Genetically Directed Syntheses of New Polymeric Materials. Efficient Expression of a Monodisperse Copolypeptide Containing Fourteen Tandemly Repeated -(AlaGly)$_4$ProGluGly- Elements

The statistical nature of chemical polymerization processes ensures that the materials prepared by such methods will consist of mixtures of chains. In contrast, ribosomal catalysis of protein synthesis on messenger RNA templates affords homogeneous products in which chain length, sequence, and stereochemistry are subject to exquisite control. With the advent of recombinant DNA methods in the 1970s, it has become possible to exploit this control in a synthetic sense, i.e., to create new, artificial polypeptides of predetermined structure. Given the diversity of functionality, sequence, conformation, and chain lengths that become accessible in this way, it appears certain that this approach will find important applications in polymer materials science.

We are exploring the use of recombinant DNA methods to prepare a wide variety of copolypeptides of periodic structure. Such sequences are of particular interest in materials science, and in fact, many of the naturally occurring structural proteins (e.g., silk, elastin, and collagen) consist in large part of highly repetitive sequences.[1] In addition, the expression of artificial genes encoding repetitive polypeptides presents unusual and intriguing problems, because repetitive DNAs may be particularly unstable in bacterial hosts[2] and because repetitive messenger RNAs may adopt folded structures that are translated inefficiently or degraded rapidly. Previous attempts at expression of large repetitive genes have met with mixed success.[3-5]

We report herein efficient expression of a periodic copolypeptide containing 14 repeats of the undecapeptide sequence 1. We are exploring this and related polymers

-(AlaGly)$_4$ProGluGly-
1

in an attempt to define the determinants of crystallization and folding in periodic chains.

Oligonucleotide 2, which encodes two repeats of undecapeptide 1, was synthesized on a Milligen Biosearch Model 8700 DNA synthesizer using CED–phosphoramidite chemistry.[6] The oligonucleotides were purified by electrophoresis, enzymatically phosphorylated at the 5′ termini, annealed, and ligated between the BamHI and EcoRI sites in the polylinker region of pUC18.[7,8] *Escherichia coli* strain DH5αF′ cells were transformed and screened by insertional inactivation of the β-galactosidase gene. Subsequent restriction analysis with ApaI and BanI confirmed the presence of the insert, and the sequence of the insert was verified by double-stranded sequencing using the T7Sequencing protocol supplied by Pharmacia. Band compressions in the sequencing gel were resolved by replacement of 2′-deoxyguanosine 5′-triphosphate (dGTP) in the sequencing reactions with 7-deaza-dGTP or with 7-deaza-2′-deoxyinosine 5′-triphosphate. BanI digestion of the amplified plasmid yielded a 66-base-pair fragment bearing nonpalindromic termini and encoding two repeats of the undecapeptide. This fragment was purified by electrophoresis in a 10% polyacrylamide gel and recovered by electroelution.

The 66-base-pair monomer was self-ligated to yield a population of multimers with a chain length distribution centered at a degree of polymerization of 20. Multimers encoding up to 200 repeats of the undecapeptide were visible on the gel. The multimer population was cloned into the unique BanI site of p937.51, a small, high copy number plasmid supplied by Protein Polymer Technologies that features an origin of replication derived from pBR322 and a gene encoding chloramphenicol acetyltransferase.[9] The resulting plasmid mixture was used to transform *E. coli* strain HB101, and transformants were screened for the presence of recombinant plasmids by colony hybridization using as a probe one of the starting synthetic oligonucleotides labeled at the 5′ terminus with ^{32}P. Plasmids were found with inserts of degrees of polymerization of $\leq$10. Future ligations with a size-fractionated multimer population should yield clones with larger inserts (while the extent of polymerization can be subject to biological limitation,[10] substantially larger multimeric inserts have been observed in our laboratories for two sequences closely related to 1). One recombinant plasmid, carrying an insert encoding 14 repeats of the undecapeptide, was selected for further analysis.

The insert was recovered from p937.51 by BamHI digestion, purified, and inserted into pET3-b,[11] an expression vector in which transcription is driven by T7 RNA polymerase. This construction yields an open reading frame encoding 14 repeats of the undecapeptide plus short N- and C-terminal extensions as shown in sequence 3.[12,13] The terminal extensions can in principle be removed efficiently by CNBr cleavage at the methionine residues flanking the repetitive portion of the polymer.

MASMTGGQQMGRDPMFKYSRDPM-
[GAGAGAGAGPEGAGAGAGAGPE]$_7$-
GARMHIRPGRYQLDPAANKARKEAELAAATAEQ
3

A recombinant plasmid containing the insert in the proper orientation, designated pET1-7, was used to transform *E. coli* strain BL21(DE3)pLysS. In this strain, a gene encoding T7 RNA polymerase is incorporated into the bacterial chromosome under control *lacUV5*, and protein production is induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). The pLyS plasmid supplies T7 lysozyme, which inhibits T7 RNA polymerase activity and suppresses the basal level of protein expression.

Expression of the target protein was monitored by the in vivo incorporation of [^{3}H]glycine into proteins synthesized during mid-log growth (A_{600} = 0.6–0.8) in M9 minimal medium supplemented with a mixture of amino acids lacking glycine.[7] The rate of cell growth was found to be normal prior to induction but then declined as a prominent new protein product appeared in electrophoretic analyses of whole cell lysates (Figure 1). This product was not found in untransformed BL21(DE3)pLysS or in BL21(DE3)pLysS transformed with pET3-b. The product migrates with an apparent molecular weight of 40 000, and although this is higher than the expected molecular weight of 17 207 similar proteins are known to migrate

```
Stop Gly Ala Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala Gly Ala Gly Ala Gly Ala Gly Pro Glu Gly Ala
AATTCG TAA GGT GCC GGC GCT GGT GCT GGG GCC GGT CCG GAA GGT GCA GGC GCT GGC GCG GGC GCG GGC CCG GAA GGT GCC G
   GC ATT CCA CGG CCG CGA CCA CGA CCC CGG CCA GGC CTT CCA CGT CCG CGA CCG CGC CCG CGC CCG GGC CTT CCA CGG CCTAG
EcoRI        BanI                                                                    ApaI        BanI    BamHI
```


0024-9297/91/2224-1213$02.50/0

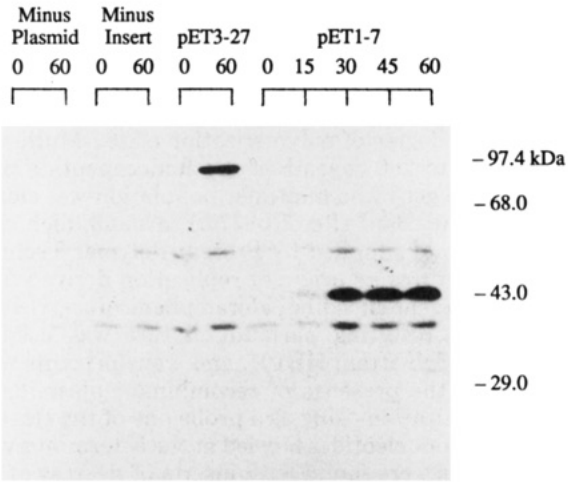

Figure 1. SDS polyacrylamide gel analysis of in vivo labeling of proteins derived from pET1-7. Lanes 1–4 are negative controls; lanes 5 and 6 represent a (positive) control expression of a pET recombinant encoding $[(AlaGly)_3ProGluGly]_{54}$, reported previously by McGrath and co-workers.[14] Lanes 7–11 represent time points (in minutes) after addition of IPTG to BL21(DE3)pLysS cells transformed with pET1-7. The gel was soaked in Enlightening autoradiography enhancer (Du Pont NEN) and dried prior to exposure.

anomalously slowly on SDS polyacrylamide gels.[14,15]

No evidence of genetic instability was observed in these experiments. Electrophoretic analysis of isolated plasmid DNA revealed no length polymorphism in the insert, even in cultures grown through ca. 35 generations and in which synthesis of **3** had been induced. The product polymer also appeared to be relatively stable in this host strain and continued to accumulate throughout the typical 2–2.5-h period between induction and harvest.[16]

For purification of **3**, the same recombinant strain was grown in a New Brunswick Scientific Microferm Fermenter in 10 L of YT medium (8 g of Bacto-tryptone, 5 g of yeast extract, 5 g of sodium chloride per liter of distilled, deionized water) containing 200 mg/L of ampicillin and 25 mg/L of chloramphenicol at 37 °C. Cells were grown to an optical density of 0.8–1.0 at 600 nm with vigorous aeration (9000 cm^3/min with 400 rpm agitation; 1 mL of Sigma Antifoam A was added to suppress foaming), and protein synthesis was induced by addition of IPTG to a concentration of 0.4 mM. Cells were harvested 2–2.5 h after induction by centrifugation at 4000*g* for 20 min at 4 °C. Pelleted cells (8–10 g dry weight) were resuspended in 100–200 mL of distilled water and lysed by freezing and thawing. The lysate was centrifuged at 10 000*g* for 30 min and the recovered supernatant adjusted to pH 4 by the addition of glacial acetic acid. After centrifugation (10 000*g*, 20 min, 4 °C), the supernatant was adjusted to 40% ethanol, centrifuged again, and finally adjusted to 80% ethanol to precipitate the protein of interest. After washing with distilled water, the product was redissolved in 0.1 M ammonium bicarbonate, dialyzed against distilled water, and recovered by lyophilization. Typical yields after drying to constant weight over P_2O_5 at 80 °C were 400 mg/10 L of fermentation. Repetition of this procedure has provided ca. 5 g of product.

The structure of **3** was confirmed in several ways. ^1H NMR spectrometry (300 MHz, D_2O) reveals only those resonances expected from the sequence as written. Amino acid compositional analysis shows the product to consist of 86.3% glycine, alanine, proline, and glutamic acid vs the expected total of 85.7% for these four residues. The purified protein runs as a single band on SDS polyacrylamide gel electrophoresis, even when overloaded (i.e., 1 mg of protein/lane). Absence of substantial non-protein contamination was verified by the absence of a 260-nm absorption due to nucleic acids and by combustion analysis. Anal. Calcd for **3** plus 12% water: C, 43.8%; H, 6.9%; N, 17.1%. Found: C, 43.8%; H, 6.0%; N, 16.7%; ash < 0.1%.

The purified polymer undergoes a reversible glass transition at 182 °C (differential scanning calorimetry, 20 °C/min), starts to decompose under nitrogen at ca. 250 °C with subsequent loss of 70% of its weight (thermogravimetric analysis, 20 °C/min), and is readily cast into coherent films from formic acid. Detailed analysis of the solid-state structure of **3** is in progress.

Acknowledgment. We thank Dr. Joseph Cappello of Protein Polymer Technologies for the gift of p937.51 and for fruitful discussions regarding the gene construction described herein. This work was supported by a grant from the Polymers and Genetics Programs of the National Science Foundation (DMR 8914359) and by the NSF Materials Research Laboratory at the University of Massachusetts.

References and Notes

(1) Walton, A. G.; Blackwell, J. *Biopolymers*; Academic Press: New York, 1973; Chapter 10.
(2) Mahajan, S. K. *Genetic Recombination*; Kucherlapati, R., Smith, G. R., Eds.; American Society for Microbiology: Washington, DC, 1988; Chapter 3.
(3) McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J. *Biotechnol. Prog.* **1990**, *6*, 188 and references therein.
(4) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198.
(5) Goldberg, I.; Salerno, A. J.; Patterson, T.; Williams, J. I. *Gene* **1989**, *80*, 305.
(6) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.
(7) All nucleic acid manipulations were accomplished via standard protocols. Cf.: Sambrook, J.; Frisch, E. F.; Maniatis, T. *Molecular Cloning. A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Press: Cold Spring Harbor, NY, 1989.
(8) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985**, *33*, 103.
(9) Ferrari, F. A.; Richardson, C.; Chambers, J.; Causey, S. C.; Pollock, T. J. U.S. Patent Appl. 927,258, 1986.
(10) Case, S. T. *Gene* **1982**, *20*, 169. Doel, M. T.; Eaton, M.; Cook, E. A.; Lewis, H.; Patel, T.; Carey, N. H. *Nucleic Acid Res.* **1980**, *8*, 4575. Gupta, S. C.; Weigh, H. L.; Somerville, R. L. *Bio/Technology* **1983**, *1* (7), 602.
(11) Studier, F. W.; Rosenberg, A. H.; Dunn, J. J. *Methods in Enzymology*; Academic Press: Orlando, FL, 1990; Vol. 185, p 60.
(12) Standard single-letter abbreviations for the amino acids may be found in many biochemical textbooks; see, for example: Mathews, K.; van Holde, K. E. *Biochemistry*; Benjamin/Cummings: Redwood City, CA, 1990.
(13) Preliminary analysis of the recovered copolypeptide by mass spectrometry indicates that the N-terminal methionine of sequence **3** is deleted.
(14) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polym. Prepr.* **1990**, *31* (1), 190.
(15) McGrath, K. P.; Cappello, J., personal communications.
(16) The continued accumulation of **3** beyond 30 min is not readily apparent in Figure 1 but is clearly observed in gels stained with Coomassie Brilliant Blue. The apparent plateau in accumulation in Figure 1 is a result of rapid consumption of $[^3H]$glycine, which in this experiment was added in a single pulse shortly before induction.
(17) Department of Polymer Science and Engineering.
(18) Department of Biochemistry.
(19) Molecular and Cellular Biology Program.

Howard S. Creel,[17] Maurille J. Fournier,[18,19] Thomas L. Mason,[18,19] and David A. Tirrell*[17,19]

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

Received October 5, 1990
Revised Manuscript Received November 30, 1990

Registry No. 2, 131406-53-6.