Rapid Assembly of Synthetic Genes Encoding Protein Polymers

R. Andrew McMillan, Terrence A. T. Lee, and Vincent P. Conticello*

Department of Chemistry, Emory University, Atlanta, Georgia 30322

Received October 23, 1998; Revised Manuscript Received March 23, 1999

ABSTRACT: A general method is described for the rapid assembly of synthetic genes encoding protein polymers based on the Seamless cloning technique. Genes encoding repeats of the elastin-mimetic polypeptides [(Val-Pro-Gly-Val-Gly)$_4$(Val-Pro-Gly-Xaa-Gly)] (Xaa = Lys **1**; Ile, **2**) were constructed using this technique. The use of this method eliminates the dependence of the concatamerization on a limited pool of nonpalindromic restriction endonucleases and reduces the number of subcloning steps. A synthetic gene of approximately 3000 base pairs in length was isolated that encoded a protein polymer based on repeat sequence **1**. An inducible expression of this gene in bacterial cultures of recombinant *E. coli* afforded a 90 kDa protein polymer of **1** in high yield (64 mg/L of bacterial culture). The protein was purified to homogeneity using immobilized metal affinity chromatography. The sequence of the protein polymer was confirmed by N-terminal amino acid sequence analysis and MALDI-TOF mass spectrometry of site-specific proteolytic cleavage fragments.

Introduction

Protein biosynthesis has emerged recently as a useful technique for the preparation of polypeptide-based materials.[1] Although polypeptides have long been considered for materials applications,[2] most investigations have been confined to synthetic homo- and random copoly(α-amino acids) or reconstituted naturally occurring proteins until the advent of techniques for the synthetic overproduction of proteins.[3] The main advantage of protein biosynthesis lies in the ability to directly produce high molecular weight polypeptides of exact amino acid sequence with high fidelity. In contrast to conventional polymer synthesis, biosynthesis of polypeptides proceeds with near-absolute control of macromolecular architecture, i.e., size, composition, sequence, topology, and stereochemistry. Biosynthetic poly(α-amino acids) can be considered as model uniform polymers and may possess unique structures, and, hence, materials properties, as a consequence of their sequence specificity. Recombinant DNA (rDNA) technology and bacterial protein expression have been employed for the biosynthesis of repetitive polypeptides based on naturally occurring protein sequences[4] as well as artificial proteins that have no direct parallels in nature.[5]

The biosynthesis of protein polymers requires the construction of large synthetic genes encoding tandem repeats of target oligopeptide sequences. The most common procedure involves synthesis of double-stranded oligonucleotide cassettes (DNA "monomers") containing nonpalindromic cohesive ends.[1,6] These DNA monomers are oligomerized exclusively head-to-tail by enzymatic ligation such that only the sense (or coding) strand is translated into polypeptide. The DNA concatamers are fractionated, enzymatically joined to an expression vector, and transformed into a suitable expression host, usually a strain of the bacterium *E. coli*. Generally, the target genes are placed under inducible control of a strong promoter, such as the phage T7 *gene 10* promoter, which directs overexpression of the heterologous polypeptide in the host.[7] In many cases good yields (multigram quantities) of the protein polymers are obtained, despite their structural incongruity with the endogenous proteins of the host.[8]

Despite the successful production of numerous synthetic protein polymers, the aforementioned approach has several drawbacks. The most significant disadvantage is the stringent dependence on a limited pool of restriction endonucleases recognizing nonpalindromic cleavage sites. These endonucleases are necessary for the generation of DNA monomers that undergo self-ligation in the correct head-to-tail orientation. Often, superfluous amino acid residues are introduced into the repeat sequence of the target polypeptide as a consequence of this requirement. In addition, such endonucleases can have relaxed specificities in recognition sequences, which increases the likelihood of cleavage of internal sites in expression plasmids. Usually, one or more additional cloning steps is required in the gene assembly to prevent this process. We describe herein a general method for the rapid assembly of synthetic genes encoding repetitive polypeptides that liberates the procedure from dependence on these endonucleases. This procedure is based on the Seamless cloning technique,[9] which utilizes the recognition/cleavage characteristics of the type IIs restriction endonuclease *Eam*1104 I. This technique permits the direct cloning of DNA concatamers into expression vectors, which reduces the number of subcloning steps required in the gene manipulation. Large synthetic genes have been constructed using this procedure (up to 3000 base pairs) that encode repetitive polypeptides of molecular mass greater than 90 kDa.

The efficacy of the seamless cloning procedure resides in two specific properties of endonuclease *Eam*1104 I: (1) cleavage of a DNA duplex at a specific position *downstream* of its recognition site and (2) inhibition of cleavage by incorporation of 5-methyldeoxycytosine into the recognition site of the enzyme. Cleavage of synthetic duplexes with *Eam*1104 I generates 5′ cohesive ends in which the identity of the three base overhangs is independent of the recognition site (Figure 1A). This general procedure can produce any triplet sequence at the 5′-termini of the duplex, which obviates the need

* To whom correspondence should be addressed at vcontic@emory.edu.

10.1021/ma981660f CCC: $18.00
Published on Web 04/28/1999

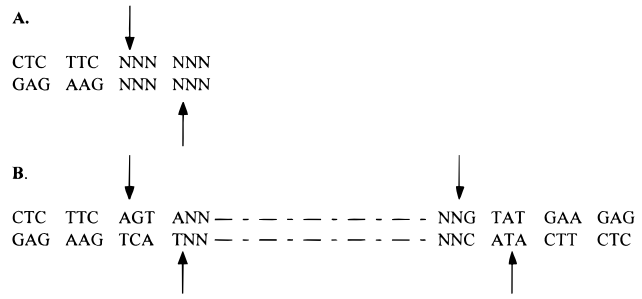

Figure 1. (A) Recognition and cleavage sequence of restriction endonuclease *Eam*1104 I (N represents any of the four possible nucleotides A, T, C, or G). (**B**) Schematic representation of the cleavage pattern for a synthetic DNA duplex flanked by inverted *Eam*1104 I recognition sites, which generates complementary 5′ cohesive ends having the sequence of the valine codon GTA. The top strand of the duplexes is oriented with the 5′ to 3′ direction proceeding from left to right in the sequence.

for an array of endonucleases with unique internal recognition/cleavage patterns. Moreover, the *Eam*1104 I restriction sites are cleaved from the DNA cassette and, hence, are not incorporated into the coding sequence of the DNA monomer. Synthetic duplexes flanked by inverted *Eam*1104 I recognition sites are enzymatically cleaved to generate ligation-competent DNA monomers with nonpalindromic, complementary cohesive ends (Figure 1B). These monomers can be enzymatically joined head-to-tail to generate concatamer libraries.

The second feature of *Eam*1104 I facilitates the insertion of concatameric genes directly into the cloning site of an expression plasmid. Synthetic primers are used to direct the amplification of an appropriate expression plasmid using the inverse polymerase chain reaction (PCR) process.[10] These primers anneal to opposite strands of the plasmid such that their 3′ termini are outwardly oriented on the circular map. Amplification from these primers affords a linear plasmid. The *Eam*1104 I recognition sites in the primers are incorporated into the termini of the plasmid upon amplification. If the PCR process is performed in the presence of 5-methyldeoxycytosine, internal *Eam*1104 I recognition sites in the expression plasmid are protected from enzymatic cleavage, *but the terminal sites derived from the primers are not*. After PCR amplification, the purified, amplified plasmid is incubated with *Eam*1104 I, which cleaves primarily at the terminal sites.[11] The primers are chosen such that the cohesive ends that are generated by *Eam*1104 I cleavage of the amplified plasmid are complementary to those of the multimers, e.g., the GTA sequence of the valine codon as in the current case. In addition, the nucleotide sequences of the primers are designed such that insertion of the concatamers occurs in the correct reading frame for expression of the desired protein polymer.

We have employed the Seamless cloning procedure for the assembly of synthetic genes encoding repetitive polypeptides based on elastin analogues **1** and **2** (Figure 2). The sequences of these polypeptides are modeled on the pentapeptide sequence (Val-Pro-Gly-Val-Gly), which is tandemly repeated 11 times in bovine tropoelastin.[12] Protein polymers based on **1** contain reactive lysine residues once every 25 residues within the repeat sequence. The ϵ-amino groups of the lysine residues on adjacent chains can be chemically cross-linked under controllable conditions into synthetic protein networks.[13] Protein polymers of **1** were designed to examine the

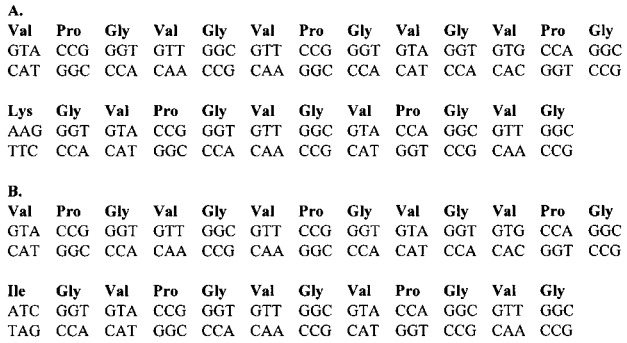

Figure 2. Design of the DNA monomers encoding elastin analogues **1** (A) and **2** (B).

effect of molecular architecture, e.g., cross-link density, on the physical properties of synthetic protein gels. The resulting networks should contain cross-links at well-defined positions determined by the sequence of the repeat. Polypeptide **2** was designed as a non-cross-linkable control for comparison with reactive analogue **1**. The extensive and innovative studies of Urry et al., on synthetic polypeptides based on the pentapeptide repeat [Val-Pro-Gly-Xaa-Gly] of elastin, provide a solid framework for evaluation of structure−property relationships in these elastin-mimetic materials.[14]

Experimental Methods

Materials. *E. coli* strain BLR(DE3), plasmid pET-19b, and T7 promoter and terminator primers were purchased from Novagen (Madison, WI). *E. coli* strains Top10F′ and Top10, plasmid pZErO-1, and Zeocin were obtained from Invitrogen, Inc. (Carlsbad, CA). *Pfu* DNA polymerase, restriction endonuclease *Eam*1104 I, and other reagents for the Seamless cloning procedure were purchased from Stratagene (La Jolla, CA). Sequenase, ThermoSequenase, and α-[^{33}P]-dATP were purchased from Amersham Life Sciences, Inc. (Arlington Heights, IL). All other reagent enzymes were obtained from New England Biolabs (Beverly, MA). Synthetic oligonucleotides and *lacZ* primers were purchased from Genosys Biotechnologies, Inc. (The Woodlands, TX).

General Methods. The procedures for manipulation of DNA, transformation of competent cells, and growth and induction of bacterial cultures were adapted from published literature[15] or from instructions supplied by manufacturers. All reagents for the manipulation of bacteria and DNA were sterilized by either autoclave or passage through a 0.2 μm filter. Enzymatic reactions were performed in the reagent buffers supplied by the manufacturer. PCR amplications were performed in a Perkin-Elmer GeneAmp 2400 thermal cycler. Manual DNA sequence analysis was performed using the ThermoSequenase procedure using radiolabeling with α-[^{33}P]-dATP, and sequencing gels were visualized by autoradiography. Automated DNA sequence analysis was performed on a Perkin-Elmer ABI Prism model 377 DNA sequencer at the Core DNA sequencing facility at Emory University. Polypeptide electrophoresis was performed using 10−15% gradient discontinuous SDS polyacrylamide gels on a PhastSystem apparatus from Pharmacia Biotechnology, Inc., and were visualized using the silver staining process supplied by the manufacturer. Amino acid compositional analysis, MALDI-TOF mass spectrometry, and N-terminal Edman sequencing services were obtained from the Microchemical Facility of the Winship Cancer Center at Emory University. Computer analyses of DNA and protein sequences were performed using the LaserGene program from DNASTAR, Inc. (Madison, WI). Gel images were captured by either a Hewlett-Packard ScanJet 3c scanner or a Kodak DC-40 digitizing camera.

Synthetic Gene Construction. Purified oligonucleotides corresponding to the sense and antisense strands[16] (10 μg each) were mixed in 1X Sequenase buffer (200 μL). The mixture was

heated to 95 °C and cooled gradually to 30 °C. A solution of dNTPs (25 mM each dATP, dTTP, dCTP, dGTP, 2 μL) was added to the reaction mixture followed by Sequenase (100 u). The reaction mixture was incubated for 1 h at 30 °C and then 10 min at 70 °C to deactivate the enzyme. The mixture was extracted with phenol/chloroform, and the synthetic duplexes were concentrated by ethanol precipitation in the presence of PelletPaint (Novagen, Inc.) as a coprecipitant. The synthetic duplexes were puified by preparative agarose gel electrophoresis (2% NuSieveGTG agarose, 1X TBE buffer). The bands were excised from the gel, and the genes were isolated by incubation of the slabs with 1X TE buffer followed by filtration through a 0.45 μm centrifuge filter. The filtrates for each gene were combined, and the DNA was isolated by ethanol precipitation in the presence of a coprecipitant. The synthetic duplexes (3.3 μg of **1**, 5.8 μg of **2**) were digested with *Bam* HI (5 u of enzyme/μg of DNA) in 1X *Bam* HI buffer (total volume equal to 100 times the volume of enzyme solution) for 2 h at 37 °C. The mixture was extracted with phenol/chloroform and the DNA isolated by ethanol precipitation. An identical procedure was employed for cleavage of the synthetic duplexes with *Eco* RI.

Plasmid pZErO-1 (10 μg) was cleaved successively with *Bam* HI and *Eco* RI as above, but the incubation time was decreased to 30 min to prevent degradation of the 5′ terminal phosphate groups. The plasmid was isolated by ethanol precipitation and dried in vacuo for 10 min. The plasmid was dissolved in TE buffer to a final concentration of 10 ng/μL. Separate cohesive end ligations of pZErO-1 with synthetic duplexes **1** and **2** were performed at a 15:1 insert/vector molar ratio (11 ng of genes **1** or **2**, 10 ng of pZErO-1) in the presence of T4 DNA Ligase (1 u) in 1X T4 DNA ligase buffer (10 μL). The mixtures were incubated at 16 °C for 2 h. Aliquots of these mixtures (2 μL) were used to transform competent cells of *E. coli* strain Top10 (50 μL), which were incubated at 0 °C for 30 min. The transformation mixture was cultured on solid low salt LB medium (1% tryptone, 0.5% yeast extract, 0.5% sodium chloride, 1.5% agar, pH 7.5) under zeocin selection (50 μg/mL) for 12 h at 37 °C.

The transformants were checked for the presence of synthetic inserts by colony screening polymerase chain reaction using forward and reverse *lacZ* PCR primers. Positive clones were amplified by growth in low salt LB medium (25 mL) under Zeocin selection at 37 °C for 12 h. Plasmid DNA was isolated using the QIAGEN miniprep procedure (Qiagen, Inc.). Two recombinant plasmids containing each DNA monomer were subjected to sequence analysis. The ThermoSequenase procedure was employed with a reverse *lacZ* sequencing primer and incorporation of radiolabeled α-[^{33}P]-dATP. The sequencing mixtures were run on a 6% polyacrylamide gel containing 50% urea at 55 °C. The gel was transferred onto a Whatman 3 MM filter paper, dried in vacuo, and exposed to a sheet of X-ray film for 48 h. One plasmid of each type with the correct sequence was isolated from this analysis. These plasmids were amplified in *E. coli* strain Top10F′ and isolated using the QIAGEN maxiprep procedure (Qiagen, Inc.).

DNA monomers **1** and **2** were cleaved from pZErO-1 by restriction digestion with *Eam*1104 I. Each recombinant vector (500 μg) was dissolved in 1X enzyme buffer (2.5 mL) in the presence of *Eam*1104 I (5 u of enzyme/μg of DNA). The mixtures were incubated at 37 °C for 12 h, and the enzyme was deactivated at 70 °C for 10 min. The mixtures were extracted with phenol/chloroform, and the cleavage products were concentrated by ethanol precipitation. The DNA monomers were separated from the other DNA fragments by preparative agarose gel electrophoresis (2% NuSieveGTG agarose, 1X TBE buffer). The desired bands were excised from the gels, and the DNA monomers were isolated as described above. This procedure afforded 7.5 and 8.1 μg of DNA monomers **1** and **2**, respectively.

Each DNA monomer (2 μg) was dissolved in 1X ligase buffer (80 μL), and T4 DNA ligase (400 u) was added to the solution; the mixtures were incubated at 8 °C for 12 h. The size distribution of the concatamers was analyzed by agarose gel electrophoresis (1% SeaPlaque agarose, 1X TBE buffer). This concatamer mixture was employed directly in the ligation reaction with the recipient plasmid.

Preparation of the Recipient Plasmid pGTA. Synthetic forward and reverse oligonucleotide primers[17] were employed in the amplification of the recipient plasmid pET-19b. The plasmid (100 ng) was dissolved in an amplification cocktail containing forward and reverse primers (100 ng each), 10X *Pfu* DNA polymerase, dNTP mixture (1 μL, 10 mM each dCTP, dATP, dTTP, and dGTP), and *Pfu* DNA polymerase (2.5 u) in a total volume of 50 μL. Typically, 24 reactions are performed simultaneously in 200 μL sample tubes in a Perkin-Elmer thermal cycler to prepare enough of the linear plasmid for subsequent manipulations. The following temperature program was employed in the initial phase of the PCR process:

cycle 1: 95 °C (3 min), 58 °C (1 min), 72 °C (12 min)

cycles 2−10: 95 °C (45 s), 58 °C (35 s), 72 °C (12 min)

final extension: 72 °C (8 min), cool to 4 °C

After completion of this program, an aliquot (50 μL) of a solution containing 10X *Pfu* DNA Polymerase buffer (5 μL), methylated dNTP mixture (1 μL, 10 mM each 5-methyl-dCTP, dATP, dTTP, and dGTP), and sterile water (44 μL) was added to each tube. The mixtures were subjected to the following program to complete the PCR amplification:

initial hold: 95 °C (4 min)

cycles 11−15: 95 °C (45 s), 58 °C (35 s), 72 °C (12 min)

final extension: 72 °C (8 min), cool to 4 °C

The plasmid DNA was isolated from the primers and dNTPs using a Qiaquick PCR purification kit (Qiagen, Inc.). The eluates of plasmid DNA were combined and isolated by ethanol precipitation. The linearized plasmid pGTA (30 μg) was incubated with *Eam*1104 I (5 u of enzyme/μg of plasmid) in 1X enzyme buffer (1.5 mL) for 12 h at 37 °C. The enzyme was inactivated at 70 °C for 10 min. The mixture was extracted with phenol/chloroform, and the DNA was precipitated with ethanol. The cleaved plasmid was dissolved in 1X NEBuffer 4 and dephosphorylated with calf intestinal alkaline phosphatase (20 u) at 37 °C for 2 h. The mixture was extracted with phenol/chloroform, and the DNA was precipitated with ethanol. The amplified plasmid was analyzed by agarose gel electrophoresis (1% SeaPlaque agarose, 1X TBE buffer) before and after these enzymatic reactions.

Construction of Expression Plasmids. Ligation reactions between linearized plasmid pGTA and the concatameric genes of **1** or **2** were performed as follows: a microcentrifuge tube was charged with the concatamer mixture (40 ng) and *Eam*1104 I-digested, dephosphorylated pGTA (20 ng), and T4 DNA ligase (400 u) in 1X ligase buffer (20 μL). The reaction mixture was incubated at 16 °C for 16 h. *E. coli* strain Top10F′ (50 μL) was transformed with an aliquot (2 μL) of the ligation mixture at 0 °C for 2 h. Transformants were cultured on LB solid media under ampicillin selection (200 μg/mL) for 16 h at 37 °C. Plasmid DNA was isolated from 5 mL cultures of positive transformants using the Qiawell 8 Ultra plasmid kit (Qiagen, Inc.). Plasmids were screened for the presence and size of multimer inserts by double digestion. Plasmid DNA (2 μg) was dissolved in 1X *Bam* HI buffer (50 μL) containing *Nde* I and *Bam* HI (20 u each) and incubated at 37 °C for 4 h. Restriction fragment analysis was performed using agarose gel electrophoresis (1% SeaPlaque agarose, 1X TBE buffer). Recombinant plasmids were isolated that contained an approximately 3000 base pair insert encoding **1** (pRAM**1**) and a 1350 base pair insert encoding **2** (pRAM**2**).

Protein Expression. Competent cells of *E. coli* strain BLR-(DE3) were transformed with pRAM**1** and pRAM**2** at 37 °C for 1 h. These mixtures were cultured on LB solid medium under ampicillin (200 μg/mL) and tetracycline (30 μg/mL) selection for 12 h at 37 °C. Single colonies of positive transformants were used to inoculate 5 mL of LB medium supple-

mented with carbenicillin (50 μg/mL) and tetracycline (30 μg/mL). These cultures were incubated at 37 °C for 12 h. The cells were isolated from the supernatant by centrifugation at 4000*g* and 4 °C for 20 min and resuspended in sterile LB medium (5 mL). This cell suspension was added to fresh LB medium (50 mL) supplemented with carbenicillin (50 μg/mL) until the $OD_{600} \approx 0.1$. The culture was incubated at 37 °C with agitation at 200 rpm until the $OD_{600} \approx 1.0$. IPTG (0.1 M) was added to a final concentration of 1 mM to induce target protein synthesis. Aliquots (1 mL) of the culture were removed immediately prior to induction and at 30 min intervals thereafter. Cells were isolated by microcentrifugation for 1 min and resuspended in sterile water (50 μL). After 3 h ($OD_{600} \approx$ 1.6), the remaining cells were isolated by centrifugation at 4000*g* and 4 °C for 20 min. The protein content of the whole cell lysates was analyzed by 10−15% gradient discontinuous SDS polyacrylamide gel electrophoresis. The protein bands were visualized by silver staining of the gels.

The remaining cell paste from the protein expression (from ca. 40 mL of original culture) was resuspended in 600 μL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole). The cells were lysed by a freeze (−78 °C)/thaw (37 °C) cycle. Solutions of phenylmethylsulfonyl fluoride (100 mM; 3 μL) and egg white lysozyme (10 mg/mL; 2 μL) were added to the lysate. The mixture was incubated at 25 °C for 30 min with agitation at 250 rpm. The lysate was sonicated at 10 s intervals for 3 min at 0 °C. Subsequently, 1 M $MgCl_2$ (1 μL) and benzonase (50 u) were added to the lysate, which was incubated at 25 °C for 30 min with agitation. The mixture was centrifuged at 18 000*g* and 4 °C for 30 min to pellet the cell debris. The soluble extract was charged onto a Ni−NTA agarose centrifuge filter (Qiagen, Inc.). The target proteins were eluted using an imidazole step gradient from 10 to 250 mM in lysis buffer. The eluates were analyzed by 10−15% gradient discontinuous SDS polyacrylamide gel electrophoresis as described above. The polypeptides were desalted and concentrated using Centricon concentrators (MWCO = 30 kD). Lyophilization of the frozen concentrates afforded the pure polypeptides as colorless fibrous solids.

Results and Discussion

Two synthetic genes were designed that encode repeat sequences [(Val-Pro-Gly-Val-Gly)$_4$(Val-Pro-Gly-Xaa-Gly)] (Xaa = Lys **1**; Ile, **2**) based on the pentapeptide repeat of elastin. These genes were used to develop a general method for the synthesis of DNA concatamers that encode protein polymers. The network precursor sequence **1** contains a reactive lysine residue regularly at 25 residue intervals. Polypeptide **2** was designed as a non-cross-linkable control, which differs in the substitution of an isoleucine residue for the lysine residue of **1**. The amino acid composition, sequence, length, and distance between reactive residues of the polypeptide **1** are exactly controlled at the molecular level by the nucleotide sequence and length of the synthetic genes. Most importantly, both synthetic polypeptides retain the sequence pattern of the consensus repeat of elastin, and therefore, should adopt the β-spiral structure postulated by Urry to account for the macromolecular properties of elastin analogues.[18]

The DNA monomers were assembled by enzymatic, mutually primed synthesis from synthetic oligonucleotides with 21 base pair overlapping 3′ ends. After restriction cleavage with *Bam* HI and *Eco* RI, the DNA cassettes encoding **1** and **2** were cloned separately into compatible sites within the polylinker of plasmid pZErO-1. The ligation mixture was used to transform *E. coli* strain Top10. Positive transformants were identified by growth on low salt LB solid media in the presence of the antibiotic zeocin. In contrast to conventional cloning plasmids, such as pUC-18/19, very low nonrecombinant background levels are observed in the ligation of foreign

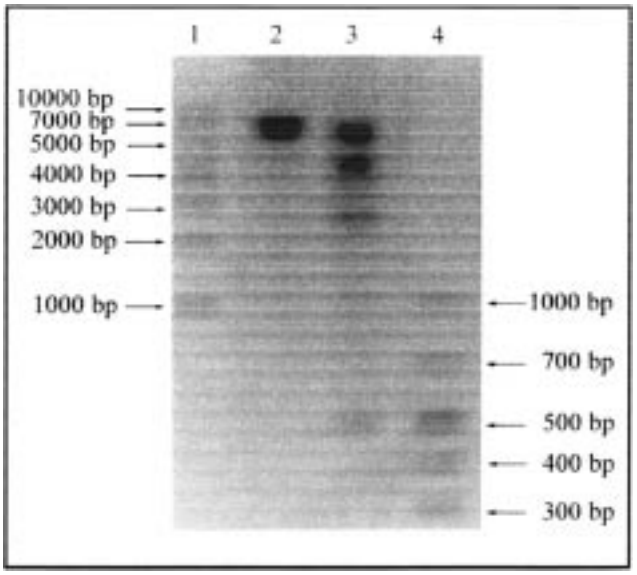

Figure 3. Agarose gel electrophoresis of pGTA before (lane 2) and after (lane 3) digestion with *Eam*1104 I. Lanes 1 and 4 are DNA size standards. DNA bands were stained with ethidium bromide and visualized by UV illumination. The image was captured on a Kodak DC40 digital camera.

DNA into pZErO-1. Insertion of foreign DNA into the multiple cloning site of pZErO-1 interrupts the reading frame of the lethal *ccdB* gene located downstream of that region.[19] This toxic gene is expressed constitutively in *E. coli* strains containing the wild-type *lac* repressor allele (*lacI*) such as Top10. Recombinants are easily scored by growth on solid media, rather than blue/white screening protocols based on α-complementation of β-galactosidase.

The identities of the sense and antisense strands of positive transformants were verified by cycle sequencing of the insert using *lacZ* reverse and forward primers, respectively. The DNA monomers were liberated from the plasmid by restriction digestion with type IIs endonuclease *Eam*1104 I, which generated ligation-competent DNA monomers with complementary 5′ cohesive ends based on valine codon GTA (Figure 1). The monomers were purified by preparative agarose gel electrophoresis and oligomerized enzymatically with T4 DNA ligase. Self-ligation of the DNA monomer afforded the expected distribution of concatameric genes that differed by integral values of the monomer insert length (75 base pairs). The average degree of oligomerization could be approximately controlled through varying the concentration of the DNA monomer and the reaction time. Concatamers of 3750 base pairs in length were observed in the reaction mixture using the conditions described above.

The recipient expression plasmid was amplified using the polymerase chain reaction in the presence of 5-methyldeoxycytosine. Purified linear plasmid was incubated with restriction endonuclease *Eam*1104 I and dephosphorylated with calf intestinal alkaline phosphatase to minimize intramolecular recircularization of the plasmid during the subsequent ligation step. Electrophoretic analysis of the plasmid after cleavage with *Eam*1104 I revealed a ladder of distinct bands due to incomplete incorporation of 5-methyldeoxycytosine into the internal sites of the plasmid pET-19b during the amplification (Figure 3). However, the restriction map clearly indicated the presence of the desired product, which migrated at the expected size (5750 base pairs)

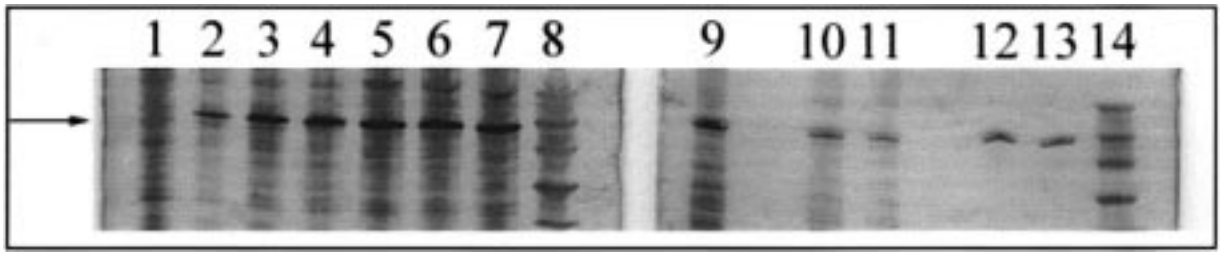

Figure 4. Biosynthesis (left) and purification (right) of the 90 kDa elastin analogue analyzed by denaturing polyacrylamide gel electrophoresis (10−15% acrylamide gradient, visualized by silver staining). Lanes 1−7: time course of target protein expression at 0, 30, 60, 90, 120, 150, and 180 min after induction. Lane 9: soluble lysate of induced *E.coli* expression strain BLR(DE3)-pRAM1. Lanes 10−13: protein fractions eluted from immobilized metal affinity chromatography of the soluble lysate on nickel-NTA resin (Qiagen, Inc.) using an imidazole step gradient in 50 mM phosphate buffer (pH 8.0) containing 300 mM NaCl. Lanes 10 and 11: 10 mM imidazole washes. Lanes 12 and 13: 250 mM imidazole elution fractions. Lanes 8 and 14 contain protein molecular weight standards of 35, 50, 90, and 150 kDa.

versus the duplex DNA standards. Computer analysis of the theoretical restriction pattern generated from nonmethylated pET-19b upon cleavage with *Eam*1104 I revealed that none of the possible products from internal cleavage of the plasmid had cohesive ends that were complementary to the DNA multimers. The presence of these dephosphorylated DNA fragments does not interfere with the subsequent insertion of the multimers into the modified cloning site of amplified pET-19b, renamed pGTA.

The population of concatamers was cloned into dephosphorylated pGTA by incubation at 16 °C in the presence of T4 DNA ligase. *E. coli* strain Top10F′ was transformed with this mixture and cultured on LB solid media under ampicillin and tetracycline selection. Numerous colonies arose from the ligation reaction between the concatamer mixture and pGTA, while few were observed from plasmid-only control ligations. Recombinant plasmids were screened for the presence and size of multimeric insert using *Nde* I/*Bam* HI restriction analysis.[20] Concatameric genes having sizes of 3000 and 1350 base pairs were isolated that encode repetitive polypeptides based on elastin-mimetic sequences **1** and **2**, respectively. The isolation of larger concatamers is possible using this process, especially for **2**, on the basis of the size distribution of the multimer mixture. The forward and reverse sequences of the recombinant plasmid pRAM**1** and pRAM**2** were verified through five repeats of **1** or **2** in both sense and antisense directions using automated DNA sequence analysis.

Plasmid pRAM**1**, which encodes a repetitive polypeptide of **1** with approximately 90 kDa molecular mass, was isolated and evaluated for level of target protein expression. The *E. coli* expression strain BLR(DE3), which contains a copy of the T7 RNA polymerase gene under *lac* control on a dormant λ prophage localized on the bacterial chromosome, was transformed with pRAM**1**. A small-scale culture of the recombinant expression strain in LB medium under carbenicillin selection was employed to assess the presence and level of accumulation of the target protein fusion. The cells were incubated at 37 °C until midlog growth was achieved, and expression of the target polypeptide was induced by addition of IPTG to a final concentration of 1 mM in the medium. Analysis of the soluble versus insoluble fractions of the whole cell lysate indicated that the protein accumulates at reasonably high levels in vivo and segregates almost exclusively in the soluble fraction of the cell lysate (Figure 4).

Expression from the plasmid pRAM**1** affords the target protein as a translational fusion to an amino terminal decahistidine tag, which is encoded upstream of the open reading frame of the synthetic gene. The presence of this oligopeptide tag provides a method for purification of the fusion protein using immobilized metal affinity chromatography.[21] The fusion protein was purified to homogeneity on a microscale using Ni−NTA agarose affinity matrix (Figure 4). The target polypeptide was selectively eluted at high concentrations of imidazole (250 mM) after the endogenous host proteins had been washed off the resin. A similar procedure was employed for the microscale synthesis and purification of polypeptides based on sequence **2**, using expression from plasmid pRAM**2**.

The identity of the 90 kDa protein polymer of **1** was confirmed by N-terminal amino acid sequence analysis of proteolytic fragments derived from cleavage with Endoproteinase Lys-C. The major cleavage product was subjected to 25 cycles of automated Edman degradation, which afforded the sequence GVPGVGVPGVGVPGV-GVPGVGVPGK. This product corresponds to the major fragment expected for digestion of poly-**1** with Endoproteinase Lys-C, which cleaves after the carboxy terminus of lysine residues in the polypeptide. The identity of this cleavage fragment is supported by observation of an $[M + H]^+$ peak at 2092.3 *m*/*z* (expected: 2093.5 *m*/*z*) in the MALDI-TOF mass spectrum of this protein fragment.

Protein polymers of **1** and **2** have been subsequently expressed under large-scale, batch fermentation conditions, which afforded unoptimized yields of 64 and 96 mg purified polypeptide per liter of culture in LB medium, respectively. These proteins were isolated and purified using the procedure of Urry et al.,[4a] which takes advantage of the reversible, temperature-dependent precipitation of elastin-mimetic proteins. Although the N-terminal decahistidine tag provides a general method for polypeptide purification, reversible precipitation procedure is preferred for large-scale preparations of elastin-mimetic polypeptides due to simplicity and freedom from lengthy and expensive chromatography steps.

The N-terminal leader sequences can be removed from the fusion proteins by cyanogen bromide cleavage to afford pure protein polymers of **1** and **2**.[22] The ^{1}H NMR spectra of the purified polypeptides in water or perdeuteriodimethyl sulfoxide are consistent with the formulation of the sequences as polymers of **1** and **2**. The details of the large-scale expression and purification of these polypeptides will be reported in a subsequent communication.[13]

These results suggest that the Seamless cloning technique can be successfully adapted for the rapid

assembly of synthetic genes encoding protein polymers. Synthetic genes were seamlessly fused into a recipient plasmid in the correct reading frame for protein expression. No restrictions are placed on the sequence of the junction zones other than those specified in the design of the synthetic gene. This feature permits considerable flexibility in gene design and cloning strategy. High levels of recombinant protein accumulation are observed upon expression from a phage T7 promoter in a conventional bacterial expression system. We have recently employed this procedure for the assembly of genes encoding polypeptides based on an 80 amino acid residue, elastin-mimetic repeat sequence **3**. This repeat has a segmented structure consisting of separate elastomeric and cross-linkable domains based on the sequence of naturally occurring mammalian tropoelastins. Polypeptides based on this sequence can be chemically cross-linked in a manner analogous to polymers of **1** and may offer insight into the self-assembly of tropoelastin prior to the enzymatic cross-linking reaction.[13]

Val-Pro-Gly-Val-Gly-(Ile-Pro-Gly-Val-Gly)$_4$−
Val-Pro-Gly-Ala-Gly-Thr-Pro-(Ala)$_5$-Lys-
(Ala)$_{10}$-Lys-Ala-Ala-Gln-Phe-Gly-Leu-
Gly-Pro-Gly-Val-Gly-(Ile-Pro-Gly-Val-Gly)$_4$ (**3**)

References and Notes

(1) (a) Heslot, H. *Biochimie* **1998**, *80*, 19−31. (b) Ferrari, F. A.; Cappello, J. In *Protein-Based Materials*; McGrath, K., Kaplan, D., Eds.; Birkhauser: Boston, 1997; pp 37−60. (c) Cappello, J.; Ferrari, F. In *Plastics from Microbes: Microbial Synthesis of Polymers and Polymer Precursors*; Mobley, D. P., Ed.; Hanser/Gardner Publications: Munich, 1994; pp 35−92. (d) Tirrell, J. G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Chem. Eng. News* **1994**, *72*, 40−51. (e) O'Brien, J. P. *Trends Polym. Sci.* **1993**, *1*, 228−232.

(2) Hudson, S. M. In *Protein-Based Materials*; McGrath, K., Kaplan, D., Eds.; Birkhauser: Boston, 1997; pp 313−337.

(3) Deming, T. J. *Adv. Mater.* **1997**, *9*, 299−311.

(4) For leading references in the biosynthesis of repetitive polypeptides based on natural protein materials, please see the following: (a) Elastin: McPherson, D. T.; Xu, J.; Urry, D. W. *Protein Expr. Purif.* **1996**, *7*, 51−7. McPherson, D. T.; Morrow, C.; Minehan, D. S.; Wu, J.; Hunter, E.; Urry, D. W. *Biotechnol. Prog.* **1992**, *8*, 347−352. (b) *N. clavipes* dragline silk: Prince, J. T.; McGrath, K. P.; DiGirolamo, C. M.; Kaplan, D. L. *Biochemistry* **1995**, *34*, 10879−10885. Fahnestock, S. R.; Irwin, S. L. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23−38. Fahnestock, S. R.; Bedzyk, L. A. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 33−39. Lewis, R. V.; Hinman, M.; Kothakota, S.; Fournier, M. J. *Protein Exp. Purif.* **1996**, *7*, 400−406. (c) *M. edulis* mussel adhesive: Salerno, A. J.; Goldberg, I. *Appl. Microbiol. Biotechnol.* **1993**, *39*, 221−226. Masiliamani, D.; Goldberg, I.; Salerno, A. J.; Oleksiuk, M. A.; Unger, P. D.; Piascik, D. A.; Bhattacharjee, H. R. In *Biotechnology and Polymers*; Gebelein, C. G., Ed.; Plenum Press: New York, 1991; pp 245−264. Strausberg, R. L.; Anderson, D. M.; Filpula, D.; Finkelman, M.; Link, R.; McCandliss, R.; Orndorff, S. A.; Strausberg, S. L.; Wei, T. In *Adhesives From Renewable Resources*; Hemingway, R. W., Conners, A. H., Branham, S. J., Eds.; ACS Symposium Series 385; American Chemical Society: Washington, DC, 1989; pp 453−483. (d) Collagen: Goldberg, I.; Salerno, A. J.; Patterson, T.; Williams, J. I. *Gene* **1989**, *80*, 305−314. (e) Glutenin: Anderson, O. D.; Kuhl, J. C.; Tam, A. *Gene* **1996**, *174*, 51−58. (f) *B. mori* silk fibroin: Cappello, J.; Crissman, J. W.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198−202. (g) Adenovirus fiber spike protein: O'Brien, J.; Hoess, R.; Gardner, K.; Lock, R.; Wasserman, Z.; Weber, P.; Salemme, F. In *Silk Polymers-Materials Science and Biotechnology*; Kaplan, D. L., Adams, W. W., Farmer, B. L., Viney, C., Eds.; American Chemical Society: Washington, DC, 1994; pp 104−117.

(5) (a) Petka, W. A,; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, *281*, 389−392. (b) Yu, S. M.; Conticello, V. P.; Zhang, G.; Kayser, C.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Nature* **1997**, *389*, 167−170. (c) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427−1432. (d) Zhang, G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1992**, *25*, 3601−3603. (e) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727−733. (f) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213−1214.

(6) Tirrell, D. A.; Fournier, M. J.; Mason, T. L. *MRS Bull.* **1991**, *16*, 23−28.

(7) Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. *Methods Enzymol.* **1989**, *185*, 60−89.

(8) Panitch, A.; Matsuki, K.; Cantor, E. J.; Cooper, S. J.; Atkins, E. D. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1997**, *30*, 42−49.

(9) Padgett, K. A.; Sorge, J. A. *Gene* **1996**, *168*, 31−35.

(10) Triglia, T.; Peterson, M. G.; Kemp, D. S. *Nucleic Acids Res.* **1988**, *16*, 8186.

(11) A percentage of internal sites on the recipient plasmid will incorporate dCTP rather than 5-methyl-dCTP and will be cleaved by *Eam*1104 I. Typically, the addition of 5-methyl-dCTP is delayed until the last five cycles of the PCR amplification of the plasmid to minimize thermal deamination of the nucleotide.

(12) Urry, D. W. *J. Protein Chem.* **1988**, *7*, 1−34.

(13) McMillan, R. A.; Caran, K.; Apkarian, R.; Conticello, V. P., manuscript in preparation.

(14) Urry, D. W. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 819−841.

(15) Sambook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989.

(16) The sequences of the oligonucleotides employed in the assembly of synthetic genes encoding **1** and **2** are given below with the region of 3′ complementarity in boldface. Sense strand for **1**: 5′-AGTTAGGATCCCTCTTCAGTACCGGGTGTTGGCGTTCCGGGTGTA**GGTGTGCCAGGCAAGGGTGTA**-3′. Antisense strand for **1**: 5′-AGTTAGAATTCCTCTTCATACGCCAACGCCTGGTACGCCAACACCCGG**TACACCCTTGCCTGGCACACC**-3′. Sense strand for **2**: 5′-AGTTAGGATCCCTCTTCAGTACCGGGTGTTGGCGTTCCGGGTGTA**GGTGTGCCAGGCATCGGTGTA**-3′. Antisense strand for **2**: 5′-AGTTAGAATTCCTCTTCATACGCCAACGCCTGGTACGCCAACACCCGG**TACACCGATGCCTGGCACACC**-3′.

(17) The sequences for the forward and reverse primers for the PCR amplification of pET-19b are given below with the regions of complementarity to the plasmid in boldface. Forward primer: 5′-AGTTACTCTTCAGTATA**AGGATCCGGCTGCTAACAAAG**-3′. Reverse primer: 5′-AGTTACTCTTCATAC**CATATGCTTGTCGTCGTCGTC**-3′.

(18) (a) Wasserman, Z. R.; Salemme, F. R. *Biopolymers* **1990**, *29*, 1613−1628. (b) Chang, D. K.; Urry, D. W. *Chem. Phys. Lett.* **1988**, *147*, 395−400.

(19) Bernard, P.; Gabant, P.; Bahassi, E. M.; Couturier, M. *Gene* **1994**, *148*, 71−74.

(20) Colony screening PCR is the preferred method for the identification of multimer genes in the recombinant plasmids as it avoids parallel preparations of plasmid DNA from multiple colonies. We use a procedure recommended by Novagen, Inc., for amplification of the polylinker region using commercially available T7 promoter and terminator primers. The details of this procedure are described in their pET System Manual, which is available in hardcopy from Novagen, Inc., or in electronic HTML or PDF format at URL: (http://www.novagen.com/html/techlit.html#pet).

(21) Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stuber, D.; Henco, K. *Methods Mol. Biol.* **1994**, *31*, 371−387.

(22) Amino acid compositional analysis of the 90 kDa protein after cyanogen bromide cleavage is in excellent agreement with the expected composition of a polymer based on repeats of **1** and **2**.

amino acid	calc for **1**	obs for **1**	calc for **2**	obs for **2**
glycine	40%	39.9%	40%	39.1%
proline	20%	19.5%	20%	19.5%
valine	36%	36.5%	36%	36.8%
lysine	4%	4.1%		
isoleucine			4%	4.2%

MA981660F