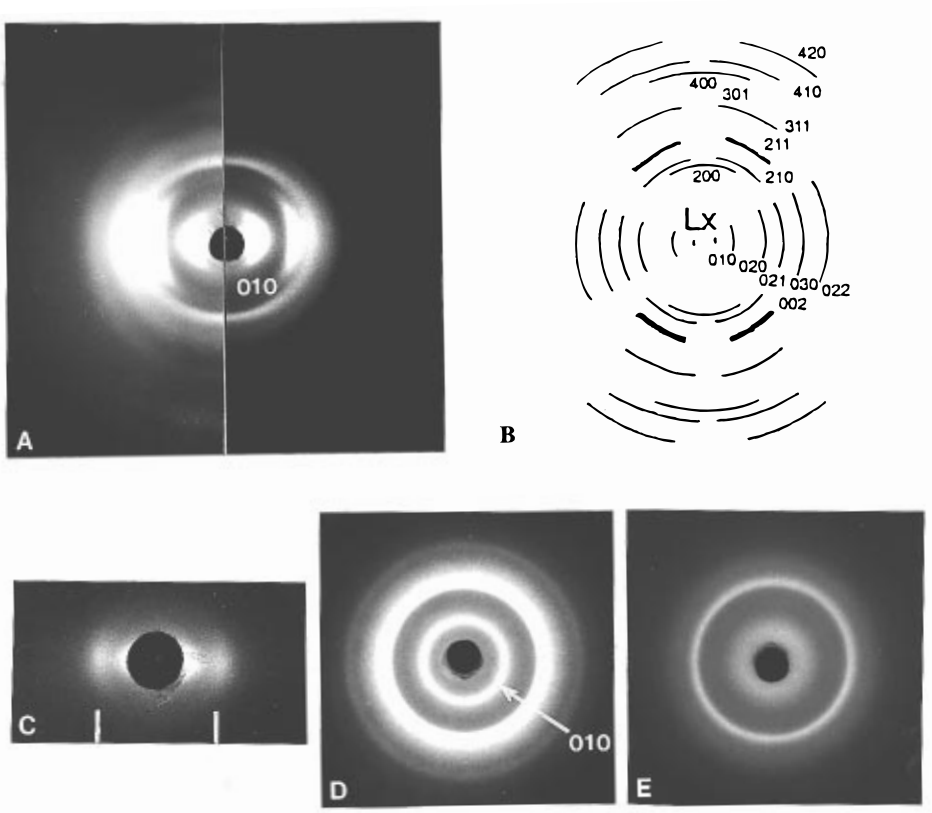

Figure 1. X-ray diffraction patterns from poly(AG)$_3$EG. (A) Wide-angle photograph with the X-ray beam directed parallel to the plane of the compressed mat; the mat normal is horizontal. The left-hand side is a more exposed version of the right-hand side. The 010 is indicated. (B) Schematic illustration of the prominent diffraction signals indexed on the orthorhombic unit cell of poly(AG)$_3$EG. The orientation direction *a* is along the meridian (vertical bisector). The low-angle diffraction arcs on the equator (horizontal bisector) are marked LX. Note the relative broadness of the 211. (C) Low-angle photograph of mat in the same orientation as in part a with peaks at 3.6 nm indicated. (D) X-ray diffraction from a powder sample. (E) Wide-angle photograph from poly(AG)$_3$EG with the beam orthogonal to the compressed mat. Note that the relative intensities of the 0*k*0 signals, the 010 in particular, are less than for the powder photograph (part D).

Table 1. Comparison of Observed Diffraction Signal Spacings (d_o) with Those Calculated (d_c), in Nanometers (Errors ±0.002 nm), for the Orthorhombic Unit Cell of Poly(AG)$_3$EG, together with an Estimate of Observed Intensities (I_o)[a]

hkl	d_o	d_c	I_o
010	1.060	1.060	vs
020	0.530	0.530	s
021	0.424	0.421	m
030	0.353	0.353	m
002		0.348	
012		0.330	
031		0.315	
022	0.298	0.291	m
040		0.265	
070	0.151	0.151	vw
006	0.116	0.116	w
200	0.474	0.474	m(sharp)
210	0.437	0.433	m
211	0.368	0.367	m(br)
301	0.285	0.288 }	m
311		0.278 }	
400	0.241	0.238	w(sharp)
410	0.230	0.231	w
420	0.212	0.216	w

[a] Poly(AG)$_3$EG: a = 0.948 nm, b = 1.060 nm, c = 0.695 nm. Key: vs = very strong; s = strong; m = medium; w = weak; vw = very weak; } : unresolved pair of diffraction signals.

estimated to be 10-fold broader than its 210 near-neighbor. These two signals are sufficiently close to each other in reciprocal space (and are not compromised by any other signals generated by the lattice) that we can make a realistic comparison of relative line widths. 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. The relative breadth of the prominent 010 and 020 diffraction signals allows us to establish that the coherent scattering length in the *b* direction is intermediate between those in the *a* and *c* directions.

Two additional wide-angle X-ray photographs, one obtained from a powder sample and the other taken with the X-ray beam directed orthogonal to the compressed mat, are shown in parts D and E of Figures 1, respectively. The diffraction signals appear as a set of concentric rings, and index on the same orthorhombic lattice that was deduced from analysis of the oriented pattern shown in Figure 1A, although the relative intensities are different. The 0*k*0 signals are weaker, most noticeably for the isolated inner 010 diffraction ring, as may be seen by comparison of parts A and D of the figure.

The low-angle X-ray diffraction photograph, taken with the X-ray beam parallel to the plane of the compressed mat, is shown in Figure 1C. This photograph shows a prominent arc, centered on the equator, corresponding to a spacing of 3.6 ± 0.1 nm, together with a weaker second order, which is reminiscent of an interparticle interference function. This low-angle diffraction disappears after swelling of the mats in glycerol. In contrast, with the exception of a minor change in the degree of orientation, the character and interplanar spacings observed for all other diffraction signals remain unchanged after swelling.

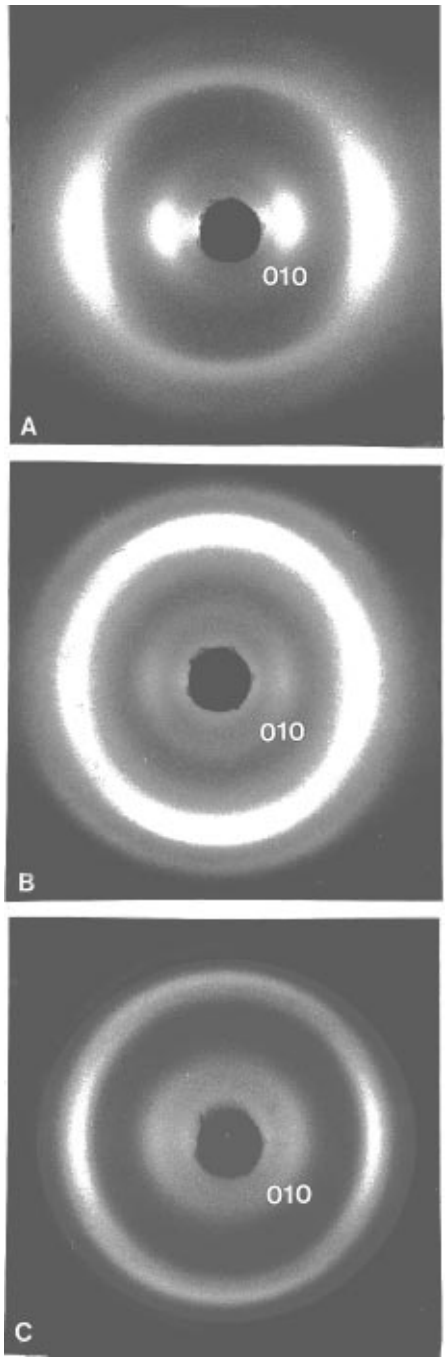

Figure 2. Wide-angle X-ray photographs taken under the same conditions as described in the caption to Figure 1A. (A) Photograph obtained from poly(AG)$_4$EG. The 211 is approximately twice as sharp as for poly(AG)$_3$EG. (B) Photograph from poly(AG)$_5$EG. The relative sharpness of the 211 is increased further. (C) Photograph from poly(AG)$_6$EG. Compared with the pattern from poly(AG)$_5$EG there is only a marginal increase in the relative sharpness of the 211.

X-ray Diffraction from Poly(AG)$_x$EG for $x = 4, 5$, and 6. The wide-angle X-ray diffraction photographs of these members of the family, taken with the X-ray beam directed parallel to the compressed mat and with the mat normal horizontal, are shown in Figure 2. The overall features of the X-ray patterns are similar to those of the X-ray photograph of poly(AG)$_3$EG shown in Figure 1A and all exhibit clear *a*-axis orientation. The degree of orientation varies within the family. We believe that this is a consequence of our relatively elementary methodology for specimen preparation (see Experimental Section) rather than some intrinsic differences among the polymers and that it has no particular significance with respect to the crystallization behavior of these materials or indeed their structural analyses which will be discussed later. The poly(AG)$_4$EG and poly(AG)$_6$EG X-ray photographs (parts A and C of Figure 2, respectively) exhibit lower degrees of orientation compared to poly(AG)$_3$EG and poly(AG)$_5$EG (see Figures 1A and 2B, respectively). The diffraction signals in all the X-ray patterns index on orthorhombic unit cells and the parameters are listed in Table 2. The *a* and *c* dimensions remain constant within the estimated experimental error of measurement (<2%; however, the value of *b* does change. There is a continual reduction in the value of *b* within the family, from poly(AG)$_3$EG to poly(AG)$_6$EG as three more alanylglycyl diads are inserted into the repeating motif; the overall decrease is 9.3%, and most of the change (8.5%) occurs between poly(AG)$_3$EG and poly(AG)$_5$EG.

Table 2. Comparison of Orthorhombic Unit Cell Dimensions (nm)

sample	*a*: hydrogen-bond direction	*b*: sheet stacking direction	*c*: chain-axis direction
poly(AG)$_3$EG	0.948	1.060	0.695
poly(AG)$_4$EG	0.948	1.028	0.695
poly(AG)$_5$EG	0.957	0.970	0.695
poly(AG)$_6$EG	0.964	0.962	0.695

More detailed examination of these X-ray photographs reveals that the diffraction signals with indices of the type *hkl* sharpen as the length of the repetitive alanylglycyl segment increases. Specifically, the diffraction signal indexed as 211 is approximately twice as sharp in poly(AG)$_4$EG as in poly(AG)$_3$EG, and the relative sharpness increases by another factor of 1.5 between poly(AG)$_4$EG and poly(AG)$_5$EG. Upon further increase in the length of the alanylglycyl repetitive segment to poly(AG)$_6$EG, no further change in the breadth of the 211 diffraction signal is evident.

Discrete low-angle arcs were observed on the equator of the X-ray photographs from poly(AG)$_4$EG and poly(AG)$_5$EG in the range 6.2−5.5 nm. No corresponding diffraction arc was observed for poly(AG)$_6$EG, but it might have occurred at a slightly larger spacing and been masked by the central beam stop in our experiments.

Vibrational and NMR Spectroscopy. The FTIR spectrum in the frequency range 3500−500 cm^{-1} for poly(AG)$_3$EG, together with an expansion of the 1800 to 1400 cm^{-1} section, which contains the amide I region, is shown in Figure 3A. The overall spectral features are the same for the whole poly(AG)$_x$EG family although the amide I vibrational mode at 1620−1623 cm^{-1} sharpens on going from poly(AG)$_3$EG to poly(AG)$_6$EG.

The Raman spectrum of poly(AG)$_3$EG in the frequency range 3500−500 cm^{-1} is shown in Figure 3B. The amide I band is observed at 1664 cm^{-1} and the amide III band is split into two components at 1260 and 1228 cm^{-1}. The observed chemical shifts from the CP/MAS ^{13}C NMR spectrum of poly(AG)$_3$EG are listed and compared with those reported for polyAG[17] in Table 3.

Discussion

We will proceed to describe the detailed structural analysis of poly(AG)$_3$EG and then discuss the differences, variations, and similarities noticed for the other members of this family of sequence-designed periodic polypeptides. We have chosen to display the comparison between the experimental data, obtained in the form of

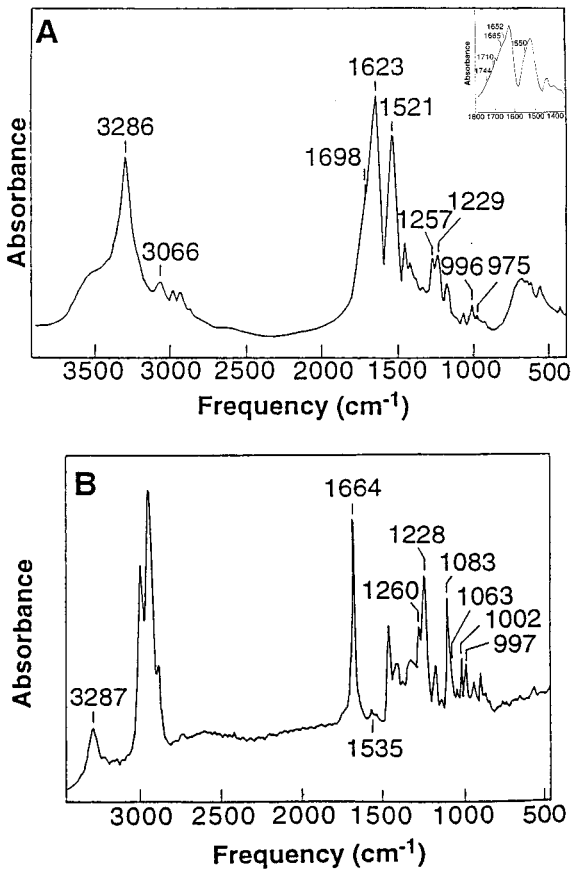

Figure 3. (A) Fourier transform infrared spectrum of poly-$(AG)_3EG$ in the frequency range 1800 to 500 cm^{-1}. The expanded amide I region (1800 to 1400 cm^{-1}) is shown as an inset. (B) Raman spectrum of poly$(AG)_3EG$. Note the 1664 cm^{-1} amide I band and the two components (at 1260 and 1228 cm^{-1}) of the amide III band, characteristic of ap β-sheet conformations.

Table 3. Chemical Shifts and Assignments for Selected Peaks Observed in the CP/MAS ^{13}C NMR Spectrum of poly$(AG)_3EG$ and Comparison with Data from the β-Form of polyAG[17]

chem shift (ppm)		
poly$(AG)_3EG$	polyAG	assignment
49.9	48.5	alanyl C_a
20.7	20.0	alanyl C_β
171.4	171.8	alanyl C=O
43.6	43.3	glycyl C_a
171.4	168.4	glycyl C=O

oriented X-ray diffraction photographs, and the calculated diffraction from the structures being tested for goodness of fit, in the form of look-alike patterns simulated using the Cerius system. This makes comparison between the observed and calculated results most convenient and instructive.

Analysis of the X-ray Diffraction Pattern from Poly$(AG)_3EG$. Basic Crystallographic Analysis. The interpretation of the oriented wide-angle X-ray diffraction pattern shown in parts A and B of Figure 1, supplemented with the interplanar spacings and estimates of observed relative intensities in Table 1, is aided by an understanding of the salient features of the crystal structures of β-silks.[8] The calculated unit cell parameters for poly$(AG)_3EG$ are commensurate with previously published X-ray diffraction results from various silk fibroins[18] and synthetic polypeptides that exhibit similar crystalline structures. These structures, first described by Pauling and Corey[19] in 1951, consist of sheets composed of relatively extended protein chains (with a 2-fold helical conformation) and with adjacent chains running in opposite polarity. The chains zip together through linear hydrogen bonds to form the so-called antiparallel β-sheet structure, as illustrated by the computer-drawn model shown in Figure 4A. The repeating distances along the chain axis (*c*) and in the hydrogen bond direction (*a*) are essentially invariant and therefore act as characteristic and diagnostic values for ap β-sheet structures. Reported values[8] for *B. mori* silk are 0.948 and 0.695 nm, respectively.

In the three-dimensional structure the sheets pack with in-phase puckering but usually with alternating shears of ca. $\pm a/4$ in the *ac* plane. The sheet stacking distance, which is related to the crystallographic axis *b*, is a function of the size and distribution of the amino acid side groups which decorate the upper and lower ap β-sheet surfaces. In the orthorhombic protein ap β-sheet structures, where *a*, *b*, and *c* are mutually perpendicular, there are two ap β-sheet stacking categories that need to be considered for the benefit of our structural analysis of poly$(AG)_3EG$. The first is when the decoration patterns on the upper and lower surfaces of the ap β-sheet are identical, as illustrated in Figure 4B. This can be described as an apolar ap β-sheet.[20] In this structure, the 0*k*0 diffraction signals for *k* odd would be systematically absent. In particular, the 010 would not be observed. The second arrangement is when the decoration patterns on upper and lower surfaces are different, which we describe as a polar ap β-sheet, as illustrated in Figure 4C. If such ap β-sheets stacked with like surfaces together, as reported for example for polyAG,[15] then a 010 diffraction signal *would* be observed; furthermore, its relative intensity would be a convenient indicator of the deviation from equally-spaced ap β-sheet stacking in the *b* direction. Thus these β-silklike structures are held together with covalent bonds in the chain (*c*-axis) direction, hydrogen bonds at right angles (*a*-axis) and van der Waals forces in the third mutually orthogonal *b* direction.

The *a* (=0.948 nm) value (Table 1) was assigned on the basis of its second diffraction order observed at a spacing of 0.474 nm. This value is the characteristic interchain hydrogen-bond distance in ap β-sheet crystalline structures.[8] It is also close to the interchain hydrogen-bond distance reported for many nylons[9a,b] and the insect cuticle amino−polysaccharide chitin.[21] Indeed it is a characteristic distance between polymeric chains held together through linear N−H···O=C hydrogen bonds. Thus, we 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. Values for this spacing have been reported from 0.344 nm for polyG,[22,8] to 0.79 nm for *Nephila senegalensis* fibroin,[18] a silk that contains a relatively high percentage of amino acid residues with bulky side chains, to 0.88 nm for polyE.[4] The values reported for poly(AG)[15] and the β-form of polyA[18] are 0.44 and 0.54 nm, respectively; thus, the spacing observed for poly$(AG)_3$-EG is in line with our expectation.

Texture. Before we can proceed to analyze the X-ray diffraction patterns in any detail and discuss possible structures, we need to understand the texture in the compressed crystalline mats. It was observed that when the X-ray beam was directed parallel to the mat plane, an oriented X-ray pattern was observed as illustrated, for example, in Figure 1A. On the basis of a single X-ray

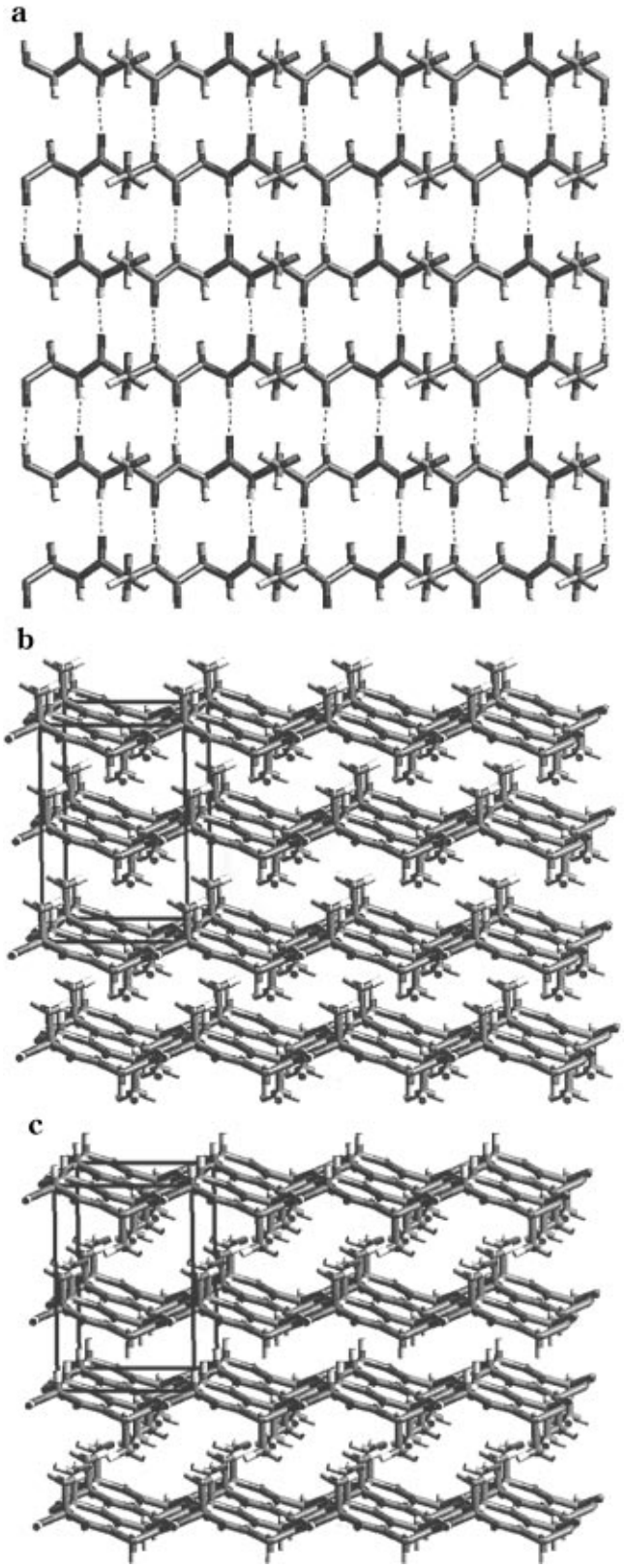

Figure 4. Computer-drawn models of ap β-sheet structures. (a) View of a piece of a single ap β-sheet orthogonal to the *ac* plane. (b) View of the structure of the stacked apolar ap β-sheets with the orthorhombic unit cell. Note the alternating $\pm a/4$ shear in the *ac* plane of successive sheets. No 010 signal would be expected. (c) View similar to view b for polar ap β-sheets with like surfaces in contact. The hydrophobic methyl groups are layered between one pair of ap β-sheets with a successive layer of hydrogen atoms. The sheets are alternatingly sheared by $\pm a/4$ in the *ac* plane. Note the deviation from the $b/2$ stacking periodicity, which would generate an 010 signal.

diffraction experiment of this kind it might be concluded that we had straightforward fiber-type symmetry, i.e., with cylindrical averaging about the *a*-axis direction only. It turns out that to make this assumption would be a serious error indeed. That this assumption is incorrect is clear from the fact that when the X-ray beam is directed along the crystalline mat normal, a set of concentric diffraction rings is observed, as shown in Figure 1E. This result demands that there is cylindrical averaging of the *a*-axis direction about the normal to the crystalline mat surface. This feature is also confirmed by the observation that the *same* X-ray fingerprint is obtained, no matter what the direction of the X-ray beam, *as long as it is parallel to the crystalline mat surface.* Thus we have a complex texture for these crystalline protein mats, the consequences of which result in cylindrical averaging about the normal to the mat surface, the *b*,*b**-axis, coupled with some degree of azimuthal disorientation about the *a*,*a**-axis. There are two worthwhile possibilities to consider regarding the nature of this interesting texture. One possibility is that the orienting forces are acting on crystalline aggregates and not on individual polymer chains. The geometry of these crystalline aggregates would need to be such that they have a substantially greater dimension in the *a* direction than in the two mutually orthogonal directions, *b* and *c*. This is wholly consistent with the observed line broadening of diffraction signals as a function of their Miller indices, as outlined in the Results. Also there is a precedent for this type of texture in a naturally occurring protein insect silk; the same *a*-axis orientation occurs in the thin fibrous egg stalks drawn by the green lace-wing fly *Chrysopa*[12] as a result of orienting preformed crystalline aggregates.

A second interpretation is that the crystalline entities grow preferentially in the *a* direction. It should be remembered that unlike the *Chrysopa* fibers, in these samples there is no unique *a* direction; it occurs at all angles in the plane of the compressed mat—that is, it is confined to a plane but radially directed. Furthermore, there is additional azimuthal averaging about each and every radius vector. This type of behavior is typical of polymers that grow and crystallize with spherulitic texture[23] or, perhaps more relevant in this context, spherulitic growth essentially confined to a plane—a spherodisc. The X-ray diffraction patterns from such a texture would be basically consistent with the X-ray observations regarding orientation and selective line broadening seen in Figures 1 and 2, since the size of the crystalline entities would be restricted in the *b*- and *c* directions and long in the *a* direction. Although at this stage, we are uncertain of the origin of the overlying texture in our samples, we do know the nature of the texture, and fortunately, as long as this textural feature is recognized and understood, it does not in any serious way interfere with the structural analyses of these protein crystals. Thus, we can effectively consider oriented patterns of the type shown in Figures 1 and 2 as if they were fiber-type X-ray diffraction patterns and analyze them accordingly. There will be implications for the computerized generation of simulated diffraction patterns from structural models, and these will be allowed for and described later. The details of the overlaying texture and its implications are given in an Appendix.

Chain Folding. The line broadening of diffraction signals with Miller index $l \neq 0$, the 211 in particular, indicates that there are interruptions to the crystal-

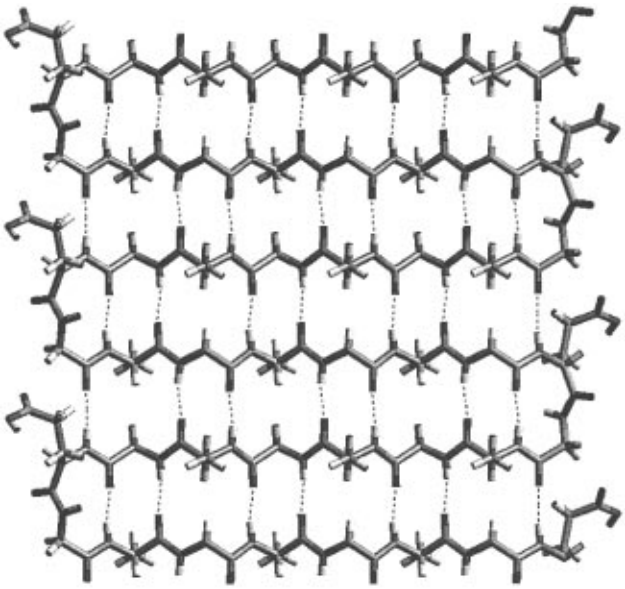

Figure 5. Computer-drawn view of a single ap β-sheet of poly-$(AG)_3EG$, folding in phase with the octapeptide periodicity and forming β-turns through the glutamic acid−glycine diads. The thickness is commensurate with the observed low-angle peak at a spacing of 3.6 nm. This is an apolar ap β-sheet with a similar distribution of hydrogen atoms and methyl side groups on both upper and lower surfaces.

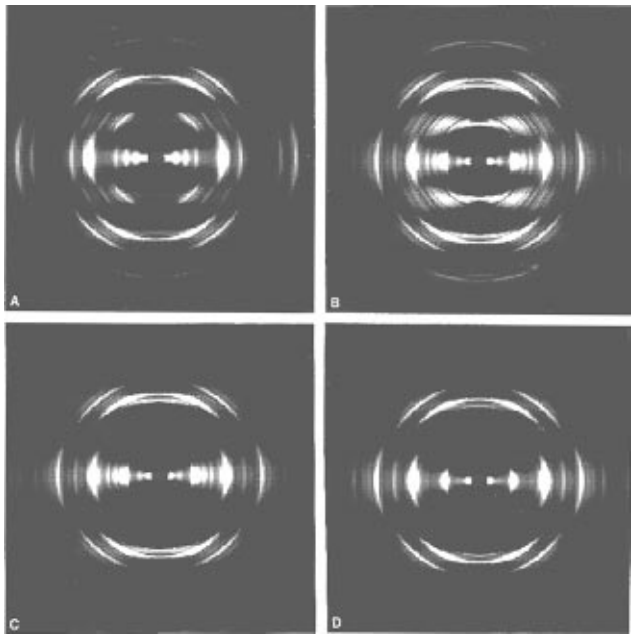

Figure 6. Simulated X-ray diffraction patterns (SXDPs) from proposed crystalline structures for the poly$(AG)_x$EG family. (A) SXDP of a crystalline structure based on the apolar ap β-sheet and β-turns shown in Figure 5. The important 010 is missing and there is some unwanted first layer-line diffraction. (B) SXDP of the starting crystalline structure using the chain-folded γ-turn model for poly$(AG)_3$EG. The ap β-sheets are sheared in an alternating manner by exactly $\pm a/4$ in the *ac* planes. There is considerable unwanted first layer-line diffraction. This can be removed by randomizing the positively and negatively directed shears, as illustrated in part C. (D) SXDP of the structurally refined model for poly$(AG)_3$EG.

lographic lattice in the *c* direction, that is, the chain direction, at distances ≤4 nm. This is ca. 25 times less than the length of the protein chain. The low-angle diffraction pattern shown in Figure 1E, with a strong peak at a spacing of 3.6 nm, resembles an inter/particle interference function. This diffraction peak can be made to disappear with swelling agents without affecting the underlying crystal structure. This behavior is reminiscent of interlamellar stacking and is similar to that reported for chain-folded lamellar crystals of polyethylene,[23] nylons,[9b] and *Chrysopa*[12] insect silk. Adjacent chains in our structure are antiparallel and the straightforward explanation for these results is that the chains in poly$(AG)_3$EG are folding regularly at a distance less than the measured interlamellar periodicity of 3.6 nm. This conclusion is bolstered by the observation in transmission electron microscopy of lamellae of the appropriate dimensions, as reported previously.[2b] The pattern of changes in the X-ray diffraction photographs from the other members of the poly$(AG)_x$EG family add considerable support to the chain-folded model. As the number of alanylglycyl diads is increased, the line width of the 211 diffraction signal systematically decreases; this change is to be expected as the straight stems in the chain-folded crystals increase in length, assuming the geometry of the folds remains intact. In addition, the measured spacing of the low-angle diffraction peak increases to accommodate the thicker crystal.

Nature of the Folds. The β-Turn Model. A model of a chain-folded ap β-sheet that is commensurate with the dimensions deduced from the X-ray diffraction data, and which folds in phase with the repetitive octapeptide sequence, is illustrated in Figure 5. The lamellar thickness is approximately 3 nm. There are three pairs of alanylglycyl units in each straight stem, and the remaining glutamylglycyl diad forms a β-turn. Thus all the glutamic acid residues occur at the fold surfaces. At first sight, the arrangement appears to be a perfect solution for our structure; however, it is an apolar sheet with a similar distribution of hydrogen atoms and methyl side groups on the upper and lower surfaces. As a consequence, we would anticipate regular stacking of the chain-folded ap β-sheets in the *b* direction and therefore the absence of the 010 diffraction signal, as discussed in the Basic Crystallographic Analysis section; contrary to observation. One possibility for modification is to consider interactions of the glutamic acid side groups at the fold surface that might induce pairing of sheets. For example, specific hydrogen-bonding patterns can be generated, by variation in the glutamic acid side groups, between near-neighbor carboxylic acid groups (the crystals were prepared in formic acid) between one pair of ap β-sheets so as to move their intersheet spacing slightly away from the average 0.53 nm stacking value. We considered the possibility that a movement of 50% of the glutamic acid side groups (accounting for 7.4% of the electron density of the crystal) off crystal lattice sites, together with the slight perturbation in the regular inter-sheet spacing, might be sufficient to generate the observed intensity of the 010 diffraction signal. Several candidate structures of this kind were examined and explored using computerized modeling procedures, and simulated X-ray diffraction patterns were generated for comparison with the observed X-ray data. We were unable to generate the necessary intensity for the 010 diffraction signal with any model of this kind, subject to the constraint that the calculated unit cell *b*-dimension equaled 1.060 nm. The simulated X-ray diffraction pattern (SXDP) shown in Figure 6A demonstrates that although the overall features match well enough, indicating the ap β-sheet is an essential element in any correct structure, the lack of the 010 diffraction signal is a serious discrepancy and we cannot accept this structure.[24] The decision to

abandon structures in this category was bolstered by the knowledge that the 010 was also the prominent diffraction signal in all the other members of the poly-$(AG)_xEG$ family (see, for example, the X-ray photograph of poly$(AG)_6EG$ shown in Figure 2C); as successive alanylglycyl diads are added, the possible perturbations at the fold surface and the percentage scattering from the glutamic acid side groups (4.2% for poly$(AG)_6EG$) will have a decreasing influence on both the interior ap β-sheet stacking and the ability to generate diffraction intensity at the 010 inter-planar spacing.

The γ-Turn Model. If the chain-folded ap β-sheets were polar in character, with one surface decorated entirely with methyl side groups and the other with hydrogen atoms and like surfaces were in contact, then the sheets would automatically stack with a displacement of sheets off the $b/2$ periodicity in an alternating manner, as shown in Figure 4c. This will produce a 010 diffraction signal. Indeed, this is precisely the model proposed for poly(L-alanylglycine)[15] 30 years ago, although in that case no chain-folding was invoked. In such an arrangement, the sheets are stacked in pairs with the hydrophobic methyl side groups sandwiched together. This raises a number of pertinent questions. Is there a folding mechanism able to accommodate apolar ap β-sheets, within the constraints imposed by the X-ray diffraction results, and keep the bulky glutamic acid residues at the fold surfaces? Is the folding geometry a consequence of the hydrophobic interactions of the methyl side groups? Returning to the first question: the computer-drawn view of the chain-folded ap β-sheet shown in Figure 7a illustrates such a solution. Each fold consists of three amino acids, rather than two. These tripeptide folds have been referred to as γ-turns[25] and are very common in globular proteins.[26] If the glutamic acid group remains in the fold there are three possible arrangements for a straight stem plus fold, namely $(AG)_2A$ GEG, $(GA)_2G$ AGE, and $(GA)_2G$ EGA. In the first instance, it seemed appropriate to choose the fold with the glutamic acid residue at the apex and with the small glycine residues on either side, as shown in Figure 7a. Close inspection of this computer-drawn view shows that folding with γ-turns, while keeping the glutamic acid residues within the fold itself, results in successive folds having slightly different fold conformations; the folds on the right-hand edge are conformationally different from those on the left. This is an inevitable consequence of having an odd number of amino acid residues in the stem segments of a chain-folded lamella. In the case of poly$(AG)_3EG$, if we use three amino acids for the fold, then, of course, we are left with only five amino acid residues for the stem. The same argument will apply to other members of the family. This is a subtle point and can be expressed in different ways. For example, at one edge of a chain-folded ap β-sheet, γ-turns connect adjacent antiparallel chains through *inward-pointing* bonds, while at the opposite edge the γ-turns connect adjacent antiparallel chains through *outward-pointing* bonds. This pattern of two conformationally different γ-turns would occur even if we chose the AGE or EGA type of γ-turn. There are other schemes of folding this polymer using combinations of GEG, AGE, and EGA, but all create a roughness to the chain-folded lamellar surface that increases the average thickness of the lamellae to an extent that is incompatible with the observed lamellar stacking periodicity of 3.6 nm. 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 model shown in parts a and b of Figure 7 as the basis of our proposed structure.

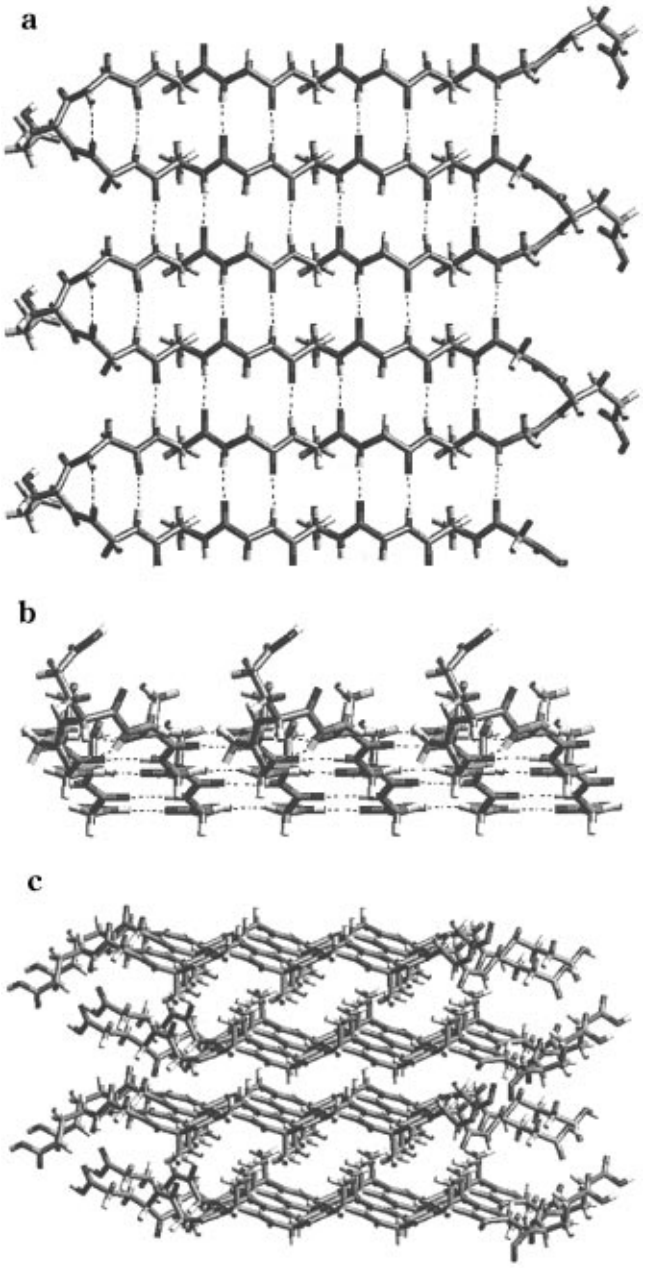

Figure 7. Computer-drawn views of the chain-folded γ-turn structure. (a) View orthogonal to (parallel to b-axis of) a single ap β-sheet of poly$(AG)_3EG$, folding in phase with the octapeptide periodicity and forming γ-turns with GEG sequences. Note the conformations of the γ-turns are different; the tripeptides connect through *inward-pointing* bonds on the left and *outward-pointing* bonds on the right. The glutamic acid side groups are at the fold apexes. (b) Oblique view of part a to illustrate the γ-turns. (c) View of the refined chain-folded lamellar crystal of the proposed structure for poly$(AG)_3EG$ with γ-turns. The methyl side groups form hydrophobic layers interspersed with layers of hydrogen atoms from the glycyl residues. Successive sheets are related by rotations of π about the a-axis. The sheets are stacked with random shears of approximately $\pm a/4$ and alternating $+c/2$ and $-c/2$ shears in the ac plane.

Stacking of γ-Turn ap β-Sheets. Care is required when considering the stacking of adjacent-chain reentry folded *polar* ap β-sheets with like surfaces in contact. Alternative ap β-sheets need to be imagined to be turned upside down prior to stacking. There are two ways of intercalating the sheets if the fold-edges

have different geometries; either a rotation of π about the *a*-axis or a rotation of π 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 conformation. In all structures, shears of approximately *a*/4 in the *ac* planes and parallel to *a* were incorporated.

Stacking considerations of these apolar ap β-sheets with like surfaces in contact introduce some subtle features. If, in all the sheets, we maintain the same amino acid sequence for the γ-turn, for example GEG, there is a consequential oscillating slip of 0.348 nm (*c*/2) between ap β-sheets; in the *ac* plane parallel to *c*. This is also a feature of having an odd number of amino acid residues within the ap β-sheet and can be seen in Figure 7c. Although it affects all the poly$(AG)_x$EG family, it is particularly acute for poly$(AG)_3$EG; it makes the thickness of the lamella almost the same as the 3.6 nm low-angle X-ray diffraction peak. In fact, the glutamic acid side groups cannot finger out from the lamellar surface in an extended conformation; they need to be in conformations more closely confined to the lamellar surface.

If, on the other hand, we pair a GEG γ-turn ap β-sheet with either of the AGE or EGA γ-turn ap β-sheets, then it is not necessary to have this 0.348 nm oscillating *c* direction slip. Although this makes for a slightly more comfortable fit of lamellae into the 3.6 nm packing periodicity, each glutamic acid residue is no longer at the apex of the γ-turn and is therefore more likely to disrupt the close-packing of the ap β-sheets. We were unable to choose between these subtly different stacking models on the basis of comparisons between calculated and observed X-ray diffraction. Thus, a structure with GEG γ-turns was chosen as a sufficiently representative model for the crystalline lamella.

Crystalline Lamellar Structure of Poly$(AG)_3$EG. Model structures were computer-generated and checked qualitatively for feasibility. Figure 7 illustrates the basic structure. In the first instance, successive *a*-axis shears of $\pm a/4$ were incorporated and lamellae were stacked on a one-dimensional superlattice of 3.6 nm spacing. The SXDP of this poly$(AG)_3$EG structure is shown in Figure 6B which should be compared with the X-ray diffraction photograph shown in Figure 1A. There is a reassuring degree of overall visual matching with respect to relative intensities, giving us confidence, at this stage in the structure analysis, that the model is basically correct. The 010 is there and obvious, and so is the 020, 021, 030, 022 equatorial diffraction set with roughly correct relative intensities. The prominent 210 and its broader 211 partner appear, and the 400 occurs on the meridian. It is opportune to mention that the intensity of the 210 results from the *a*/4 shears in the *ac* planes, as already pointed out in the structural analysis of polyAG.[15]

Now for the discrepancies and the structural changes necessary to reduce or eliminate them. An obvious feature in Figure 6B is the appearance of a set of diffraction signals on first layer-line with an unacceptably high level of intensity, contrary to observation. This diffraction can be eliminated by generating random shears of $\pm a/4$ in the *ac* planes, rather than alternating shears of $\pm a/4$. Random shears of $\pm a/4$ would be expected to occur since shears of $+a/4$ and $-a/4$ produce structures with identical energies; thus, these shears would have equal probability of occurrence. This was also discovered to be the case in the analyses of the polyA[18] crystalline structure, which showed no 1*kl* diffraction signals. Interestingly, first-layer diffraction signals do occur in *Chrysopa* insect silk, which demanded alternating $\pm a/4$ shears.[12] Before showing the results of this change, we will remove another discrepancy between the experimental and calculated diffraction patterns. Close inspection of the SXPD in Figure 6B shows that the 200 diffraction signal is absent. The intensity of the 200 diffraction signal can be raised by allowing the ap β-sheets to shear by only approximately $\pm a/4$ in the *ac* plane. This can be considered to be a natural consequence of the localized stacking arrangement. Any chosen *chain* in an ap β-sheet will be flanked by *two chains* in the adjacent ap β-sheet: one of which will have the same chain direction and the other the chain direction opposite to that of the chosen chain, resulting in asymmetry in the localized chain environment. Hence, in the minimum energy structure, the ap β-sheets will not have *exact* $\pm a/4$ shears in the *ac* plane. The intensity can be reduced to yield a better match with the observed X-ray photograph if the random $\pm a/4$ shears are only approximately *a*/4 in magnitude, with deviations of a few percent. Thus the structure is best described as one with successive ap β-sheets shearing randomly by approximately $\pm a/4$ in the *ac* planes. Two other differences between the SXDP pattern in Figure 6B and the observed X-ray diffraction photograph in Figure 1A require comment. The low-angle diffraction pattern can be changed, essentially independently of the wide-angle pattern, since the former originates from the one-dimensional interlamellar stacking and crystalline−amorphous electron density profile. Thus, it is straightforward to quench orders of the low-angle 3.6 nm spacing and control the relative intensity by introducing packing and lamellar surface disorder to fit the experimental data. The relative sharpness of diffraction signals embodying Miller indices *h*,*k* and *hk* combinations can be adjusted, with discretion, by changing crystal size and/or van der Waals stacking disorder (in the case of index triads involving $k \neq 0$) to give a better fit without compromising the basic crystalline structure of poly$(AG)_3$EG. Figure 6C illustrates the effect of randomizing the approximately $\pm a/4$ shears; the unwanted first layer-line diffraction signals have been removed, and the intensity match of the 200 and 400 diffraction signals is improved. However, three prominent unwanted diffraction signals, at 0.9, 0.78, and 0.68 nm spacing, remain. These signals result from the alternating shears of $+c/2$ and $-c/2$ in the *ac* plane. This condition was introduced into the structure when we decided that *all* of the γ-turns should have the GEG sequence. The three unwanted signals can be removed by allowing random $\pm c/2$ shears to occur between successive ap β-sheets. Random $\pm c/2$ shears would be favored by arguments similar to that outlined above for the random $\approx\pm a/4$ shears. Figure 6D illustrates the SXDP generated from a structure incorporating both random $\approx\pm a/4$ shears and random $\pm c/2$ shears. This SXDP is in good agreement with the experimental X-ray pattern from poly$(AG)_3$EG, shown in Figure 1A and is the basis for our proposed chain-folded lamellar structure illustrated in Figure 7.

Vibrational and NMR Spectroscopy. The infrared spectrum of poly$(AG)_3$EG exhibits amide I, II, and III vibrational modes at 1623, 1521, and 1229 cm^{-1}, respectively, characteristic of the β-sheet conformation,[15,27] and the weaker amide I component at 1698 cm^{-1} indicates the regular alternation of chain direction

that defines the ap β-sheet architecture[28]. Detailed examination of the amide I region (1800−1400 cm^{-1}), shown in the Figure 3 (insert), reveals the presence of four vibrational modes unrelated to ap β-sheet architecture. The weak vibrations at 1710 and 1744 cm^{-1} can be ascribed to carbonyl stretching of the side chain carboxyl groups of glutamic acid; however, the amide I component observed at 1652 cm^{-1} and the weaker shoulder at 1665 cm^{-1} indicate that a fraction of the polypeptide chain adopts a conformation different from the ap β-sheet. Normal mode calculations[29] suggest that these amide I components, together with the shoulder at 1550 cm^{-1} in the amide II region, may arise from reverse turns. The sharpening of the 1623 cm^{-1} band and the enhancement of the 996 and 975 cm^{-1} pair of bands, on passing through the family from poly$(AG)_3$EG to poly$(AG)_6$EG, indicate that the ap β-sheet structure is becoming better defined as the oligoalanylglycine repeating lengths increase. One further feature is the relative weakening of the bands at 1744 and 1710 cm^{-1}, associated with the glutamic acid residues (γ-C=O stretch), again from poly$(AG)_3$EG through to poly$(AG)_6$EG. This is consistent with a diminishing proportion of glutamic acid residues in the chemical sequence.

The Raman amide I band at 1664 cm^{-1} and the splitting of the amide III band into two components at 1260 and 1228 cm^{-1} are characteristic of the ap β-sheet structure,[27,30] and the broad, weak amide II band observed at 1535 cm^{-1} is expected for this structure.[31] It is plausible to associate the bands in the 1330−1300 cm^{-1} region with reverse turn conformations.[29]

The observed chemical shifts from the CP/MAS ^{13}C NMR spectrum of poly$(AG)_3$EG are compared with those reported for the β-form of polyAG in Table 3. All of these assignments are consistent with an ap β-sheet architecture,[2b] except for the presence of a shoulder at 16.8 ppm on the β-carbon signal from alanyl residues. The chemical shift of this shoulder is close to that observed for the β-carbon signal from alanyl residues in the silk-I structure,[17] but the absence of the corresponding carbonyl signal at 176 ppm suggests that the 16.8 ppm signal is not associated with the silk-I structure. The intensity of this signal was not reduced by washing the sample with formic acid, a good solvent for the silk-I polymorph, reinforcing the conclusion that it is not associated with silk-I. It would be wholly consistent with the X-ray diffraction results, and IR and Raman analyses, to propose that this 16.8 ppm signal arises from reverse turn structures.

Overall Structural Considerations and Implications. All of the experimental evidence supports a regular chain-folded ap β-sheet architecture for the crystalline form of poly$(AG)_x$EG. The X-ray diffraction analyses argue strongly for stacking of polar ap β-sheets with like surfaces in contact. This requirement demands an odd number of amino acids in each fold, *irrespective* of whether the glutamic acid residues remain on the fold surfaces. The γ-turn is the only serious contender. (We note in passing that a mixture of γ- and β-turns will not meet the requirements.) The *b*-dimension of poly$(AG)_3$EG of 1.060 nm is 19.5% larger than that reported for polyAG[15] (and 14% larger than *B. mori* silk fibroin[32]) and so this may be used to raise the question of whether glutamic acid residues enter the crystalline core of the lamellae; that is, whether the folding periodicity is out of phase with the octapeptide sequence. This is an important issue, since the choice of the amino acid sequence was part of our strategy for creating crystalline chain-folded lamellae with useful groups—in this case glutamic acid residues—wholly decorating the lamellar surfaces. Clearly the chain-folding periodicity is confined within certain limits, as imposed by the relative line width of the 211 diffraction signals, by the occurrence of low-angle diffraction peaks, and by the experimental evidence in general. On the other hand, if a mismatch of two amino acids (equivalent to a *c* direction translation of 0.695 nm) were permitted, it would be enough to allow the bulkier glutamic acid side groups to impregnate the crystalline lattice in the form of spatially distributed localized defects.[33]

There are a number of pieces of evidence against the decoupling of folding periodicity from the repetitive amino acid sequence, that is, in favor of keeping the glutamic acid residues at the fold surfaces. Although the *b*-parameter value is 19.5% greater for poly$(AG)_3$EG than for polyAG, this difference reduces, as more alanylglycyl diads are incorporated, to only 8.5% for poly$(AG)_6$EG (see Table 2). Thus, with the folds taking three-eighths of the available amino acids in poly$(AG)_3$EG, it is perhaps not surprising that the rather short stems (five amino acids) cannot develop sufficient van der Waals interactions to pull the ap β-sheets closer together against the envisioned overcrowding near the fold surfaces.[34] As the stems get progressively longer throughout the series, the *b*-parameter approaches that of polyAG,[15,35] a pattern of crystalline behavior consistent with the evidence. There is a counterargument, of course, that the relative dilution of glutamic acid residues, in passing from poly$(AG)_3$EG to poly$(AG)_6$EG, results in a corresponding reduction in the number of localized defects, allowing closer packing. This counterargument is weakened by the detailed computer analyses; if the glutamic acid units enter the lattice they occur selectively in the inter-alanyl decorated layers. Thus, the middle ap β-sheet in the unit cell (see Figure 7c) would move proportionally closer to the *b*/2 position in passing from poly$(AG)_3$EG to poly$(AG)_6$EG. This would change the ratio of the relative intensities of the 010 and 020 reflections, and to a lesser extent other diffraction signals, contrary to the X-ray results.[36] It is also worthwhile to consider what the lowest expected *b* value would be for such a structure. It would be 1.124 nm; the sum of 0.344 nm for polyG (see review by Fraser and MacRae[8]) and 0.78 nm (the lowest value reported by Keith et al[4]). This *b* value is 6% and 16.8% greater than the values measured for poly$(AG)_3$EG and for poly$(AG)_6$EG, respectively. The consequences of allowing glutamic acid residues to penetrate the lamellar core are shown in Figure 8.

The X-ray diffraction results show that, in passing from poly$(AG)_3$EG to poly$(AG)_6$EG, the coherent scattering length in the *c* direction increases substantially, as highlighted by the relative sharpening of the 211 diffraction signal; this is supported by the increased spacing of the low-angle signal. If glutamic acid residues can feed through and be accommodated in the crystalline core of the lamellae, there would be no reason why a particular member of this family should have a fold length different from those of any of the others. Thus the model fails to offer an explanation for the increase in fold length as alanylglycyl diads are added.

If glutamic acid residues feed through the crystalline cores of the lamellae we would expect to see a measurable improvement in the relative sharpness of diffraction signals containing Miller index *k*, the 010 and 020 in particular, on passing from poly$(AG)_3$EG to poly-

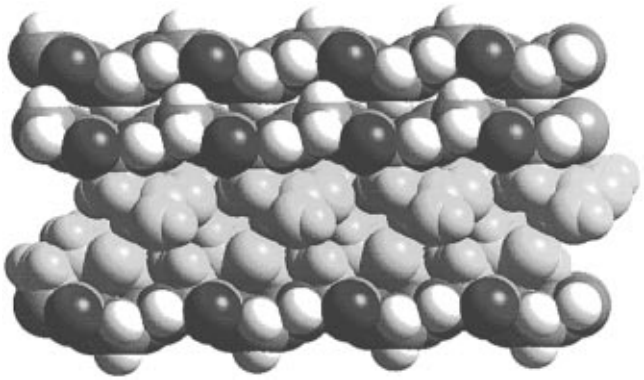

Figure 8. Projection, parallel to the *a*-axis, of stacked polar ap β-sheets with the glutamic acid side groups shaded in light gray. If glutamic acid residues were to penetrate the lattice, they can only do so in the inter-alanine layer. If they occur out of phase with the folding periodicity they systematically decorate all sites in the inter-alanine layer and would force every other pair of ap β-sheets apart. The minimum *b* value would be greater than the measured values.

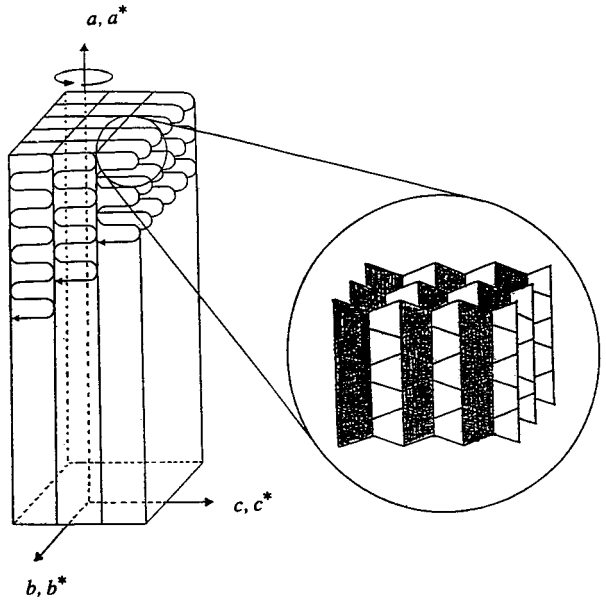

Figure 9. Schematic illustration of the basic crystalline entity giving rise to the X-ray diffraction and spectroscopic results. The long axis of the roughly square cross-sectioned rod is parallel to the crystallographic *a*-axis or hydrogen bond direction. The chain-folded lamellae stack in the *c* direction.

$(AG)_6EG$, as the proportion of glutamic acid residues is reduced. No such improvement is noticed.

Although the analysis of the X-ray diffraction results provides considerable support for an in-phase relationship between chain-folding periodicity and amino acid sequence, so as to keep the glutamic acid residues at the lamellar fold surfaces, no direct evidence of fold surface confinement of glutamic acid has yet been obtained. Experiments are being designed to selectively locate the glutamic acid residues within the crystalline architecture.

The model of the crystalline entity that is the basis for the diffraction studies is shown in Figure 9. It is a semicrystalline rod, of roughly square cross section, and is constructed from a stack of chain-folded lamellae, the overall thickness of which is similar to the dimension in the orthogonal *b* direction; the axis of the rod is along the *a*-axis hydrogen-bond direction.

Now that we have a model structure for these polypeptides, it is worthwhile, and instructive in terms of future strategy, to consider questions concerning the interplay of the intermolecular forces and protein chain geometry in these structures. For example: Is it the overwhelming hydrophobic interactions of the methyl side groups of alanyl residues, forming a set of hydrophobic sandwiches by pairing in contact, that force a polar architecture on the ap β-sheets, and, as a consequence, demand γ-turns for these amino acid sequences? Or, is it that only the γ-turns will allow a chain-folded architecture of this type to be generated systematically? In our original strategy, we expected the polypeptide to fold via β-turns. The β-turn has been a common element in globular protein structures[37] and, at first glance, is not too different from the proven bend[38] in nylon 46 which forms similar regular chain-folded crystals.[9a] However, more careful consideration reveals that β-turns would encounter difficulties in an adjacent reentry chain-folded crystal structure; this is because the two chain segments that emanate from the β-turn in globular proteins prefer to run inclined to each other. Even in globular protein structures that incorporate a few consecutive β-turns, the patch of ap β-sheet formed is distorted—twisted in its plane—and would not be expected to stack comfortably.[39] However, as the interior of the lamella develops with stacking of successive sheets, the resulting crystalline lattice forces will be expected to dominate, leading to distortion of the turn. If, as a result, β-turns are not compatible with the overall structure, the chains will fold through γ-turns. Recent structure surveys indicate that the γ-turn is the most common sharp fold in protein structures.[27] The consequence of incorporating γ-turns into our structures is to produce polar ap β-sheets, as dicussed previously.

To help answer these questions, and others, different amino acid sequences have been synthesized and crystallized and are in the final stages of analysis, for example, where the glutamic acid residues have been replaced by a variety of natural and artificial amino acids. Also more subtle changes have been investigated, as in $poly(AG)_3EG(GA)_3EG$, where successive AG segments are sequence-reversed. These results, and others, will be published elsewhere and will add to the debate about the interplay of intermolecular forces controlling the folding pattern in crystalline polymers and protein structures.

Conclusions

The family of $poly(AG)_xEG$ polypeptides, where the integer x varies from 3 to 6, can be crystallized by controlled precipitation from aqueous formic acid solutions. Analysis via X-ray diffraction and spectroscopic techniques supports an ap β-sheet architecture within the crystals. The X-ray diffraction results allow us to go further, enabling an analysis of alternative three-dimensional models and selection of the structure(s) giving the best fit with the experimental data. Line broadening of the wide-angle X-ray diffraction signals associated with the chain direction and the appearance of one-dimensional particle-interference peaks favor chain-folded lamellae. The thickness and packing periodicity are approximately 3.6 nm for $poly(AG)_3EG$, and are somewhat larger for $x = 4-6$.

The lamellar thickness, for the whole family of these polymers, is always $\leq$28 times the chain length. Thus, the chains fold back at the lamellar surfaces and reenter the crystalline lamellae at adjacent lattice sites, as generally accepted for sufficiently flexible polymers.[24] A direct consequence of this adjacent reentry mecha-

nism is the creation of ap β-sheets within the lamellar crystal as demanded by the X-ray diffraction and spectroscopic analyses.

The crystalline core of the lamellar crystals for poly-$(AG)_3EG$ is constructed from an orthorhombic unit cell with the following parameters: $a = 0.948$ nm (hydrogen-bond direction), $b = 1.060$ nm (ap β-sheet stacking direction) and $c = 0.695$ nm (chain-axis direction). This unit cell is closely related to the ap β-sheet structures of polyAG and β-silk fibroins. The unit cells of the other members of the family are essentially the same, except for a progressive decrease in the value of b as the number of alanylglycyl diads in the repetitive amino acid sequence increases. We interpret this result in terms of a closer stacking of the ap β-sheets as the crystalline lamellar core becomes thicker.

X-ray diffraction structure analyses and refinements favor a structure composed of polar ap β-sheets, one surface decorated with methyl side groups and the other with hydrogen atoms. The ap β-sheets stack with like surfaces in contact. This creates an alternation in sheet spacing, similar to that proposed for polyAG, and accounts for the observed strong 010 diffraction signal. The structure demands γ-turns (three amino acid turns) and rules out β-turns. The evidence favors a chain-folding periodicity in-phase with the repetitive amino acid sequence; thus, we propose that the glutamic acid residues occur at the folds and decorate the crystalline lamellar surfaces.

Acknowledgment. This investigation was supported by grants from the Polymers and Genetics Programs of the National Science Foundation (NSF), the NSF Materials Research Science and Engineering Center at the University of Massachusetts, the Engineering and Physical Sciences Research Council, and the Biotechnology and Biological Sciences Council (Bristol University Molecular Recognition Center). We thank Joseph Cappello of Protein Polymer Technologies, Inc., for a gift of one of the plasmids used in cloning and expression, and for helpful discussions of methods of gene design and construction. NMR spectra were acquired with the assistance of L. K. Thompson.

Appendix: Texture of Samples Prepared as Mats[40]

As stated in the main paper, the crystalline samples of the family of periodic polypeptides that were used to obtain the textured X-ray diffraction photographs (Figures 1A,F, and 2) were prepared by compressing and drying swollen gels into mats. Only two types of X-ray diffraction fingerprint are observed from these mats: (1) when the X-ray beam is directed orthogonal to the mat surface, diffraction rings are obtained (e.g., Figure 1E); (2) with the X-ray beam in any direction parallel to the mat, surface-oriented diffraction patterns are obtained (e.g., Figure 1A). The overall picture is illustrated in Figure 10. Thus, this method of specimen preparation, necessary because of the relatively small quantities of crystalline material available, produces a textured structure.

The texture can be described by considering the pole figure diagrams shown in Figure 11. The lattice is orthogonal so that each pair of reciprocal, real axes—a^*,a; b^*,b; c^*,c—are coincident and mutually orthogonal. In Figure 11 they have been arranged to match the oriented X-ray patterns shown in Figures 1A,B and 2. The set of three pole figure diagrams in Figure 11a

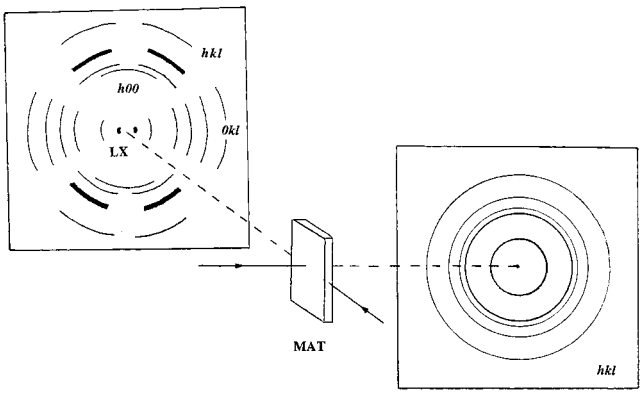

Figure 10. Diagram illustrating the relationship between the crystalline mat sample and the two types of X-ray diffraction pattern observed. LX represents the observed low angle diffraction peak.

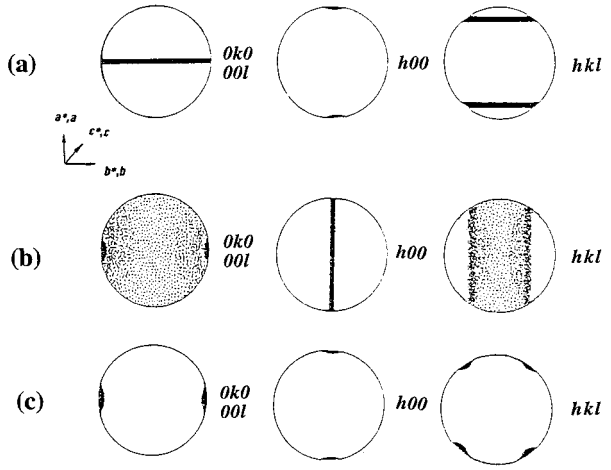

Figure 11. Pole figure diagrams corresponding to the mat texture shown in Figure 1A. (a) Projection of poles with crystallites cylindrically averaged about a^*,a-axis. The black bars are projections of circles in planes perpendicular to a^*,a direction. (b) The same poles as in part a cylindrically averaged about the b^*,b direction. The graduated shading is meant to give an indication of the relative distribution of poles on this set of three-dimensional pole figure diagrams. (c) Sections of these pole distributions in the diametric plane (in the plane of the paper). These are the pole figure sections that correspond to the X-ray diffraction patterns obtained when the X-ray beam is directed parallel to the surface of the mat. X-ray patterns, with the incident X-ray beam orthogonal to the mat surface, i.e., orthogonal to the a^*,a-axis, will give diffraction rings consistent with the experimental observations.

illustrates the behavior of the 0k0/00l, h00, and hkl poles, respectively, when there is cylindrical averaging about the a^*,a direction. In this case, the 0k0/00l poles are swept into equatorial circles; the h00 poles are confined to "polar" caps, and hkl poles trace out latitude circles. Figure 11b shows the effect when additional cylindrical averaging takes place about the b^*,b direction. Although the 0k0/00l poles are distributed over the whole polar diagram, they are concentrated at the extrema of the horizontal diameter; the h00 poles trace out a "polar" great circle, and the hkl poles are distributed within the band shown but with a pronounced concentration at the edges. The diametric plane sections of these pole distributions (in the plane of the paper) are shown in Figure 11c; they are equivalent to the sections of diffraction space captured in the X-ray photographs taken with the X-ray beam parallel to the mat surface (Figures 1A and 2). In contrast, the X-ray diffraction pattern of this texture, taken with the

incident beam orthogonal to the mat surface, is a set of concentric rings (imagine vertical sections of Figure 11b perpendicular to the paper) as illustrated in Figure 1E. The interplanar spacings will, of course, match those of the powder X-ray diffraction photograph (Figure 1E), but the relative intensities will not be the same; this difference will be most noticeable for the intrinsically strong 0*k*0 diffraction signals. Thus, this texture, with cylindrical averaging about the *a**,*a*-axis coupled with cylindrical averaging about a direction orthogonal to the *a**,*a*-axis, accounts for the X-ray experimental results.

Allowance has to be made for texture in the computation of the relative intensities of the diffraction signals. In the SXDPs, derived using the Cerius 2 software system, only fiber and powder textures can be accommodated; thus, modifications were made to compensate for the additional textural features that we have in our crystalline mats. These modifications are, by their very nature, a respectable approximation and not an exact calculation. Detailed quantitative comparison between calculated and measured relative intensities is therefore inappropriate.

References and Notes

(1) See, for example: Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680, and references therein.

(2) (a) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727. (b) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427. (c) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471. (d) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213. (e) Deguchi, Y.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Macromol. Sci., Pure Appl. Chem.* **1994**, *A31*, 1691.

(3) Keller, A. *Philos. Mag.* **1957**, *2*, 1171. Till, P. H. *J. Polym. Sci.* **1957**, *24*, 301. Fischer, E. W. *Z. Naturforsch., A* **1957**, *12*, 753.

(4) Keith, H. D.; Paddon, F. J.; Giannoni, G. *J. Mol. Biol.* **1969**, *43*, 423.

(5) Keith, H. D.; Giannoni, G.; Paddon, F. J. *Biopolmers* **1969**, *7*, 775.

(6) Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 195.

(7) Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 201.

(8) Fraser, R. D. B.; MacRae, T. P. *Conformation in Fibrous Proteins*; Academic Press: New York, 1973.

(9) (a) Nylons with amide in folds. Nylon 4,6: Atkins, E. D. T.; Hill, M. J.; Hong, S. K.; Keller, A.; Organ, S. J. *Macromolecules* **1992**, *25*, 917. Nylon 4: Bellinger, M. A.; Waddon, A. J.; Atkins, E. D. T.; MacKnight, W. J. *Macromolecules* **1994**, *27*, 2130. (b) Nylons with polymethylene segments in folds: Atkins, E. D. T.; Keller, A.; Sadler, D. M. *J. Polym. Sci.* **1972**, *A2*, 863. Hill, M. J.; Atkins, E. D. T. *Macromolecules* **1995**, *28*, 604. Atkins, E. D. T.; Hill, M. J.; Velluraja, K. *Polymer* **1995**, *36*, 35.

(10) Single-letter amino acid notation: A = alanine ($-CH_3$ side group), G = glycine ($-H$ side group), E = glutamic acid ($-CH_2CH_2COOH$ side group).

(11) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C. *J. Mol. Biol.* **1966**, *19*, 508.

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

(13) Walsby, A. E. *Sci. Am.* **1977**, *237*, 90. Chain-folded structure: Atkins, E. D. T. Inaugural Lecture, University of Bristol, U.K., 1983.

(14) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211.

(15) Fraser, R. D. B.; MacRae, T. P.; Steward, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706.

(16) Itoh, K.; Foxman, B. M.; Fasman, G. D. *Biopolymers* **1976**, *15*, 419.

(17) Ishida, M.; Asakura, T.; Yokoi, M.; Saito, H. *Macromolecules* **1990**, *23*, 88.

(18) Warwicker, J. O. *J. Mol. Biol.* **1960**, *2*, 350.

(19) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729.

(20) Takahashi, Y.; Gehoh, M.; Yuzuriha, K. *J. Polym. Sci., Part B: Polym. Phys.* **1991**, *29*, 889.

(21) Calstrom, D. *J. Biophys. Biochem. Cytol.* **1957**, *3*, 669.

(22) Lucas, F.; Shaw, J. T. B.; Smith, S. G. *J. Mol. Biol.* **1960**, *2*, 339.

(23) Geil, P. H. Polymer Single Crystals; Interscience: New York, 1963. Keller, A. *Rep. Prog. Phys.* **1968**, *31*, 624.

(24) The appearance of first layer−line diffraction signals and an excessive number of low-angle diffraction orders (the lamellae were positioned on a one-dimensional superlattice, parallel to *c*) should be ignored in this instance since these features can be removed by adjustments to the model as explained later in the case for the γ-turn structure. Since we abandon the β-turn model as a consequence of more fundamental discrepancies with the experimental evidence, there is little point in describing these modifications, although in practice we investigated variations in this type of structure.

(25) Nemethy, G.; Printz, M. P. *Macromolecules* **1972**, *5*, 755.

(26) A recent survey of globular protein structures suggests that γ-turns are the most common: Blundell, T. L. Private communication.

(27) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2483.

(28) Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712.

(29) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1968**, *38*, 181.

(30) Frushour, B. G.; Koenig, J. L. *Biopolymers* **1975**, *14*, 2115.

(31) Frushour, B. G.; Painter, P. C.; Koenig, J. L. *J. Macromol. Sci. Rev. Macromol. Chem.* **1976**, *C15*, 29.

(32) Marsh, R. E.; Corey, R. E.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1.

(33) A mismatch of one amino acid would create apolar ap β-sheets, which we have already ruled out on the basis of the observed intensity of the 010 diffraction signal. We judge that to change the lamellar thickness by ±19.3% for poly-$(AG)_3EG$ is excessive, but since the change for poly$(AG)_6EG$ is ±12.2%, we will discuss the consequences of the folding periodicity being out-of-phase with the amino acid sequence.

(34) In adjacent reentry chain-folded lamellae, the straightforward way to relieve overcrowding at the fold surfaces is to allow contiguous sheets to progressively shear, adjacent chains in a sheet to progressively shear, or both; thus, the chains would run oblique to the lamellar normal. This can only be accomplished in the case of the distinctly puckered ap β-sheets if the shears are multiples of 0.695 nm.

(35) Fraser *et al.*[15] do not mention chain-folding in their structural investigation of polyAG, but, in view of the considerations discussed in the Appendix, there is every likelihood it existed in their samples. Because they did not notice any low-angle intensity it is reasonable to assume that if it exists, the peak is hidden behind the beam stop; this implies that the straight stems are long and therefore allow closer inter/sheet stacking. Irrespective of the length of the stems, the chain-folded ap β-sheets of polyAG would stack closer than those of the polypeptides investigated herein since there would be no bulky groups in the folds.

(36) To imagine that as the middle ap β-sheet moves closer to the *b*/2 position the relative intensity of the 010 diffraction signal would drop in comparison with the 020, (as might be erroneously supposed in casually comparing Figure 1A (poly-$(AG)_3EG$) with Figure 2C (poly$(AG)_6EG$)) neglects the additional scattering of the off-set layer of methyl side groups and proves to be an incorrect simplification.

(37) Sibanda, B. L.; Blundell, T. L.; Thorton, J. M. *J. Mol. Biol.* **1989**, *206*, 759.

(38) The amide-in-the-bend folds[9a] for nylon 46 are not exactly the same as β-turns, since they have *two* pairs of CH_2 units either side of the $-NH-CO-$ group; this extra flexibility allows the in-coming and out-going stems to remain parallel.

(39) Extensively stacked ap β-sheets do not occur in globular protein structures and so a β-turn/twisted β-sheet is perfectly satisfactory. In those structures where a few ap β-sheets are associated together there is no indication of sheet alignment or register.

(40) Appendix prepared by E.D.T.A.

MA9614050