Journal of Macromolecular Science, Part A
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597274

Periodic Polypeptides Based on Poly(L-Alanylglycine): Biological Synthesis and Verification of the Structure of A Series of Polymers Containing Tandem —(Alagly)$_x$Glugly— Repeats

Yoshikuni Deguchif; Maurille J. Fournier; Thomas L. Mason; David A. Tirrell

To cite this Article Deguchif, Yoshikuni , Fournier, Maurille J. , Mason, Thomas L. and Tirrell, David A.(1994) 'Periodic Polypeptides Based on Poly(L-Alanylglycine): Biological Synthesis and Verification of the Structure of A Series of Polymers Containing Tandem —(Alagly)$_x$Glugly— Repeats', Journal of Macromolecular Science, Part A, 31: 11, 1691 — 1700
To link to this Article: DOI: 10.1080/10601329408545877
URL: http://dx.doi.org/10.1080/10601329408545877

J.M.S.—PURE APPL. CHEM., A31(11), pp. 1691–1700 (1994)

PERIODIC POLYPEPTIDES BASED ON POLY(L-ALANYLGLYCINE): BIOLOGICAL SYNTHESIS AND VERIFICATION OF THE STRUCTURE OF A SERIES OF POLYMERS CONTAINING TANDEM —(AlaGly)$_x$GluGly— REPEATS

YOSHIKUNI DEGUCHI†

Department of Polymer Science and Engineering

MAURILLE J. FOURNIER and THOMAS L. MASON

Department of Biochemistry and Molecular Biology

DAVID A. TIRRELL*

Department of Polymer Science and Engineering

University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

The experiments reported herein constitute part of an investigation of the relationship between primary sequence and higher order structure in repetitive polypeptides. Three series of artificial DNAs encoding polypeptides represented by the general formula $\dashv(\mathrm{GlyAla})_n\mathrm{GlyGlu}]_m$— were constructed and cloned in *Escherichia coli*. Protein expression was monitored by in-vivo labeling with ^{3}H-glycine. Protein products were isolated from the cell lysates by stepwise acidification in yields of 20–50 mg/L of

†Present address: Central Research Laboratories, Kaneka Corporation, 1-2-80 Yoshida-cho, Hyogo-ku, Kobe 652, Japan.

culture, and fusion fragments derived from the cloning vectors were removed by cyanogen bromide cleavage. Amino acid sequence analysis verified the expected sequences through the first 40–50 N-terminal residues.

INTRODUCTION

Artificial proteins constitute a new class of macromolecular materials which can be engineered with precise control of chain length, sequence, and stereochemistry [1]. The uniformity of structure that characterizes these materials distinguishes them from conventional polymers, which are always heterogeneous.

Much of our own work on artificial proteins has been motivated by the premise that uniform polymer chains will prove to be useful in the engineering of materials properties on short-length scales, as in crystals or in liquid crystals, or at surfaces [2–4]. We have recently reported the synthesis of a class of polymers of Sequence **1** with Z = Ala, Asn, Asp, Glu, Leu, Met, Phe, Thr, Tyr, and Val [5]. The design of these sequences was based on the known propensity of poly(alanylglycine) to adopt sheetlike structures in the solid state [6,7] and on the notion that suitably chosen intervening units (Z) might be exploited to control the folding periodicity of the chain and the surface functionality of crystals formed from it. In fact, several variants of Sequence **1** have been found to crystallize in β-sheet arrangements [8], and that with Z = Glu appears to fold with the expected periodicity [9].

$$\dagger(\text{AlaGly})_3\text{ZGly}\dagger_x \qquad \mathbf{1}$$

In the present paper we describe the biosynthesis of the related series of polymers represented by Sequence **2**. The overall objective of this work is an exploration of the relation between chemical sequence periodicity and crystal thickness. We chose to examine chains comprising 8–14 residues per repeating unit for two reasons: i) work by Dreyfuss and Keller [10] and by Atkins and coworkers [11] suggests that aliphatic polyamides fold with a periodicity corresponding to 6–8 lateral hydrogen bonds per crystalline stem, and ii) the egg-stalk protein of *Chrysopa flava* is known to adopt a cross-β-structure in the solid state, with a folding periodicity of 8 residues [12]. Thus chains such as **1** might be expected to be optimally designed for folding "in register" with the chemical sequence, whereas in the family of polymers **2** the kinetic factors favoring short stems may be counterbalanced by a requirement that the polar, bulky glutamic acid residues, separated by distances greater than the preferred stem length, must be accommodated in the folded structure. Detailed analyses of the solid-state structures of **2** are ongoing and will be published elsewhere; we report herein the preparation of these polymers and the verification of their molecular structures.

$$\dagger(\text{AlaGly})_x\text{GluGly}\dagger_n$$

2a: $x = 4, n = 28$
b: $x = 4, n = 18$
c: $x = 5, n = 20$
d: $x = 5, n = 14$
e: $x = 6, n = 14$
f: $x = 6, n = 10$

EXPERIMENTAL SECTION

Materials. All restriction enzymes, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs (Beverly, Massachusetts). RNase A and DNase I were from Sigma Chemical Co. The T7 Sequencing Kit was supplied by Pharmacia LKB Biotechnology AB. *E. coli* strains DH5αF′ and HB101 were obtained from Bethesda Research Labs (Bethesda, Maryland). Plasmid p937.51 was a gift from Protein Polymer Technologies, Inc. (San Diego, California). The expression vector pET3-b and *E. coli* strain BL21(DE3)pLysS were provided by Dr. William Studier at the State University of New York, Stony Brook. ^{3}H-glycine and autoradiography enhancer Enlighten were purchased from Du Pont (Wilmington, Delaware). Water was doubly distilled, and other chemicals were reagent grade unless otherwise noted.

Preparation of Oligonucleotide Duplexes. Oligonucleotides were synthesized on a Biosearch Model 8700 DNA Synthesizer on a 1-μmol scale. Crude oligomers cleaved from the support were purified by polyacrylamide gel electrophoresis (PAGE) on a 12% gel containing 7 M urea. 5′-Termini of purified oligonucleotides were phosphorylated enzymatically using T4 polynucleotide kinase, and complementary oligomers were annealed at 80°C and cooled to room temperature over 3 hours. The annealed duplexes were precipitated with salt/ethanol, dried in vacuo, and quantified on a 2% agarose gel containing ethidium bromide.

Cloning and Verification of Sequence. Duplexes prepared as above were ligated between the *Eco*RI and *Bam*HI sites of the cloning vector pUC18, and the resulting recombinant plasmids were used to transform cells of *E. coli* strain DH5αF′. Colonies were selected by color identification and coding sequences were verified using the T7 Sequencing Kit protocol.

Multimerization of the DNA. Plasmid DNA was amplified in 1 L of 2xYT medium with ampicillin selection and digested with *Ban*I. The resulting DNA "monomers" were purified by PAGE and then self-ligated to generate repetitive DNA multimers.

Cloning of Multimers. Multimer populations were inserted into the unique *Ban*I site of the transfer vector p937.51 [13] with or without fractionation in advance. Recombinant plasmids were used to transform *E. coli* strain HB101, and transformants were examined for insert size by *Bam*HI digestion and gel electrophoresis of plasmid DNA.

Preparation of Bacterial Expression System. DNA multimers of known degree of polymerization (DP) were recovered by electroelution and inserted into *Bam*HI-digested vector pET-3b [14]. The ligation mixture was used to transform *E. coli* strain HB101, and portions of the plasmid DNA were digested with *Bam*HI and *Ava*I, respectively, to determine the presence and orientation of the insert. Recombinant plasmids containing multimer inserts in the proper orientation were then used to transform expression strain BL21(DE3)pLysS.

Protein Expression. Expression was monitored by the incorporation of ^{3}H-glycine into the target protein. Cells were grown overnight in 5 mL M9AA medium [15] lacking glycine in the presence of appropriate antibiotics. A portion of each overnight culture was used to inoculate 10 mL fresh minimal medium with antibiotics to adjust the initial optical density (OD_{600}) to 0.1. These cultures were agitated at 37°C, and the OD_{600} was recorded every 30 minutes. IPTG (20 μL, 200 mM) was added when the OD_{600} reached 0.5–0.6, and ^{3}H-glycine (75 μL, 1 Ci/μL) was added 15 minutes later. Agitation was continued for another 2 hours. Cells were harvested periodically and lysed in SDS and mercaptoethanol, and the lysates were boiled for 3 minutes and loaded on a 10% discontinuous polyacrylamide gel [16]. After electrophoresis at 15 mA for 12 hours, the gel was stained with Coomassie Brilliant Blue R-250 solution, destained, treated with autoradiography enhancer (Enlighten), and dried on filter paper. The dried gel was exposed to x-ray film at −70°C for 2 days.

Protein Production. Protein production was carried out as above except that rich medium (YT) was used in place of minimal medium. Cells were grown in YT containing 200 mg/L ampicillin and 25 mg/L chloramphenicol until OD_{600} reached 0.8, and then IPTG (final concentration 95 mg/L) was added. Cells were harvested by centrifugation 2 hours after induction. Collected cells were suspended in 50 mM Tris/HCl containing 1 mM EDTA (pH 7.5) and lysed by freezing and thawing of the suspension. The lysate was sonicated on ice, and cell debris was removed by centrifugation. Each cell extract was transferred to a new centrifuge bottle, and glacial acetic acid was added gradually to adjust the pH to 5.0, 4.5, and 4.0 sequentially. Between pH adjustments, the vessel was incubated at 4°C for 2 hours and then centrifuged to remove precipitated solids. Aqueous sodium hydroxide was added to pH 7.0. RNase A, DNase I, and $MgCl_2$ (final concentration 10 mM) were added, and the mixture was incubated at 37°C for 1 hour. The solution was adjusted to pH 4.0 with acetic acid, and a small amount of precipitate was removed by centrifugation and filtration. Ethanol was added to 40 v/v %, and the protein was precipitated in the freezer for 2 days. The crude protein was recovered by centrifugation and washed repeatedly with water until the absorbance of the wash at 260 nm was essentially eliminated. The product was then washed twice with acetone and dried in vacuo at room temperature.

Cyanogen Bromide Cleavage. Fusion fragments derived from the cloning and expression vectors were removed by cyanogen bromide cleavage. A large excess of cyanogen bromide was added to the fusion protein in 90% formic acid, and the reaction mixture was stirred under nitrogen overnight at room temperature. Excess cyanogen bromide and solvent were removed in vacuo at room temperature. The resulting pellet was washed repeatedly with water, rinsed with methanol, and dried under reduced pressure at room temperature.

Sequence Analysis. Amino acid sequence analysis was performed via Edman degradation on an Applied Biosystems 477A Protein Sequencer.

RESULTS AND DISCUSSION

Gene Design. The overall strategy used in the design and expression of protein-based materials is shown schematically in Figure 1 [8a]. To implement this

strategy, the repeating units of target sequences **2a–f** were encoded into the corresponding oligonucleotides **3a–c**. Each of these oligonucleotides was outfitted with terminal sequences that allow insertion into cloning vectors digested with the common restriction endonucleases *Eco*RI and *Bam*HI. Internal *Ban*I sites allow retrieval of the coding fragments following amplification and sequence verification, and the stop codon (TAA) at the 5′-end of each coding sequence ensures effective color identification of transformed cells harboring recombinant plasmids. *Ban*I digestion generates nonpalindromic cohesive ends on the coding fragments and reduces the probability of undesirable head-to-head polymerization of the DNA monomers (Fig. 1). The choice of codons in **3a–c** reflects the relative frequencies of codon use in *E. coli* [17], with GGT and GGC used predominantly to encode glycine and GCT and GCG to encode alanine.

$-(GlyAla)_4-GlyGlu-$ 76bp

```
5′ ATT   TCG TAA GGT GCC GGC GCT GGT GCG GGC GCT GGT
  EcoRI  GC  ATT CCA CGG CCG CGA CCA CGC CCG CGA CCA
                 BanI
         GAA GGT GCA GGC GCT GGT GCG CGC GCT GGT
         CTT CCA CGT CCG CGA CCA CGC CCG CGA CCA

         GAA GGT GCC G    BamHI   3′
         CTT CCA CGG CCTAG
              BanI
```

$-(GlyAla)_5-GlyGlu-$ 88bp **3**

```
5′ ATT   TCG TAA GGT GCC GGC GCT GGT GCG GGC GCT GGT GCG GGC
         GC  ATT CCA CGG CCG CGA CCA CGC CCG CGA CCA CGC CCG

         GAA GGT GCA GGC GCT GGT GCG CGC GCT GGT GCG GGC
         CTT CCA CGT CCG CGA CCA CGC CCG CGA CCA CGC CCG

         GAA GGT GCC G   3′
         CTT CCA CGG CCTAG
```

$-(GlyAla)_6-GlyGlu-$ 100bp

```
5′ ATT   TCG TAA GGT GCC GGC GCT GGT GCG GGC GCT GGT GCG GGC GCT GGT
         GC  ATT CCA CGG CCG CGA CCA GCG CCG CGA CCA CGC CCG CGA CCA

         GAA GGT GCA GGC GCT GGT GCG CGC GCT GGT GCG GGC GCT GGT
         CTT CCA CGT CCG CGA CCA CGC CCG CGA CCA CGC CCG CGA CCA

         GAA GGT GCC G   3′
         CTT CCA CGG CCTAG
```

Subsequent to polymerization, DNA multimers were cloned in the transfer vector p937.51, a high copy-number plasmid that features an origin of replication derived from pBR322 and a gene encoding chloramphenicol acetyltransferase [13]. This step equips the coding sequence with flanking *Bam*HI sites for transfer to the expression plasmid and allows selection of specific multimers encoding the chain-length variants of interest. Evidence of genetic instability was observed at this stage, in that *Bam*HI digestion of recombinant plasmids obtained from isolated clones generated "ladders" of DNA multimers rather than the expected discrete insert sizes. Following transfer to the expression plasmid pET3-b [14], no such instability was apparent.

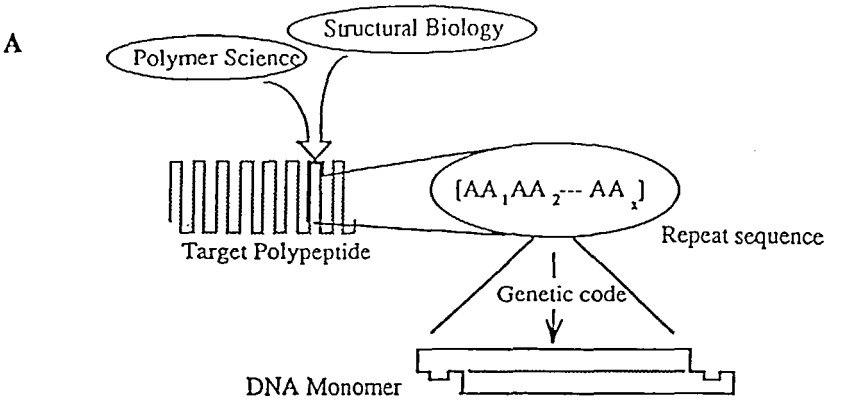
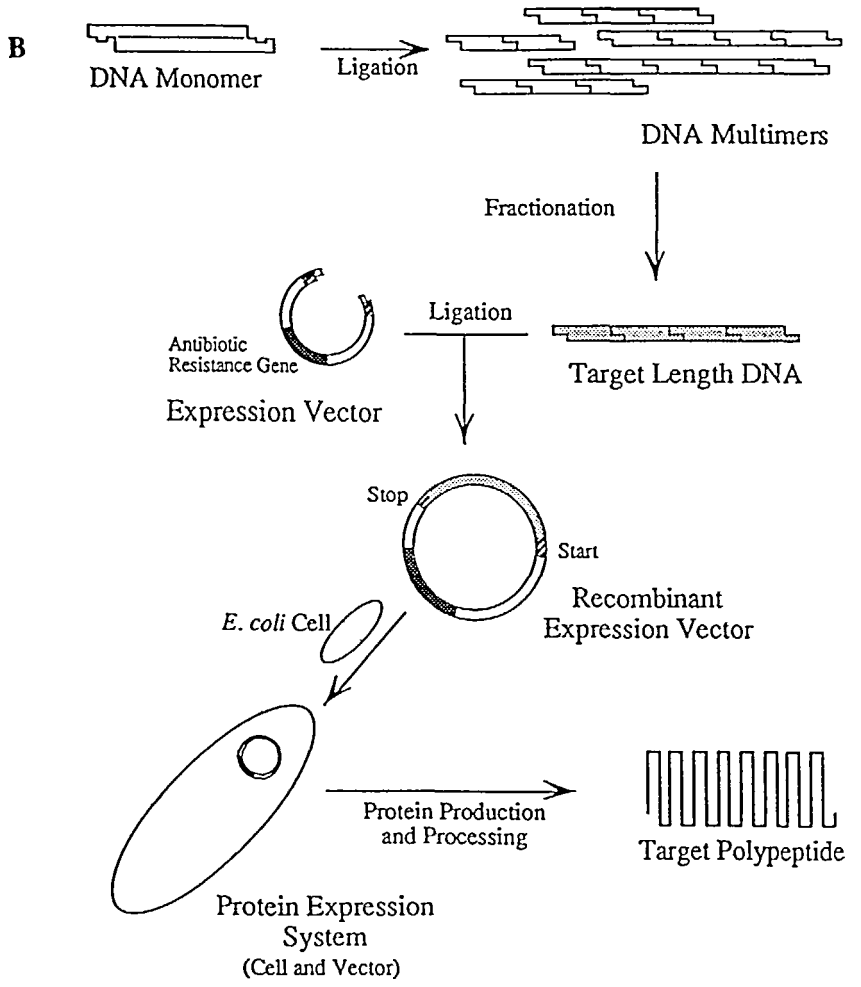


FIG. 1. (A) Schematic of the gene design process. A repetitive amino acid sequence is designed using structural criteria derived in part from structural biology and in part from polymer materials science. A DNA sequence encoding the target protein repeat is then determined using the genetic code; this sequence is referred to as the DNA monomer. (B) The purified DNA monomer is enzymatically polymerized to yield a family of DNA multimers, each encoding different lengths of the target polypeptide. A multimer of the desired length is selected and cloned into an expression vector. In the present study the multimer fractionation and expression vector construction were facilitated by an intermediate cloning step using the "transfer vector" p937.51. The recombinant expression vector, containing the DNA encoding the target polypeptide and appropriate transcription sequence elements, is used to transform an *E. coli* strain containing the requisite protein translation machinery.

Protein Expression. In the genetic constructions just described, target protein sequences **2a-f** are flanked by plasmid-derived N- and C-terminal extensions of 23 and 32 residues, respectively (4). The bacterial host used for expression was *E. coli* strain BL21(DE3)pLysS, which carries a gene encoding bacteriophage T7 RNA polymerase incorporated into the bacterial chromosome under *lacUV5* control. This configuration allows protein production to be initiated by the chemical inducer isopropyl β-D-galactopyranoside (IPTG). The ancillary plasmid pLysS provides a low level of T7 lysozyme, which inhibits T7 RNA polymerase and suppresses the basal level of protein expression. Suppression of basal expression is particularly important for efficient expression of proteins that are toxic to the host and that would cause plasmid loss or preclude accumulation of cell masses sufficient to produce useful quantities of polymer.

MASMTGGQQMGRDPMFKYSRDPM—[(GA)$_n$GE]$_m$—
GARMHIRPGRYQLDPAANKARKEAELAAATAEQ 4

Figure 2 shows the results of expression of proteins **4c-f**. Results for **4a-b** are similar and are not shown. Proteins were separated by PAGE and visualized by autoradiography following in-vivo labeling with ^{3}H-glycine, since the products of interest were found to stain poorly with the dye Coomassie Blue. In each expression (Lanes 1-12, Fig. 2) a strong new protein band appears within 15 minutes of induction. The mobilities of the products were significantly lower than expected on the basis of their molar masses (e.g., for **4c** the expected molar mass is 22,731 Da, as



compared to an apparent value of ~50 kDa inferred from PAGE), in accord with previous observations [2, 18]. Two additional bands appear in each expression (e.g., at ~40 kDa in Lanes 1–3) that are not apparent in uninduced cultures (Lane 13); we propose that these correspond to conformational isomers of the main protein band, though we cannot rule out the possibility of proteolytic cleavage of the product.

Proteins **4a–f** were isolated from cell lysates by a stepwise acidification procedure and recovered by precipitation with ethanol (see Experimental Section). Typical yields of isolated proteins were 20–50 mg/L of fermentation medium, which corresponds to ~15–30% of cellular protein. Treatment with cyanogen bromide followed by extensive washing with water was used to remove the N- and C-terminal fusion fragments. The resulting polymers were soluble in aqueous alkali, aqueous lithium halides, hexafluoroisopropanol, and formic acid.

Verification of Structure. The structures of the target proteins were verified by direct protein sequencing via Edman degradation subsequent to cyanogen bro-

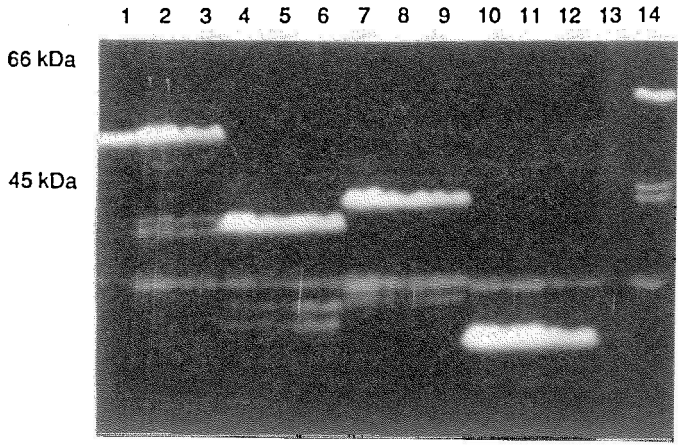


FIG. 2. Expression of proteins **4c–f**. Lanes 1–3: **4c**; Lanes 4–6: **4d**; Lanes 7–9: **4e**; Lanes 10–12: **4f**. In each case successive lanes correspond to cells harvested 15, 30, and 60 minutes respectively, after induction of protein expression. Lane 13: **4c**, no induction. Lane 14: expression of **1**, Z = Glu, x = 36. Molar masses of protein standards are indicated at left.

mide cleavage. No sequence errors were detected up to 60 residues from the N-terminus in any of the samples examined, though in most cases unambiguous sequence identification was limited to 40–50 residues.

The expected β-sheet structures of **2a–f** were confirmed by infrared spectroscopy and wide angle x-ray diffraction methods for films cast from formic acid (Fig. 3A). The β-structure is clearly indicated by the strong amide I vibration at 1625

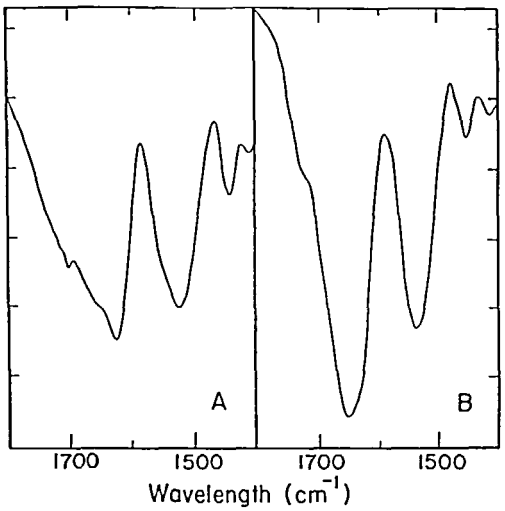


FIG. 3. Fourier transform infrared spectra of **2a**. (A) Film cast from formic acid. (B) Sample precipitated from formic acid by addition of water.

cm^{-1} in the infrared spectrum. Precipitation from formic acid by addition of water, on the other hand, gives a disordered solid characterized by a relatively broad, featureless amide I band centered at 1650 cm^{-1}. Detailed analyses of solid-state structure will be published separately.

ACKNOWLEDGMENTS

This work was supported by a grant from the Polymers and Genetics Programs of the National Science Foundation. We thank Mr. Koji Asahi and Mr. Michio Nomura of the Biochemical Research Laboratories of Kaneka Corporation for the amino acid sequence analysis, and Dr. Howard S. Creel for the use of Fig. 1.

REFERENCES

[1] D. A. Tirrell, M. J. Fournier, and T. L. Mason, *MRS Bull.*, *7*, 23 (1991).
[2] K. P. McGrath, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *J. Am. Chem. Soc.*, *114*, 727 (1992).
[3] G. Zhang, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *Macromolecules*, *25*, 3601 (1992).
[4] A. Parkhe, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *Ibid.*, *26*, 6691 (1993).
[5] E. J. Cantor, H. S. Creel, Y. Deguchi, M. J. Dougherty, S. Kothakota, M. T. Krejchi, K. Matsuki, K. P. McGrath, A. D. Parkhe, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *ACS Symp. Ser.*, *55*, 98 (1994).
[6] R. D. B. Fraser, T. P. MacRae, F. H. C. Stewart, and E. Suzuki, *J. Mol. Biol.*, *11*, 706 (1965).
[7] J. M. Anderson, H. H. Chen, W. B. Rippon, and A. G. Walton, *Ibid.*, *67*, 459 (1972).
[8] (a) H. S. Creel, Ph.D. Dissertation, University of Massachusetts, 1993; (b) M. J. Dougherty, Ph.D. Dissertation, University of Massachusetts, 1992; (c) E. J. Cantor, M.S. Dissertation, University of Massachusetts, 1993.
[9] M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *Science*, In Press.
[10] P. Dreyfuss and A. Keller, *J. Macromol. Sci.—Phys.*, *B4*, 811 (1970).
[11] E. D. T. Atkins, M. Hill, S. K. Hong, A. Keller, and S. Organ, *Macromolecules*, *25*, 917 (1992).
[12] A. J. Geddes, K. D. Parker, E. D. T. Atkins, and E. Beighton, *J. Mol. Biol.*, *32*, 343 (1968).
[13] J. Cappello, J. Crissman, M. Dorman, M. Mikolajczak, G. Textor, M. Marquet, and F. Ferrari, *Biotechnol. Prog.*, *6*, 198 (1990).
[14] F. W. Studier, A. H. Rosenberg, and J. J. Dunn, *Methods Enzymol.*, *185*, 60 (1990).
[15] Media were prepared as described in J. Sambrook, E. F. Frisch, and T. Maniatis, *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, New York, 1989.

[16] U. K. Laemmli, *Nature*, *227*, 680 (1970).
[17] S. Aota, T. Gojobori, F. Ishibashi, T. Maruyama, and T. Ikemura, *Nucleic Acids Res.*, *16*, r315 (1988).
[18] R. C. Beavis, B. T. Chait, H. S. Creel, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *J. Am. Chem. Soc.*, *114*, 7584 (1992).