J. Biochem. 122, 217-225 (1997)

Effects of Amino Acid Side-Chain Volume on Chain Packing in Genetically Engineered Periodic Polypeptides[1]

Eric J. Cantor,* Edward D.T. Atkins,† Sharon J. Cooper,† Maurille J. Fournier,* Thomas L. Mason,* and David A. Tirrell‡,[2]

*Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, MA 01003, USA; †H.H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK; and ‡Department of Polymer Science and Engineering, University of Massachusetts, Amherst, MA 01003, USA

Received for publication, March 17, 1997

The fidelity of bacterial protein synthesis allows the production of architecturally well-defined polymeric materials through precise control of chain length, sequence, stereochemistry, and interchain interactions. In the present paper, we examine the relation between amino acid residue volume and crystalline unit cell dimensions, in a set of periodic protein polymers of repeating unit sequence -(AlaGly)$_3$-X-Gly-, where X is Asn, Phe, Ser, Val, or Tyr. The proteins were overexpressed in *Escherichia coli*, purified by simple procedures based on acid/ethanol precipitation or insolubility in aqueous sodium dodecyl sulfate, and processed to form oriented crystalline mats by precipitation from formic acid under mechanical shear. X-ray diffraction analyses revealed that the basic structures of the -(AlaGly)$_3$-X-Gly- polymers are identical to that previously reported for [(AlaGly)$_3$-GluGly]$_{36}$, [Krejchi, M.T., Atkins, E.D.T., Waddon, A.J., Fournier, M.J., Mason, T.L., and Tirrell, D.A. (1994) *Science* 265, 1427-1432], with the oligoalanylglycine segments forming antiparallel β-sheets and the substituted amino acids occurring within three-residue folds at the lamellar surfaces. The X-ray diffraction signals for each member of the family index on an orthorhombic unit cell; the *a*-axis (hydrogen bond direction) and *c*-axis (chain direction) spacings remain invariant but the *b*-axis (sheet stacking direction) spacing increases with increasing volume of the substituted amino acid. The results obtained from a variant with alternating Glu and Lys substitution at the X position, together with the results previously reported for poly(L-alanylglycine) [Panitch, A., Matsuki, K., Cantor, E.J., Cooper, S.J., Atkins, E.D.T., Fournier, M.J., Mason, T.L., and Tirrell, D.A. (1997) *Macromolecules* 30, 42-49] are included for comparison. The average intersheet stacking distance ($b/2$) increases linearly with the volume of the amino acid inserted at position X. Because the chain-folded lamellar architecture adopted by these periodic polypeptides accommodates a wide range of residues differing in charge, steric bulk, and hydrophobicity, these results illustrate a new approach to the engineering of intermolecular interactions in polymeric solids.

Key words: artificial proteins, biomaterials, crystallization, fibrous proteins, X-ray diffraction.

Recombinant DNA methodology provides a new approach to the synthesis of well-defined macromolecular materials. Because biological protein synthesis allows precise control of critical molecular parameters such as chemical composition, sequence, stereochemistry, and molecular weight, protein-like polymers can be produced with pre-determined structures and predictable physical properties.

Several genetically engineered periodic polypeptides have been designed in our laboratories to adopt chain-folded lamellar architectures in the solid state. The design of these polypeptides, based on previous work on silk and silk analogs (*1-3*) and on the database of reverse turn sequences in globular proteins (*4-6*), features repetitive dyads of the amino acids alanine and glycine to form extended β-sheets, and regularly interspersed bulky, polar residues (*e.g.*, glutamic acid) to aid in chain reversal (*7*). The larger steric size of glutamic acid, compared to alanine and glycine, was envisioned to obstruct inclusion of the acidic side chains into the interior of the lamellae, and to cause segregation to the lamellar surface during crystal growth. Therefore, the predicted lamellar structure would have the hydrophilic polar moiety on the lamellar surface and a relatively hydrophobic stem region formed exclusively by β-strands.

[1] 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).

[2] To whom correspondence should be addressed. Fax: +1-413-577-1610, E-mail: tirrell@polysci.umass.edu

Abbreviations: IPTG, isopropyl β-D-thiogalactopyranoside; DNaseI, deoxyribonuclease I from bovine pancreas.

These design goals were met with the production of $[(AlaGly)_3GluGly]_{36}$, an artificial protein with an eight residue repetitive unit and glutamic acid in the proposed turn. Vibrational spectroscopy, nuclear magnetic resonance spectroscopy, and wide-angle X-ray diffraction have confirmed the architecture of this material (*7*) and recent infrared experiments have demonstrated that the chain reversal of a related protein polymer, $[(AlaGly)_3GluGly-(GlyAla)_3GluGly]_{10}$, follows an adjacent re-entry model (*8*). Subsequent structural analysis of $-(AlaGly)_nGluGly-$ proteins ($n=4$, 5, 6) with increased stem lengths showed that all of these materials are characterized by an adjacent re-entry, chain-folded lamellar structure composed of antiparallel β-sheets with polar orientation and three-residue turns. Indirect evidence has been obtained for in-phase folding periodicity and for confinement of Glu to the lamellar surfaces (*9*).

Here we examine how variation in the size and charge of the residues in the putative turn positions affects the chain-folded geometry and the intersheet packing distance in the lamellae. Substitution of different amino acids at the proposed turn is not expected to perturb the formation of the chain-folded lamellae, but the intersheet packing dimension would be expected to be sensitive to the volume of the substituted residue, as in other β-form materials. For example, the intersheet stacking periodicity for the β-form of polyglycine (*10*) has been reported as 0.34 nm, whereas polyglutamic acid (*11*) and the silk fibroin (*12*) from *Nephila senegalensis* (which contains a large proportion of bulky amino acids) are characterized by intersheet spacings of 0.78 and 0.79 nm, respectively. To test this possibility, a series of polypeptides of repeating unit sequence $(AlaGly)_3$-X-Gly, with single residue substitutions at the putative turn positions, was produced and analyzed for structural changes in the chain-folded lamellae. The amino acids chosen for substitution [X= Asn, Phe, Ser, Val, and Tyr] represent a broad range of residue volumes as well as hydrophobicity and chemical functionality. The analysis provides a critical test for chain packing as a function of residue volume, and yields a linear relation between the volume of residues at the lamellar surface and the intersheet distance in crystalline $(AlaGly)_3$-X-Gly polymers.

MATERIALS AND METHODS

Protein polymers were produced in *Escherichia coli* using the T7 expression system developed by Studier (*13*). Artificial genes encoding the protein polymers were inserted behind the T7 promoter of a pET expression plasmid and transformed into *E. coli* strain BL21(DE3)pLysS, which carries the gene for T7 RNA polymerase regulated by the *lacUV5* promoter. In these cells, induction of T7 RNA polymerase expression with IPTG activates expression of the T7-regulated gene for the protein polymer which accumulates in the cytoplasm of the bacteria. The polymers were purified from cell lysates for solid state structural analysis.

Construction of DNA Monomers—DNA "monomers" encoding two repeats of the $(AlaGly)_3AsnGly$ sequence (*14*) or one repeat of the corresponding Phe, Ser, Val, or Tyr variants (*15*) were formed from synthetic oligonucleotides, ligated into the pUC 18 vector and amplified in *E. coli* strain DH5αF′. Plasmid DNA was isolated and sequenced to verify the nucleotide sequence of the monomer. Plasmid DNA was digested with *Ban*I to release the monomer which was then separated from vector fragments by electrophoresis in an 8% acrylamide gel and recovered by "crush and soak" elution (*16*).

Construction of Expression Plasmids—DNA monomers with non-palindromic *Ban*I ends were self-ligated head-to-tail to form multimers for subsequent ligation into the unique *Ban*I restriction sites of the cloning vectors p937.51 (*17*; Asn variant) or pECL.1 (Phe, Ser, Val, and Tyr variants). The resulting plasmids were used to transform *E. coli* strain TOP 10F′ (Invitrogen). Plasmid DNA from individual transformants was isolated and digested with *Bam*HI to determine the sizes of the multimers. Plasmids were selected for further analysis primarily on the basis of insert size. The selected plasmids with their repetitive insert sizes and predicted numbers of repetitive segments are: p937.51-GN8 [0.5 kb, 16 repeats of $(AlaGly)_3AsnGly$]; pECL.1-GF9-2 [0.4 kb, 13 repeats of $(AlaGly)_3PheGly$]; pECL.1-GS2-16 [1.3 kb, 54 repeats of $(AlaGly)_3SerGly$]; pECL.1-GV8 [1 kb, 41 repeats of $(AlaGly)_3ValGly$]; and pECL.1-GS2aGY [1.4 kb, 59 repeats of $(AlaGly)_3TyrGly$]. The multimeric DNA sequences were subcloned as *Bam*HI fragments into the pET-3b expression vector (*13*), generating plasmids p3bGN8, p3bGF9-2, p3bGS216-1b, p3bGV8-3, and p3bGS2-1aGY, respectively. The plasmids were transformed into *E. coli* strain BL21(DE3)pLysS (*13*) for protein expression.

Protein Expression and Purification—Expression of the Asn polymer was detected by *in vivo* labeling as well as immunoblot analysis using the T7-Tag (Novagen) antibody (Cantor *et al.*, unpublished observations). Immunoblot analysis of lysates from small-scale expression experiments was used to document synthesis and determine optimal times for induction for the Phe, Ser, Val, and Tyr polymers (Cantor *et al.*, unpublished observations).

The Asn polymer was produced in 12-liter fermentations in YT media containing ampicillin (50 μg/ml) and chloramphenicol (25 μg/ml) as described by McGrath (*14*). The recombinant protein was purified from the soluble fraction of the cell lysate by acidification with acetic acid in three steps to pH 6.0, 5.0, and 4.0. Contaminating proteins precipitated at each pH drop were removed by centrifugation. At pH 4.0 ethanol was added to 40% (v/v) and additional precipitated material was removed. The target protein was precipitated by increasing the ethanol concentration to 80% (v/v) and incubating at -20°C. The protein was collected by centrifugation, washed extensively in double-distilled H_2O and lyophilized. The yield of the target protein was approximately 40 mg/liter.

The repetitive portion of the $(AlaGly)_3AsnGly$ polypeptide was separated from the N- and C-terminal flanking sequences by cleavage with CNBr. The fusion protein was dissolved in 70% formic acid at a concentration of 20 mg/ml and a 50-fold molar excess of CNBr was added. The reaction vessel was purged with nitrogen and stirred in the dark at room temperature for 24 h. Solvent was removed by vacuum evaporation. The remaining film was washed with double-distilled H_2O to remove the soluble N- and C-terminal peptide fragments. The insoluble repetitive segment was collected by centrifugation and washed several more times with H_2O. After the supernatant reached

pH 7, the protein was lyophilized. Amino acid and combustion analyses were consistent with the expected composition of the protein. Calculated: glycine, 49.2 mol%; alanine, 37.1 mol%; asparagine, 12.1 mol%. Observed: glycine, 49.7 mol%; alanine, 37.1 mol%; asparagine, 12.5 mol%. Calculated (includes 3% H_2O): C, 44.0%; H, 6.2%; N, 22.0%. Observed: C, 42.4%; H, 5.6%; N, 20.4%.

The Phe, Ser, Val, and Tyr polymers were produced in 12-liter batch fermentations in LB media at 37°C. Cells were induced for protein expression for 3 h after the addition of IPTG, and then harvested by centrifugation, resuspended in 200 ml double-distilled H_2O, and frozen at −20°C. The presence of T7 lysozyme in the BL21(DE3) pLysS cells caused lysis when the cell suspension was thawed at 37°C with occasional mixing. DNase I (20 μg/ml) and phenylmethylsulfonyl fluoride (1 mM) were added to the frozen cell suspension to reduce viscosity and prevent protein degradation. The lysates were centrifuged for 20 min at 10,000 × *g* to pellet the insoluble material containing the protein polymers. The pellets were resuspended in 1% Triton X-100 and rocked at room temperature for 1 h. The Triton-insoluble material was recovered by centrifugation and the wash procedure was repeated with 5% sodium dodecyl sulfate. These washes removed essentially all contaminants from the sodium dodecyl sulfate-insoluble protein polymers, which were then washed extensively with double-distilled H_2O and lyophilized. The yield of each polymer was between 35 and 40 mg/liter. The repetitive portions of the polymers were separated from the N- and C-terminal flanking regions by CNBr cleavage as described for the Asn polymer. Amino acid and combustion analyses were consistent with the expected compositions. *Phe polymer*: calculated: glycine, 49.8 mol%; alanine, 37.4 mol%; phenylalanine, 12.4 mol%. Observed: glycine, 43.9 mol%; alanine, 38.9 mol%; phenylalanine, 11.8 mol%. Calculated (includes 3% H_2O): C, 51.5%; H, 6.3%, N, 18.5%. Observed: C, 48.7%; H, 6.3%; N, 15.5%. *Ser polymer*: calculated: glycine, 49.8 mol%; alanine, 37.4 mol%; serine, 12.4 mol%. Observed: glycine, 49.2 mol%; alanine, 37.1 mol%; serine, 10.9 mol%. Calculated (includes 3% H_2O): C, 44.0; H, 6.3, N, 20.5. Observed: C, 46.7; H, 6.7; N, 15.7. *Val polymer*: calculated: glycine, 49.8 mol%; alanine, 37.4 mol%; valine, 12.4 mol%. Observed: glycine, 43.3 mol%; alanine, 38.5 mol%; valine, 14.1 mol%. Calculated (includes 3% H_2O): C, 47.3%; H, 6.9%, N, 20.1%. Observed: C, 48.6%; H, 7.1%; N, 14.6%. *Tyr polymer*: calculated: glycine, 49.9 mol%; alanine, 37.5 mol%; tyrosine, 12.4 mol%. Observed: glycine, 46.2 mol%; alanine, 37.7 mol%; tyrosine, 12.7 mol%. Calculated (in-

TABLE I. **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 $[(AlaGly)_3AsnGly]_{16}$ (a=0.950, b=0.974, c=0.695 nm), together with an estimate of observed intensities (I_o).** [VS=very strong; S=strong; M=medium; W=weak.]

hkl	d_o	d_c	I_o
010	0.975	0.974	VS
020	0.487	0.487	S
021	0.407	0.399	M
030	0.326	0.325	M
031	0.294	0.294	M
022		0.282	
040		0.244	
200	0.475	0.475	M(sharp)
210	0.439	0.427	M
211	0.372	0.364	M(broad)
301	0.282	0.288	M
311		0.276	
400	0.241	0.238	W(sharp)
410	0.235	0.231	W

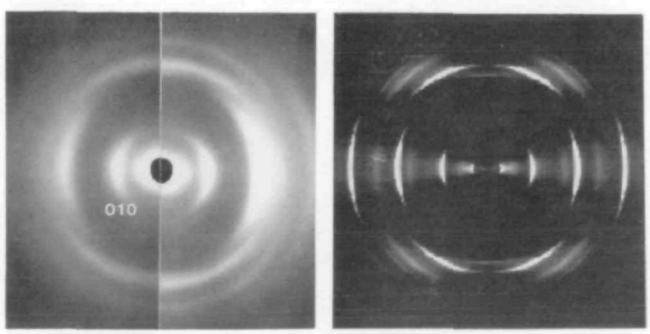

Fig. 1. **(a) X-ray diffraction photograph from $[(AlaGly)_3AsnGly]_{16}$ with the X-ray beam directed parallel to the plane of the compressed mat and with the mat normal horizontal.** Note the strong 010 signal on the horizontal bisector. The pattern is divided vertically into two halves; the right-hand side is a more exposed version of the left-hand side in order to show weaker diffraction signals. **(b) Simulated X-ray diffraction pattern (SXDP) of the crystalline lamellar structure proposed for $[(AlaGly)_3AsnGly]_{16}$ and shown in Fig. 3.**

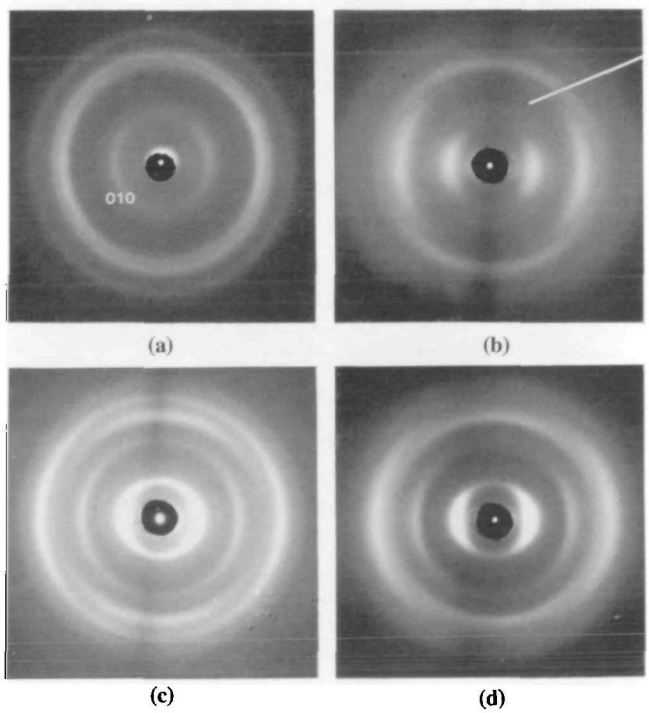

Fig. 2. **X-ray diffraction photographs obtained from compressed mats (using same arrangement as that for Fig. 1a) of: (a) $[(AlaGly)_3SerGly]_{54}$; (b) $[(AlaGly)_3ValGly]_{41}$; (c) $[(AlaGly)_3PheGly]_{13}$, and (d) $[(AlaGly)_3TyrGly]_{59}$.** Note that the 010 diffraction peak on the equator moves to lower diffraction angle in going through the series from serine to tyrosine. Weak first layer line diffraction is observed for $[(AlaGly)_3ValGly]_{41}$ (arrowed).

cludes 3% H_2O): C, 50.2%; H, 6.2%, N, 18.0%. Observed: C, 47.5%; H, 6.0%; N, 16.6%.

Orientation of -(AlaGly)$_3$-X-Gly- Polymers—For crystallization, the protein polymers were dissolved in 70% formic acid at a concentration of 20 mg/ml and then stirred at room temperature under methanol vapor. Each of the polymers precipitated in the form of a gel, which was recovered by centrifugation. After several washes in methanol, the protein was suspended in methanol and transferred to a 1 ml syringe barrel and allowed to settle onto Whatman 40 ashless filter paper. The resulting tablet of material was sandwiched between Teflon filter papers and compressed between glass plates secured by binder clamps. The apparatus was stored at room temperature for 16 h before the resulting mat was subjected to X-ray analysis. In some cases the mats were annealed at 121°C under steam for 30 min prior to structural analysis.

X-Ray Diffraction—X-ray diffraction photographs from compressed crystalline mats were obtained using a Statton-type X-ray camera, evacuated to backing pump pressure to reduce air scatter. The nickel-filtered CuKα sealed beam source was collimated with a system of 200 μm pinholes. The diffraction photographs were recorded on X-ray film

TABLE II. **Comparison of orthorhombic unit cell dimensions, in nanometers (errors ±0.002 nm), for the (AlaGly)$_3$-X-Gly family, in order of increasing volume of substituted amino acid.**

Substituted amino acid residue	*a* hydrogen-bond direction	*b* sheet stacking direction	*c* chain-axis direction
Alanine[a]	0.948	0.922	0.695
Serine	0.948	0.926	0.695
Asparagine	0.950	0.974	0.695
Valine	0.949	1.016	0.695
Glutamic acid[b]	0.948	1.060	0.695
Glutamic acid & Lysine (alternating)[c]	0.949	1.111	0.695
Phenylalanine	0.948	1.188	0.695
Tyrosine	0.948	1.198	0.695

[a]For chain-folded poly(alanylglycine) (*3*); a *b*-value (0.887 nm), 3.8% lower, was reported for oriented films of poly(alanylglycine) (*2*). [b]Refs. *7* and *9*. [c]Cantor *et al.*, unpublished observations.

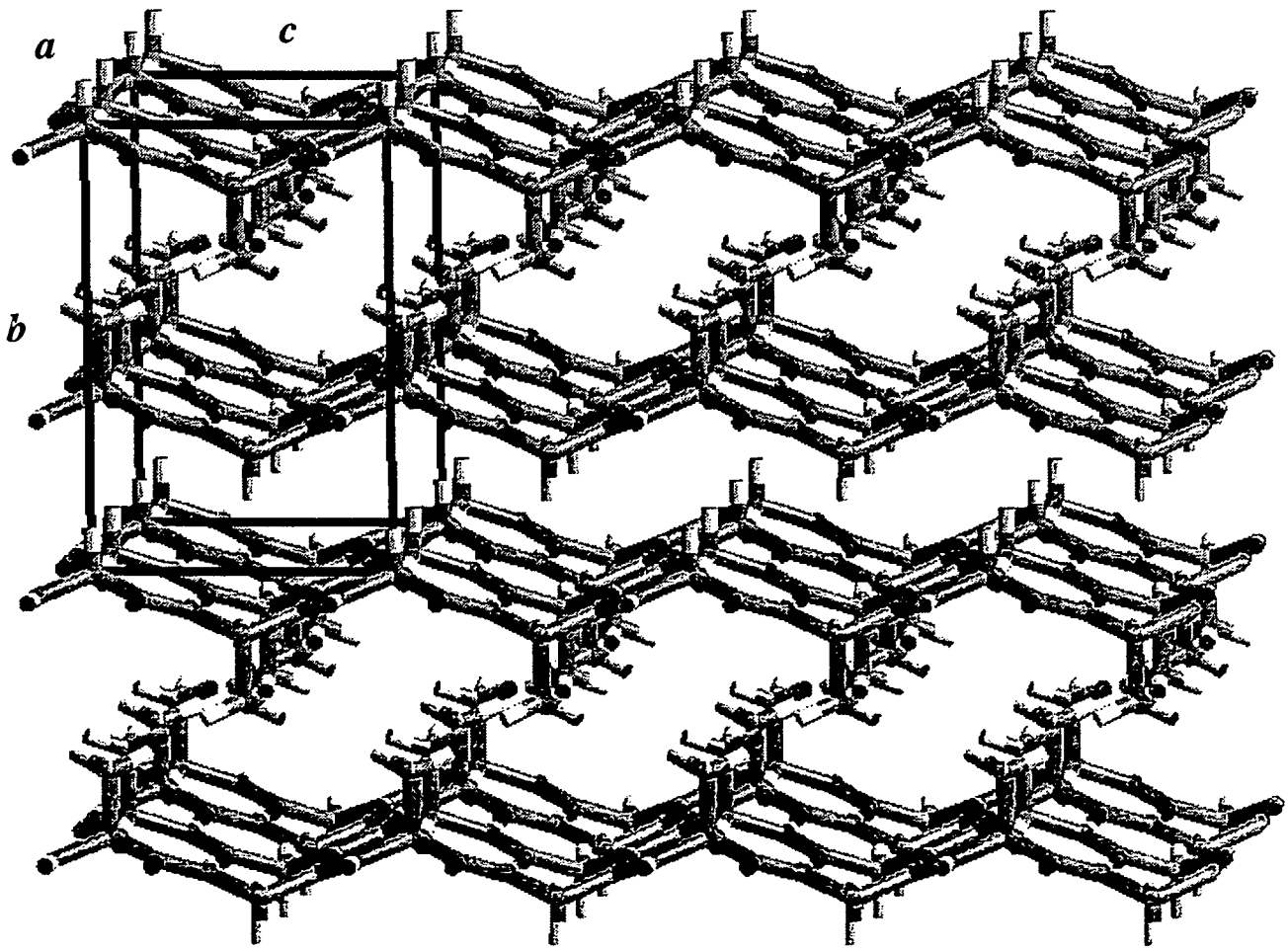

Fig. 3. **A computer-drawn projection of the stacked polar antiparallel β-sheets forming the interior of the [(AlaGly)$_3$AsnGly]$_{16}$ lamellae.** The orthorhombic unit cell is shown; *a* [oblique] = 0.950 nm (hydrogen bond direction), *b* [vertical] = 0.974 nm (sheet stacking direction), *c* [horizontal] = 0.695 nm (chain axis direction). Note that the sheets are stacked with like-surfaces together; thus, the intersheet distances are unequal and alternate in spacing. The top pair of sheets sandwich the hydrophobic alanyl groups and the second and third sheets down sandwich the glycyl groups. Examination of the portion of the center sheet within the unit cell shows that it is sheared parallel to the *a*-axis in the *ac* plane (actually ≈a/4).

with specimen-to-film distances in the range 40-50 mm.

Computer Modeling and Simulated X-Ray Diffraction Patterns—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). The simulated X-ray diffraction patterns were generated on a Silicon Graphics Indigo R4000 workstation using the Diffraction 1 module of the Cerius 2, release 1.6, program. The appropriate Lorentz and polarization correction factors are incorporated in the displayed intensities. The degree of arcing and the temperature factor were chosen to match the observed X-ray diffraction photographs.

RESULTS

X-Ray Diffraction from -(AlaGly)$_3$AsnGly-—The wide angle X-ray diffraction photograph shown in Fig. 1a was taken with the X-ray beam directed parallel to the plane of the compressed mat and with the mat normal horizontal. The 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 $a=0.950$, $b=0.974$, and $c=0.695$ nm. The measured and calculated interplanar spacings are listed for comparison in Table I together with an estimate of the observed intensities. The orientation direction is along a and the $h00$ diffraction signals appear as arcs centered on the meridian. The families of $0k0$ and $00l$ diffraction signals appear on the equator. This type of texture-oriented

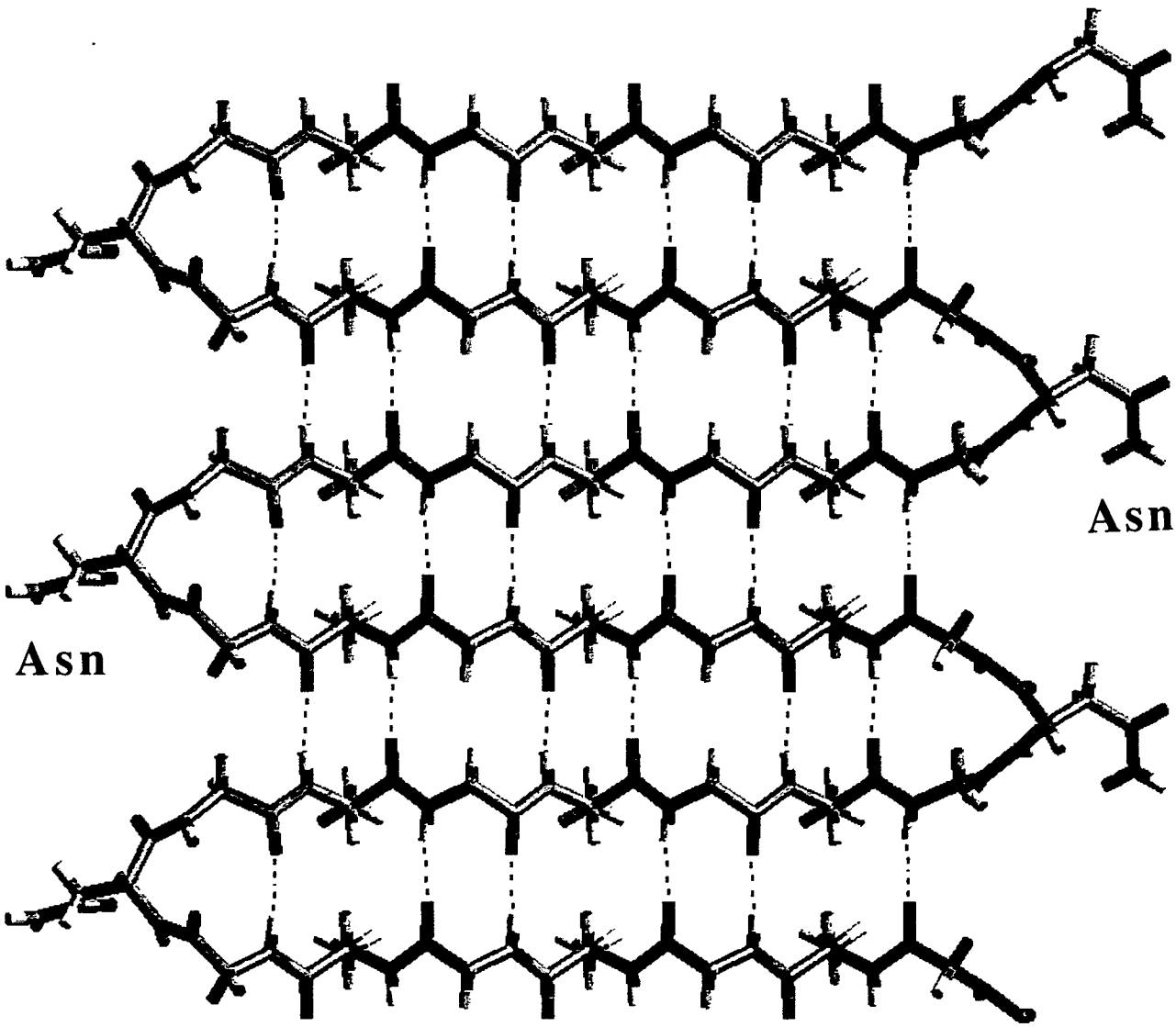

Fig. 4. **Computer drawn projection, orthogonal to the *ac* plane; with *c*-axis horizontal and *a*-axis vertical, of a single chain-folded antiparallel β-sheet of (AlaGly)$_3$AsnGly with GlyAsnGly turns.** Note that the conformations of the turns are different on left-hand and right-hand edges of the diagram.

X-ray diffraction pattern is similar to that reported for $[(AlaGly)_3GluGly]_{36}$ (*7, 9*).

We have discussed previously the line broadening of the various diffraction signals in patterns of this kind (*7, 9*). By such arguments, we conclude that the coherent scattering length in the *c* direction is small; we estimate a length of 2 to 4 nm allowing for instrumental line broadening. This is at least 16 times less than the length of the polypeptide, and demands chain folding in the lamellar aggregate.

The wide-angle X-ray photograph obtained with the X-ray beam directed orthogonal to the compressed mat exhibits a series of concentric rings that index on an orthorhombic lattice identical to that deduced from analysis of the oriented pattern shown in Fig. 1a. A detailed explanation using pole figure diagrams to account for this type of texture has been given previously (*9*).

X-Ray Diffraction from Other Members of the -(AlaGly)$_3$-X-Gly- Series—X-ray diffraction photographs taken with the X-ray beam directed parallel to the plane of the compressed mat and with the mat normal horizontal, for $[(AlaGly)_3SerGly]_{54}$, $[(AlaGly)_3ValGly]_{41}$, $[(AlaGly)_3PheGly]_{13}$, and $[(AlaGly)_3TyrGly]_{59}$ are shown in Fig. 2, a, b, c, and d, respectively. The general features are similar to those described for $[(AlaGly)_3AsnGly]_{16}$; however, there are noticeable differences. In particular, the prominent 010/020 pair of diffraction signals on the equator, representing the intersheet stacking distance, moves to lower angle (larger spacing) as the volume of the substituted amino acid increases. In all cases, the X-ray photographs obtained with the X-ray beam directed orthogonal to the compressed mat, exhibit series of concentric rings, indexing on the appropriate orthorhombic lattices.

The orthorhombic unit cell dimensions for the polypeptides already discussed, together with those for poly(L-alanylglycine) (*18*) and $[(AlaGly)_3GluGly]_{36}$ (*7, 9*), are listed in Table II in order of increasing volume (*19*) of the substituted amino acid.

DISCUSSION

The overall features of the X-ray diffraction patterns are similar for all five periodic polypeptides. The diffraction signals for each member of the family index on an orthorhombic unit cell; the *a*-axis (hydrogen bond direction) and *c*-axis (chain direction) spacings remain invariant but the *b*-axis (sheet stacking direction) spacing increases with increasing volume of the substituted amino acid. We have chosen to describe only $[(AlaGly)_3AsnGly]_{16}$ in detail, since it is centrally placed in terms of *b*-axis distance and amino acid residue size; the other members are described by comparison.

Analysis of X-Ray Diffraction Pattern from $[(AlaGly)_3AsnGly]_{16}$—The texture-oriented wide-angle X-ray diffraction pattern shown in Fig. 1a, is similar to that obtained for $[(AlaGly)_3GluGly]_{36}$; thus its interpretation is helped considerably by comparison with the published crystal structure of $[(AlaGly)_3GluGly]_{36}$ (*7, 9*). The salient features of this structure are as follows.

The chains fold repetitively through three-residue turns which are in-phase with the octapeptide repeat to form a

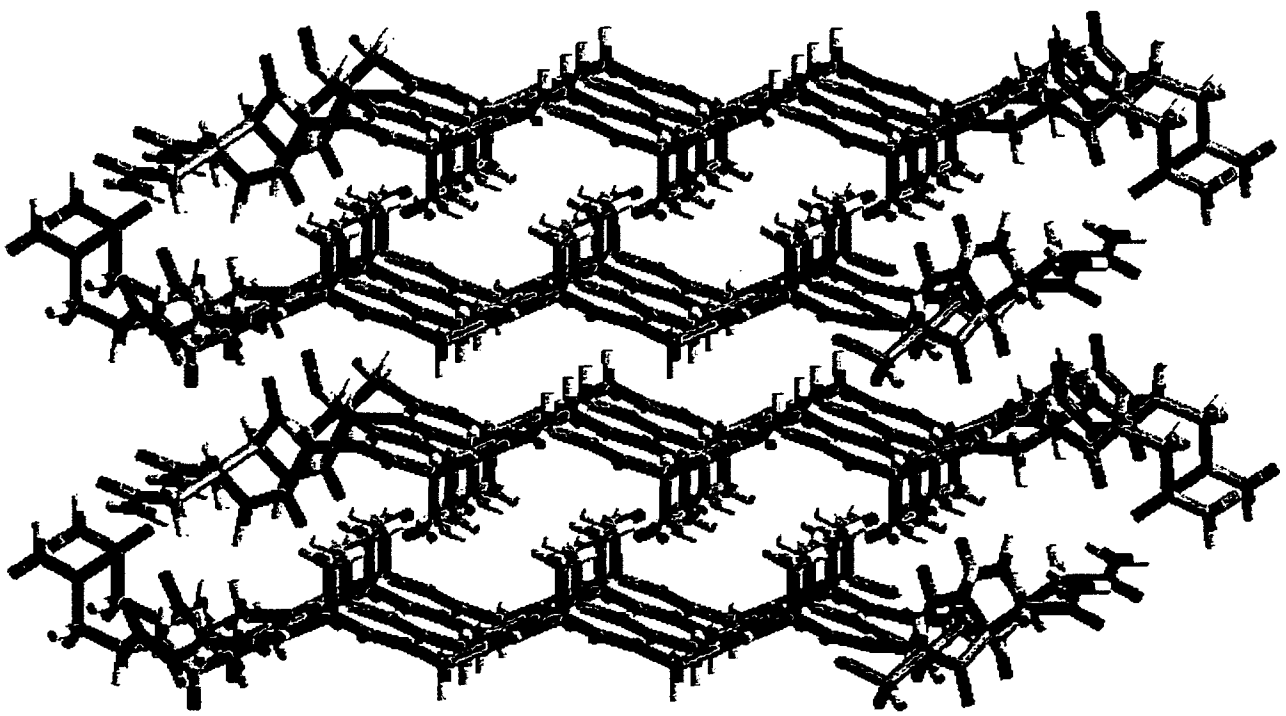

Fig. 5. **Computer drawn perspective view of the proposed structure for a chain-folded (AlaGly)$_3$AsnGly lamella.** This structure gave the best fit with the experimental X-ray diffraction results. Although both *ab* surfaces (left- and right-hand fold surfaces) of the lamellae are similar, it is instructive to note that successive turns (for example, on right-hand surface) have different conformations. This is because successive β-sheets are generated by a rotation of π about the *a*-axis.

polar antiparallel β-sheet with one surface covered with methyl groups and the other with hydrogen atoms. The sheets stack in pairs with like surfaces together, resulting in chain-folded crystalline lamellae that are approximately 3.6 nm in thickness. The diffraction signals index on an orthorhombic unit cell with dimensions $a=0.948$ nm (hydrogen-bond direction), $b=1.060$ nm (chain-stacking direction), and $c=0.695$ nm (chain-axis direction). Successive sheets are displaced in the *ac* plane.

The main difference between the structure of $[(AlaGly)_3GluGly]_{36}$ and that of $[(AlaGly)_3AsnGly]_{16}$ is a reduction in the sheet-stacking *b*-value by 8%; from 1.060 to 0.974 nm. The pairing of sheets is responsible for the occurrence of the strong 010 diffraction signal (*7, 9*); as indicated in Fig. 1. The three-dimensional structure of the stacked, polar antiparallel β-sheets, and the orthorhombic unit cell for $[(AlaGly)_3AsnGly]_{16}$ are shown in Fig. 3. Note the $\pm a/4$ shear of successive sheets in the *ac* plane.

An orthogonal view (parallel to the *b*-axis) of a single chain-folded sheet of $[(AlaGly)_3AsnGly]_{16}$ is shown in Fig. 4. The asparagine residues are placed at the apices of the three-residue turns. There are of course three possible sequences (incorporating asparagine) for the three-residue turns; namely GlyAsnGly, AlaGlyAsn, or AsnGlyAla. 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, the choice of the GlyAsnGly turn sequence as the basis for our model is arbitrary. Note that the folds on the right-hand edge of the sheet in Fig. 4 are conformationally different from those on the left. The sheets are stacked by making a rotation of π about the *a*-axis; thus, in the chain-folded lamellar crystal both surfaces are equally populated with the two types of fold conformation. The alternative mode of stacking (rotation of successive sheets by π about the *c*-axis) results in a chain-folded lamellar crystal with non-equivalent surfaces,

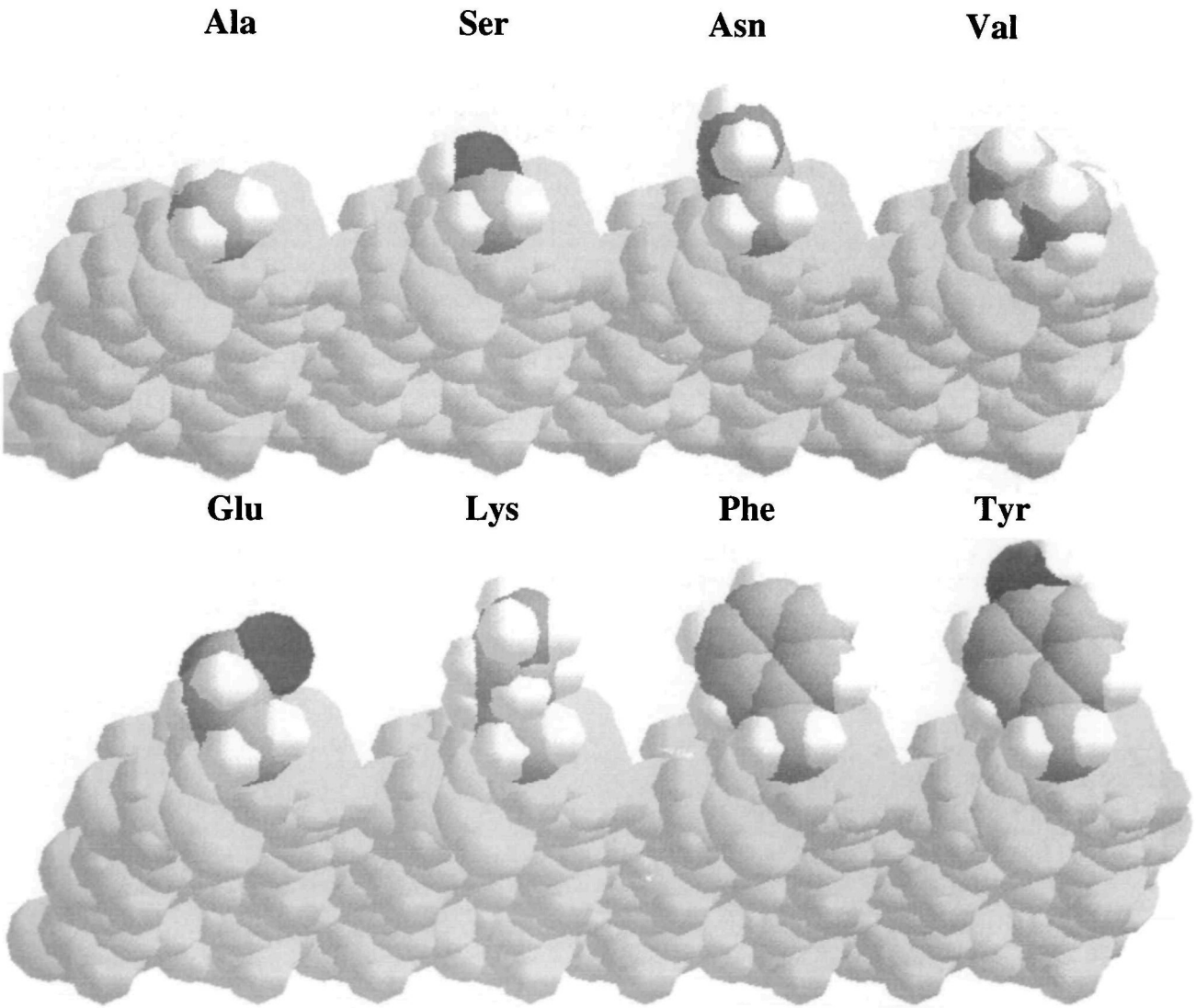

Fig. 6. **Perspective views of three-residue Gly-X-Gly turns decorated with X=alanine, serine, asparagine, valine, glutamic acid, lysine, phenylalanine, and tyrosine, respectively, illustrating the increasing size of the amino acid side group.** In an attempt to highlight the side groups of the substituted amino acids the other atoms in the turns are shaded in a uniform gray.

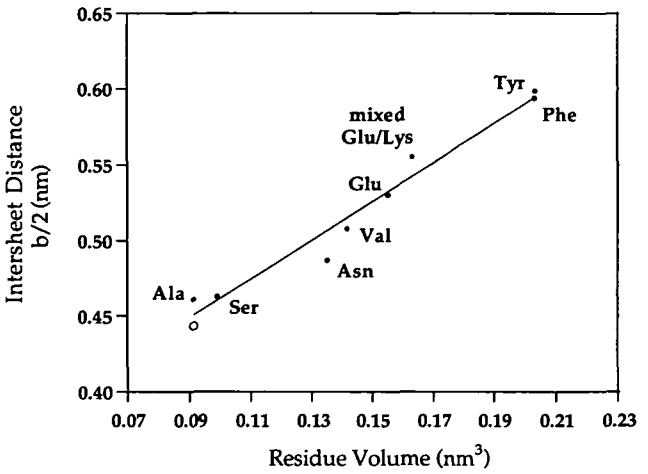

Fig. 7. **Correlation between the average intersheet distance ($b/2$) and volume (*9*) of substituted amino acids in the (AlaGly)$_3$-X-Gly series.** The single open circle on the bottom left is based on the value of $b/2$ reported for oriented films of poly(L-alanylglycine) (*2*) where no chain-folding was invoked, and therefore might be considered to be a lower limiting value for $b/2$. The value for the mixed Glu/Lys variant deserves some comment. The figure for the volume chosen is the average of those cited for Glu and Lys (0.155 and 0.171 nm^3, respectively). If the larger lysine residue were to preferentially dominate the intersheet stacking distance, then the plotted point would move horizontally to the right up to a maximum of 0.171 nm^3; almost on the drawn line.

and is thought to be less likely. These considerations have been discussed in more detail previously (*9*).

Structures for [(AlaGly)$_3$AsnGly]$_{16}$, based on this model, were computer-generated and the stereochemical feasibility of each was scrutinized using modeling procedures. We have chosen to display the comparison between experimental X-ray data, obtained in the form of texture-oriented X-ray diffraction photographs, and the calculated diffraction from the structures being tested for goodness of fit, in the form of simulated X-ray diffraction patterns (SXDPs). This makes comparison between the observed and calculated results most convenient and instructive. The structure shown in Fig. 5 represents the model that gave best agreement with the experimental X-ray data. Matching of the observed intensities required that the antiparallel β-sheets be randomly sheared by approximately $\pm a/4$ and $\pm c/2$ in the ac plane (Krejchi *et al.*, in press). The SXDP calculated from this structure is shown in Fig. 1b and can be compared with the experimental X-ray photograph in Fig. 1a.

Comparison of X-Ray Diffraction Results from Related Periodic Polypeptides—The X-ray diffraction pattern obtained from [(AlaGly)$_3$SerGly]$_{54}$ is shown in Fig. 2a. The overall features are similar to those of the pattern from [(AlaGly)$_3$AsnGly]$_{16}$, with two observable differences. First, the relative intensities of the 010 and 020 diffraction signals are different, with the 010 somewhat less intense for the serine variant. The observed changes in relative intensity cannot be accounted for by simply off-setting the central polar β-sheet further from the $b/2$ position in the unit cell, but instead reflect the relative contributions of the turn residues and the oligoalanylglycine segments to the observed scattering; as more strongly scattering groups are placed on the fold surface the relative intensity of the 010 diffraction peak increases. A similar trend was evident in our earlier work as the lamellar thickness was reduced in going from (AlaGly)$_6$GluGly to (AlaGly)$_3$GluGly (*9*); a process that effectively increases the dominance of the folds relative to the antiparallel β-sheet core region of the lamellae. Second, the spacing of the 010 diffraction signal on the equator is 4.9% less for the serine variant, which means that the average intersheet distance ($b/2$) is reduced by the same proportion. These features are the most obvious differences in a comparison of the X-ray diffraction patterns from this family of periodic polypeptides, and are illustrated by the four patterns shown in Fig. 2. The weak first layer line diffraction observed for [(AlaGly)$_3$ValGly]$_{41}$ indicates that the approximate $a/4$ shears in the ac plane are not completely random in this case (*9, 20*). Furthermore, the X-ray results for the related polymers, poly(L-alanylglycine) (*18*), [(AlaGly)$_3$GluGly]$_{36}$ (*7, 9*), and (AlaGly)$_3$GluGly(AlaGly)$_3$LysGly (Cantor *et al.*, unpublished observations) fit the same pattern of behavior. The unit cell dimensions determined for all of these materials are summarized in Table II.

The results in Table II show that the average intersheet spacing ($b/2$) increases with increasing volume (*19*) of the putative lamellar surface residue. The relative sizes of these residues are illustrated in Fig. 6, which presents Gly-X-Gly views of turns with X = Ala, Ser, Asn, Val, Glu, Lys, Phe, and Tyr, respectively. Figure 7 illustrates the correlation between intersheet spacing and residue volume, which extends over a range encompassing a 30% increase in the intersheet spacing. These results demonstrate that substitution of appropriate amino acid residues can be used to control the intersheet stacking in periodic polypeptides in a predictable manner.

CONCLUSIONS

Monodisperse populations of precisely defined protein polymers can be generated using recombinant DNA methodology. We have used this approach to produce a series of protein polymers that adopt chain-folded lamellar architectures. Within this series, the substituted "X" residues are proposed to decorate the lamellar surfaces despite differences in steric bulk, hydrophobicity, and charge. In all cases, the interior of the lamellae consist of polar antiparallel β-sheets, folding through three-residue turns, and stacking with like surfaces together. A linear correlation was discovered between the average intersheet stacking distance and the volumes of the amino acids substituted at position X. Thus, fine-tuning of the molecular architecture can be achieved by appropriate choice of the substituted amino acid in crystalline periodic polypeptides.

We thank Joseph Cappello of Protein Polymer Technologies, Inc., for a gift of the p937.51 cloning vector. We also thank Kevin McGrath, Michael Dougherty, and Mary McTernan for early contributions to this work.

REFERENCES

1. Marsh, R.E., Corey, R.B., and Pauling, L. (1955) An investigation of the structure of silk fibroin. *Biochim. Biophys. Acta* **16**, 1-34
2. Fraser, R.D.B., MacRae, T.P., Stewart, F.H.C., and Suzuki, E.

(1965) Poly-L-alanylglycine. *J. Mol. Biol.* **11**, 706-712
3. Lotz, B., Brack, A., and Spach, G. (1974) β-Structure of periodic copolypeptides of L-alanine and glycine. *J. Mol. Biol.* **87**, 193-203
4. Chou, P.Y. and Fasman, G.D. (1977) β-Turns in proteins. *J. Mol. Biol.* **115**, 135-175
5. Sibanda, B.L., Blundell, T.L., and Thornton, J.M. (1989) Conformation of β-hairpins in protein structures. *J. Mol. Biol.* **206**, 759-777
6. Mattos, C., Petsko, G.A., and Karplus, M. (1994) Analysis of two-residue turns in proteins. *J. Mol. Biol.* **238**, 733-747
7. Krejchi, M.T., Atkins, E.D.T., Waddon, A.J., Fournier, M.J., Mason, T.L., and Tirrell, D.A. (1994) Chemical sequence control of β-sheet assembly in macromolecular crystals of periodic polypeptides. *Science* **265**, 1427-1432
8. Parkhe, A.D., Fournier, M.J., Mason, T.L., and Tirrell, D.A. (1993) Determination of the chain-folding pattern in the crystalline domains of the repetitive polypeptide $\{(AlaGly)_3GluGly-(GlyAla)_3GluGly\}_{10}$ by FTIR studies of its blends with a ^{13}C-enriched analogue. *Macromolecules* **26**, 6691-6693
9. Krejchi, M.T., Deguchi, Y., Fournier, M.J., Mason, T.L., Tirrell, D.A., Cooper, S.J., and Atkins, E.D.T. (1997) Crystal structures of chain-folded antiparallel β-sheet assemblies from sequence-designed periodic polypeptides. *Macromolecules*, in press
10. Lucas, F., Shaw, J.T.B., and Smith, S.G. (1960) Comparative studies of fibroins: I. The amino acid composition of various fibroins and its significance in relation to their crystal structure and taxonomy. *J. Mol. Biol.* **2**, 339-349
11. Itoh, K., Foxman, B.M., and Fasman, G.D. (1976) The two β forms of poly (L-glutamic acid). *Biopolymers* **15**, 419-455
12. Warwicker, J.O. (1960) Comparative studies of fibroins: II. The crystal structures of various fibroins. *J. Mol. Biol.* **2**, 350-362
13. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89
14. McGrath, K.P. (1991) Genetically controlled syntheses of novel polymeric materials. Ph.D. thesis, University of Massachusetts, Amherst
15. Dougherty, M.J. (1993) Genetic synthesis of repetitive polypeptides: self-assembly of $[(GlyAla)_3GlyAsp]_{31}$ and *in vivo* incorporation of selenomethionine into $[(GlyAla)_3GlyMet]_9$. PhD thesis, University of Massachusetts, Amherst
16. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*; Vol. 1 pp. 6.32, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
17. Ferrari, F.A., Richardson, C., Chambers, J., Causey, S.C., and Pollack, T.J. (1993) Construction of synthetic DNA and its use in large polypeptide synthesis. U.S. Patent 5,243,038
18. Panitch, A., Matsuki, K., Cantor, E.J., Cooper, S., Atkins, E.D. T., Fournier, M.J., Mason, T.L., and Tirrell, D.A. (1997) Poly-(L-alanyl-glycine): multigram-scale biosynthesis, crystallization, and structural analysis of chain-folded lamellae. *Macromolecules* **30**, 42-49
19. Richards, F.M. (1977) Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176
20. Geddes, A.J., Parker, K.D., Atkins, E.D.T., and Beighton, E. (1968) "Cross-β" conformation in proteins. *J. Mol. Biol.* **32**, 343-358