SYNTHESIS AND CHARACTERIZATION OF PERIODIC POLYPEPTIDES CONTAINING REPEATING —$(AlaGly)_xGluGly$— SEQUENCES

YOSHIKUNI DEGUCHI*, MARK T. KREJCHI*, JANOS BORBELY*, MAURILLE J. FOURNIER**, THOMAS L. MASON** and DAVID A. TIRRELL*

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

ABSTRACT

We have expressed in E. *coli* a series of periodic polypeptides represented by sequence **1.** Our objective has been an understanding of the role of chemical sequence in determining the chain folding behavior of periodic macromolecules. Molecular organization has been examined by infrared spectroscopy and ^{1}H and ^{13}C NMR methods and a preliminary model of the folded structure has been developed.

$$-[-(AlaGly)_xGluGly-]_n \qquad \mathbf{1}$$

a: x=3, n=36
b: x=4, n=18
c: x=4, n=28
d: x=5, n=14
e: x=5, n=20
f: x=6, n=10
g: x=6, n=14

INTRODUCTION

We have been studying a series of periodic polypeptides represented by the general formula *2,* wherein the repeating alanylglycine dyads are expected to bias the chain toward an anti-parallel β-sheet conformation [1,2]. Our objective is control of chain folding in the solid state, by incorporating periodic β-sheet breakers into the sequence. Our first attempt used a variant of sequence **2** in which x=3, Z=ProGlu, and n=54 [3]. X-ray diffraction analysis of this protein afforded no evidence of crystalline structure. In this paper we report recent progress in characterization of the similar family of polymers represented by sequence **1.**

$$-[-(AlaGly)_xZGly-]_n \qquad \mathbf{2}$$

EXPERIMENTAL SECTION

Synthetic strategy. The overall strategy for the production of the target material is shown in Figure 1. Detailed experimental conditions are described elsewhere [3,4].

Protein expression. Artificial genes encoding proteins containing sequences **1a-g** were constructed and expressed in the pET3-b vector and E. *coli* strain BL21(DE3)pLysS developed by Studier and coworkers [5]. Expression was initially monitored by the incorporation of ^{3}H-glycine into the target protein. A rich medium (YT) was used for sample preparation. β-Isopropylthiogalactoside (IPTG, final concentration 95 mg/L) was added to the culture to induce protein production when OD at 600 nm reached ca. 0.8.

Purification. Cell extracts were prepared by freezing, thawing, sonication and spinning out of cell debris. The extract was adjusted sequentially to pH 5.0, 4.5 and 4.0 with glacial acetic acid. After each pH adjustment, precipitated material was removed by centrifugation. After enzymatic removal of contaminating nucleic acids, the target protein was precipitated by addition of ethanol to 40 v/v%. Crude product was washed with water repeatedly until the absorbance of the supernatant at 260 nm disappeared.

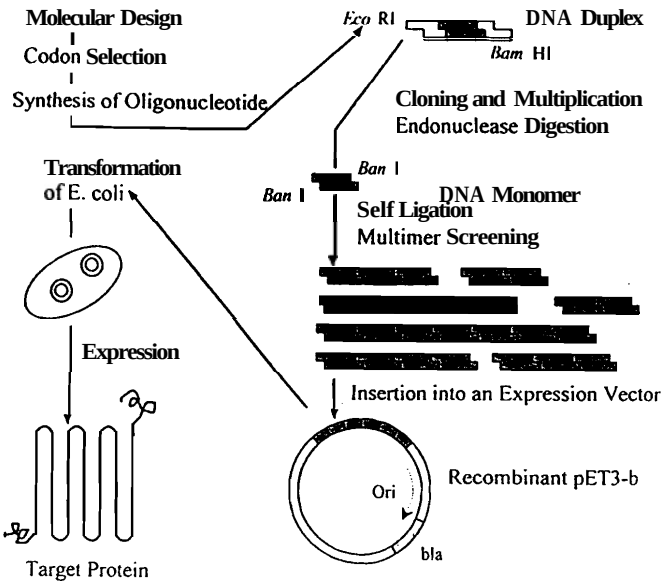

Figure 1. Overall strategy for **biosynthesis** of **polymers of** sequence **1.**

Cyanogen bromide cleavage. Fusion ends were chemically removed by selective cyanogen bromide cleavage at room temperature in formic acid.

Infrared spectroscopy. A portion of cleaved protein was placed in an Eppendorf tube and 90% formic acid was added thereto so that 1 μL of solution would contain 100 μg of solid. The solution was kept at room temperature until use. After an appropriate time period, a portion of the solution was spread on a NaCl plate and air dried.

Computer graphics. Molecular modeling was performed using Biograf version 2.2 (BioDesign, Inc.) running on a Silicon Graphics Iris work station.

NMR spectroscopy. NMR analyses were conducted at room temperature on Varian XL-300, Varian 500 MHz Unity, and Bruker 200 AF, 300 MSL and 500 AMX spectrometers. Formic acid formyl-d_1 was used as solvent.

RESULTS and DISCUSSION

Protein expression. The expression strain, BL21(DE3)pLysS, carries a chromosomal gene encoding bacteriophage T7 RNA polymerase under *lacUV5* control [5]. The pLysS plasmid produces a low level of T7 lysozyme, which suppresses leaky T7 RNA polymerase activity prior to induction with IPTG. Each of the proteins 1a-g was expressed in E. ***coli*** successfully. None of these proteins stained well with the conventional dye Coomassie Blue, so ***in vivo*** labelling techniques were used to monitor accumulation.

Purification. Contaminating proteins were removed readily from the product by stepwise acidification of whole cell lysates, owing to the fact that the target proteins remain soluble in mildly acidic solutions that precipitate native E. ***coli*** proteins. Purification leads to products of general sequence 3, in which the repetitive oligopeptide building blocks are flanked by leader and trailer sequences derived from the cloning and expression vectors. Typical yields of purified protein were 20-70 mg/L, decreasing slightly with increased spacing between

glutamic acid residues. The leader and trailer sequences were removed by cyanogen bromide digestion prior to structural analysis by infrared and NMR methods.

MASMTGGQQMGRDPMFKYSRDPM[(GA)$_x$GE]$_n$ **3**
GARMHIRPGRYQLDPAANKARKEAELAAATAEG

Infrared spectroscopy. Figure *2* shows IR absorption spectra of films of polymer **lc** cast from formic acid after aging of the casting solution for various times. The broad band in the amide I region is apparently composed of two main peaks, one centered near 1660 cm^{-1} and the other around 1625 cm^{-1}. The former band decreased in intensity as the solution aged, while the intensity of the latter band increased. The rate of change in the relative intensities of these two bands was found to depend on the sample concentration, and was reduced in more dilute solutions. Gelation of each solution was observed concurrently with the rise in intensity at 1625 cm^{-1}. We attribute the 1625 cm^{-1} band to an antiparallel β-sheet structure and the higher frequency absorption to a disordered conformation.

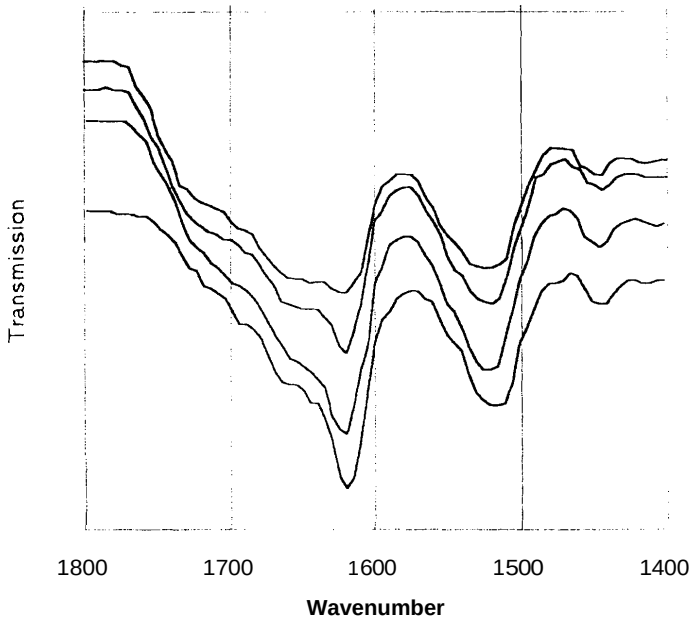

Figure 2. Infrared spectra of cast films. Cleaved lc protein was dissolved in formic acid and films were cast on NaCl. From the top, curves correspond to 5 min, 2 hr, 8 hr, and 26 hr of solution aging prior to film casting.

NMR spectroscopy. The infrared results discussed above, coupled with the observed gelation of the formic acid solutions upon aging, suggest that an ordered β-structure develops with time in these polymer solutions. This structure was probed more fully in preliminary high-field NMR experiments. Of special interest were questions of: i). whether or not the chain folds in regular fashion in solution, and ii). if regular folding does occur, what the local structure of the fold may be.

The ^{13}C NMR spectrum of polymer lc dissolved in formic acid-dl is shown in Figure 3. Chemical shifts relative to formic acid (6166.31 ppm) and peak assignments are listed in Table I. Assignments were made on the basis of repeating unit sequence 4, in which the designation of 3 distinct glycyl units and 2 distinct alanyl units is suggested by the appearance of multiple carbonyl resonances for each of these residues.

—(GlyAla)$_3$Gly*GluGly**Ala*— **4**

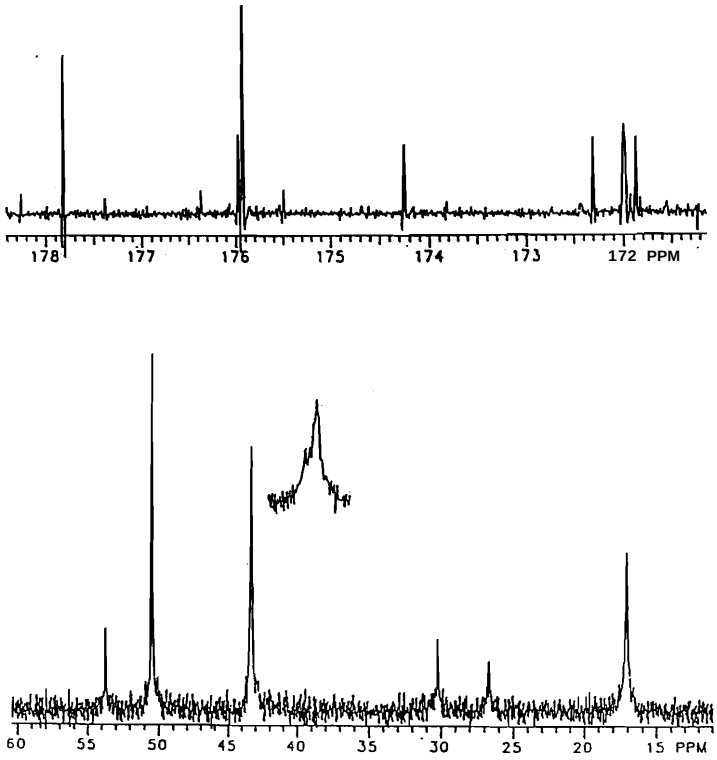

Figure 3. ^{13}C NMR spectra **of cleaved lc. Upper:** carbonyl carbons; **lower: aliphatic carbons.**

Table I. Assignment of ^{13}C NMR Spectrum of Polymer lc

δ in ppm	*Assignment*
17.03	β-CH_3 of all Ala residues
26.63	β-CH_2 of **Glu**
30.21	γ-CH_2 **of** Glu
43.28	a-CH2 of Gly
43.33	α-CH_2 of Gly* or **Gly****
43.37	a-CH2 of Gly* or **Gly****
50.51	a-CH of Ala and **Ala***
53.76	a-CH of Glu
171.88	**Carbonyl** of Gly*
172.00	**Carbonyl** of Gly
172.31	**Carbonyl** of Gly**
174.26	**Carbonyl** of Glu
175.94	**Carbonyl** of **Ala**
175.98	**Carbonyl** of Ala*
177.83	Side chain **carboxyl group** of Glu

The relative intensities of the carbonyl signals are consistent with this assignment, and the presence of sharp resonances suggests that the chain adopts a limited set of conformations in formic acid solution. In contrast, the ^{13}C NMR spectrum **obtained** in HFIP-d_2 solution shows very broad **carbonyl** peaks, suggesting that in HFIP the polypeptide **chain** exists as a **random** coil.

Figure 4 shows a partial NOESY spectrum of 1c taken in formic acid solution. The observation of strong sequential a-proton-NH NOEs and absence of amide-amide NOEs suggest that pleated β-sheets are present. This is supported by the appearance of H_α-H_α and H_α-H_β NOEs which are to be expected in a regular antiparallel β-pleated sheet, in which the interstrand H_a-H_a distance is approximately 2.4 Å. Furthermore, the NH peak of Gly** indicates a short interresidue interaction with the a-CH and a weak interaction with the β-CH_2 of Glu. This observation is consistent with the occurrence of a β-turn. The very strong correlation of Gly_i**NH -> Glu_{i-1}αCH suggests a turn of type II rather than type I or I'.

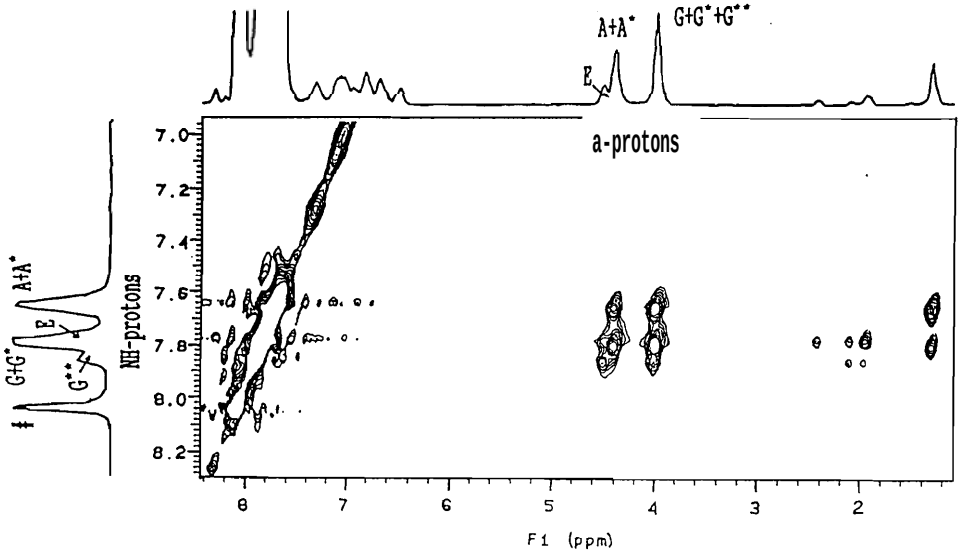

Figure 4. Partial NOESY spectrum of cleaved 1c.

CONCLUSIONS

Although this is a preliminary analysis, the ^1H and ^{13}C NMR data obtained in formic acid are consistent with a chain conformation characterized by a predominance of β-strands interrupted by reverse turns. In contrast, the polymer in HFIP appears to be a statistical coil, giving rise to broad resonances in the ^{13}C NMR spectrum. Given these differences, it is intriguing that polymer films cast from formic acid exhibit infrared spectra characteristic of P-sheet structure, while those from HFIP are disordered. We are pursuing refinement of the solution structure of 1c and related polymers, with the objective of understanding the mechanism of structure formation during solidification.

ACKNOWLEDGMENT

This work was supported by a grant from the Polymers and Genetics Programs of the National Science Foundation (DMR-8914359) and by the Materials Research Laboratory at the University of Massachusetts. J. Borbély thanks the Fulbright Foundation for financial support. Dr. Borbély is on leave from the Department of Applied Chemistry, Kossuth Lajos University, H-4010 Debrecen, Hungary. Yoshikuni Deguchi was on leave from Central Research Laboratories, Kaneka Corporation, 1-2-80, Yoshida-cho, Hyogo-ku, Kobe 652, JAPAN.

REFERENCES

1. D.A. Tirrell, M.J. Fournier and T.L. Mason, *MRS Bulletin* 16(7), *23* (1991).
2. R.D.B. Fraser, T.P. MacRae, F.H.C. Stewart and E. Suzuki, J. *Mol. Biol.* 11, 706 (1965).
3. K.P. McGrath, M.J. Fournier, T.L. Mason and D.A. Tirrell, J. Am. *Chem. Soc.* 114, 727 (1992).
4. Y. Deguchi, M J. Fournier, T.L. Mason and D.A. Tirrell, in preparation.
5. F.W. Studier, A.H. Rosenberg and J.J. Dunn, *Methods Enzymol.* 185, 60 (1990).
6. P.Y. Chou and G.D. Fasman, J. *Mol. Biol.* 115, 135 (1977).