Article

A New Cloning Method for the Preparation of Long Repetitive Polypeptides without a Sequence Requirement

Jong-In Won, and Annelise E. Barron

Macromolecules, **2002**, 35 (22), 8281-8287 • DOI: 10.1021/ma020892i

A New Cloning Method for the Preparation of Long Repetitive Polypeptides without a Sequence Requirement

Jong-In Won and Annelise E. Barron*

Department of Chemical Engineering, Northwestern University, Evanston, Illinois 60208

Received June 10, 2002; Revised Manuscript Received August 9, 2002

ABSTRACT: We describe and illustrate a new cloning method for the production of long, size-controlled DNA concatemers, which can be expressed as protein polymers without an amino acid sequence requirement. Synthetic genes encoding different numbers of repeats of the amino acid sequence -(Gly-Lys-Gly-Ser-Ala-Gln-Ala)$_3$- were constructed using this cloning technique. The method enables the production of extremely long synthetic genes (e.g., up to 48 DNA repeats, 3024 bp), in a highly controlled fashion by providing higher-order multimers via the reconcatemerization of pre-multimerized genes and also eliminates the specific sequence requirement of the DNA monomer for restriction endonuclease cleavage. Moreover, the effect of intramolecular cyclization, which has a high probability of occurring during the self-ligation reaction and prevents the cloning of long DNA concatemers, is minimized by this approach. DNA multimers encoding a ladder of four differently sized protein polymers (6, 12, 24, and 48 repeats of a 63-bp segment) were expressed in *E. coli* upon induction with IPTG. Two of the resulting proteins, composed of 277 and 529 amino acid monomers, and each comprising a 10X histidine tag, were purified by immobilized metal affinity chromatography. During the course of the work, the sequence of the DNA concatemers was analyzed and confirmed to be correct by capillary electrophoresis, and the obtainment of the desired protein polymers was confirmed by MALDI−TOF mass spectrum analysis and amino acid compositional analysis. The resultant protein polymers are shown by CD to adopt a random-coil conformation in aqueous solution and to be water-soluble between 20 and 95 °C. This polymer has lysine residues spaced evenly along the polymer backbone, which can be utilized either as derivatizable groups or as cross-linking sites for tissue engineering applications.

Introduction

Continuing research on nonnatural, repetitive polypeptides or "protein polymers" produced via genetic engineering has provided various biomaterials that serve as analogues to extracellular matrix proteins[1,2] and spider silk proteins[3−5] or as precursors to smectic liquid crystals.[6,7] Work in this area has not only enabled the fundamental study of biomaterials that have well-defined higher order structure but also shown the practical application of these polypeptides to the development of biomimetic materials which can replace natural tissue.[2,8]

Cappello et al.[9] and McGrath et al.[10,11] were the first to demonstrate that repetitive synthetic genes can be produced by head-to-tail enzymatic linkage (concatemerization) and that subsequently the corresponding polypeptides can be expressed in an *E. coli* system. This result showed the feasibility of generating useful quantities of repetitive artificial proteins using a biological system. Since that time, various repetitive polypeptides have been produced on the basis of this approach.[12−14] However, despite the successful production of numerous synthetic protein polymers, obtaining long DNA concatemers of a particular desired chain length remains a challenge to current researchers. Most of the recombinant plasmids produced from published cloning procedures and hosted in *E. coli* contain short DNA concatemers, so that generally the yield of long concatemers is too low to be of practical use in subsequent molecular biological manipulation. Thus, it is difficult to obtain a long protein polymer with a desired chain length. Another challenge in this method is that concatemerization of synthetic DNA monomers is limited by the small pool of nonpalindromic restriction endonucleases available (e.g., *Ban*I). Thus, a DNA monomer requires a 3′ and/or 5′ terminal sequence determined by the available nonpalindromic restriction enzymes, and this requirement restricts the production of diverse and arbitrary polypeptide sequences.

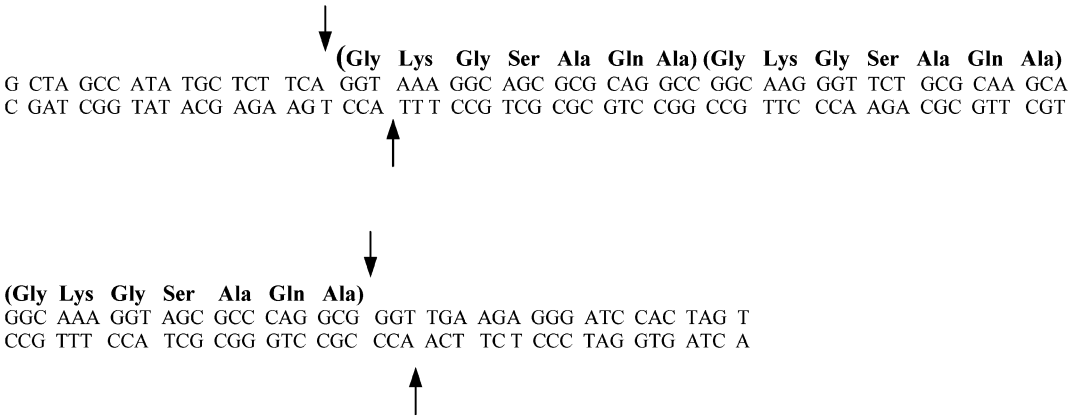
To approach the first limitation, Prince et al.[15] and Winkler et al.[16] have used two different restriction enzymes (*Nhe*I and *Spe*I) to generate concatemers that maintain the desired, correct DNA sequence orientation over the entire concatemerized gene. This approach enables control of the size of concatemerized genes by providing higher-order multimers through the reconcatemerization of pre-multimerized genes. However, to produce concatemers via this method, a DNA monomer must contain a *Nhe*I cleavage site at one terminus and a *Spe*I cleavage site at the other terminus; this requirement limits the peptide sequence which can be expressed.

In 1999, McMillan et al.[17] developed the seamless cloning technique, which eliminates the dependence of the concatemerization on a limited pool of nonpalindromic restriction endonucleases. This technique obviates the sequence requirement of the DNA monomer for nonpalindromic restriction endonuclease cleavage, thus allowing the production of a protein of any sequence. However, it is still difficult to obtain a long concatemer insert of desired length via this technique, since it is not designed to accommodate reconcatemerization of pre-multimerized genes. To facilitate the cloning of long concatemers, the concatemerization of a synthetic gene monomer must have a sufficient yield of long concatemers, which is experimentally difficult to achieve in any synthetic protein polymer cloning technique.

* Corresponding author: phone (847) 491-2778, Fax (847) 491-3728, e-mail a-barron@northwestern.edu.

10.1021/ma020892i CCC: $22.00
Published on Web 09/19/2002

Scheme 1. DNA Sequence of Gene 1, Which Was Designed de Novo, and the Amino Acid Repeat That the Central Region Encodes



Very recently, Meyer et al.[18] developed recursive directional ligation (RDL) for the synthesis of repetitive polypeptides of a specified chain length. A goal of this approach is coincident with that of ours in that both methodologies involve controlled recursive multimerization of a DNA monomer until a gene of a desired length is obtained and that other genes encoding different polypeptide sequences can be incorporated to form multidomain repetitive polypeptides (e.g., block copolymers). However, the PCR-based methodologies described below greatly facilitate the overall process by reducing cost and the time required for more intensive cloning strategies.

Here we present a modified cloning method that combines the advantages of the previously published approaches described above and enables the production of long synthetic genes (so far, in our hands, up to 48 repeats of a 63-bp DNA "monomer") in a controlled and reproducible fashion. Further, this novel method allows any triplet sequence at the 5′-terminus of the DNA monomer, which obviates the specific sequence requirement of any restriction enzyme. To illustrate the method, we present a detailed synthetic protocol for the preparation of long repetitive polypeptides in pure form without an amino acid sequence requirement.

Materials and Methods

Materials. *E. coli* strain BLR(DE3) and plasmid pET-19b were purchased from Novagen (Madison, WI). *E. coli* strain Top10, plasmid pUC18, and Ni-chelating resin were obtained from Invitrogen, Inc. (Carlsbad, CA). *Pfu* DNA polymerase and restriction endonuclease *Eam*1104 I were purchased from Stratagene (La Jolla, CA), and *Sap* I restriction enzyme was obtained from New England Biolabs (Beverly, MA). *Taq* DNA polymerase, T4 DNA ligase, T4 DNA kinase, and all other restriction enzymes were purchased from Promega (Madison, WI). Synthetic oligonucleotides were supplied by Oligos Etc. (Wilsonville, OR), and all other primers were obtained from the Northwestern University Biotechnology Laboratory (Chicago, IL).

General Methods. The procedures for growth and induction of bacterial culture, DNA manipulation, and transformation conditions were adopted from published literature[19] or from instructions provided by manufacturers. DNA sequence analyses were performed in our laboratory with a MegaBACE 96-capillary DNA sequencing instrument (Amersham-Pharmacia BioSciences, Sunnyvale, CA). Protein electrophoresis was performed on 12% discontinuous SDS polyacrylamide gels with bands visualized via silver staining or Coomassie Blue staining.

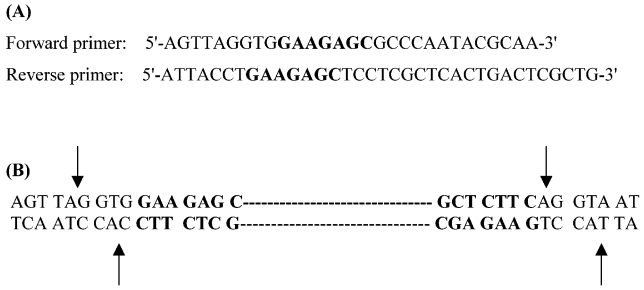
Preparation of Synthetic DNA. A single-stranded, synthetic oligonucleotide (104 bases, gene **1**, see Scheme 1), which encodes three tandem repeats of -(Gly-Lys-Gly-Ser-Ala-Gln-Ala)-, was amplified via PCR with two corresponding oligonucleotide primers [5′-GCTAGCCATATGCTCTTCAGG-3′; 5′-ACTAGTGGATCCCTCTTCAAC-3′]. This reaction was performed using an MJ Research DNA thermal cycler with 30 cycles at 95 °C for 1 min (denaturing), 55 °C for 1 min (annealing), and 72 °C for 2 min (elongation). A high concentration of dsDNA was obtained from ssDNA by PCR, and these dsDNAs were digested with *Eam*1104 I endonuclease (37 °C, 12 h) to produce DNA monomers with cohesive termini. The 63-bp DNA monomer was then fractionated and recovered via 3% agarose gel electrophoresis onto a (diethylamino)ethyl (DEAE)−cellulose membrane. The isolated DNA monomer was purified with standard phenol/chloroform extraction and ethanol precipitation.

Generation of Multimers. Purified DNA monomer was dissolved in 1X ligase buffer, and T4 DNA ligase was added to the solution. In the optimized ligation protocol, the mixture was incubated at 16 °C for 16 h. The size distribution of the concatemers was analyzed via 1.5% agarose gel electrophoresis.

Preparation and Construction of the Protein Expression Plasmid, pGGT. All procedures that were used to construct the recipient plasmid pET-19b, or pGGT, were performed on the basis of the seamless cloning technique of McMillan et al.[17] To protect the internal *Eam*1104 I recognition sites in plasmid pET-19b from enzymatic cleavage, the PCR process was performed in the presence of 5-methyldeoxycytosine. The recipient plasmid pGGT was generated by PCR with primers containing *Eam*1104 I recognition sites as well as a stop codon [forward primer: 5′-AGTTACTCTTCAGGTTAAGGATCCGGCTGCTAACAAAG-3′; reverse primer: 5′-AGTTACTCTTCAACCCATATGCTTGTCGTCGTCGTC-3′]. After PCR amplification, the linearized plasmid was incubated with *Eam*1104 I endonuclease for 12 h at 37 °C to generate cohesive termini complementary to those of the DNA multimers. The *Eam*1104 I recognition sites at both termini, which were derived from the primers, were cleaved off by the enzyme reaction whereas the internal recognition sites were protected from enzymatic cleavage due to the prior incorporation of 5-methyldeoxycytosine. The linearized plasmid pGGT was then isolated from the reaction mixture by gel electrophoresis with a DEAE−cellulose membrane. The plasmid was purified with standard phenol/chloroform extraction and precipitated with ethanol. The cohesive-ended plasmid was dephosphorylated with calf intestinal alkaline phosphatase (CIP) for 2 h at 37 °C to minimize intramolecular recircularization of the plasmid during the subsequent ligation step. The plasmid was purified again with phenol/chloroform extraction and ethanol precipitation.

The ligation reaction between pGGT (20 ng) and the concatemer mixture (60 ng) was performed with T4 DNA ligase at 16 °C for 16 h. *E. coli* strain Top10 (50 μL) was transformed with an aliquot (2 μL) of the ligation mixture using an electroporator from BioRad (Hercules, CA). Transformants

Scheme 2. Primers Used for the Amplification of PUC18 Plasmid by PCR (A) and Schematic Representation of the Cleavage Pattern of PUC18 Recipient Vector (B); Bold Letters and Arrow Represent *Sap* I Recognition Sites and Cleavage Sites, Respectively



were cultured overnight at 37 °C on LB solid media under carbenicillin selection (50 μg/mL). Each colony was isolated by toothpick and inoculated into a sterile tube containing 2 mL of LB liquid media under carbenicillin selection. After overnight culture, plasmids were isolated from the cells using Quantum Prep Kits (BioRad). Isolated plasmids were screened for the presence and size of multimer inserts by double digestion with the restriction enzymes *Nde*I and *Bam*HI.

Preparation of the pUC18 Recipient Vector. The plasmid pUC18 was obtained from Invitrogen and linearized by *Sap* I digestion (37 °C, 6 h). The linearized vector was amplified by PCR with *Pfu* DNA polymerase and synthetic forward and reverse primers, which anneal to the 3′ and 5′ termini of the vector (Scheme 2A). Because both primers encode the *Sap* I recognition site, which lies downstream of the cleavage sites, both recognition sites remain at the termini of the linear vector after *Sap* I digestion (Scheme 2B). After PCR amplification, the linearized vector was incubated with *Sap* I restriction enzyme for 6 h at 37 °C to generate cohesive ends complementary to those of the multimers. The pUC18 recipient vector was isolated from the reaction mixture by gel electrophoresis with a DEAE−cellulose membrane. The vector was purified with phenol/chloroform extraction and precipitated with ethanol. The cohesive-ended vector was dephosphorylated with CIP for 2 h at 37 °C. The vector was purified again with phenol/chloroform extraction and ethanol precipitation. Ligation and transformation were performed following the same procedure as described above. The size of cloned concatemers was determined by *Sap* I digestion of the recombinant plasmid and gel electrophoresis with comparison to DNA size standards.

Protein Expression and Purification. Competent cells of *E. coli* strain BLR(DE3) were transformed with the recombinant protein expression plasmid using an electroporator. These mixtures were cultured on LB solid media under carbenicillin (50 μg/mL) and tetracycline (30 μg/mL) selection overnight at 37 °C. Single colonies of positive transformants were used to inoculate 5 mL of LB medium supplemented with carbenicillin (50 μg/mL) and tetracycline (30 μg/mL). The culture was incubated at 37 °C with agitation at 250 rpm until the $OD_{600} \approx 0.6$. Protein synthesis was induced with isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 1 mM. After an additional 3 h, the cells were isolated by centrifugation at 6000*g* and 4 °C for 10 min. The protein content of whole cell lysates was analyzed by 12% discontinuous SDS polyacrylamide gel electrophoresis (SDS-PAGE).

The cell paste was resuspended in 5 mL of the lysis buffer (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride; pH 7.8) per gram of wet cell weight. The cells were lysed by several freeze (−80 °C)/thaw (37 °C) cycles and subsequently sonicated at 30 s intervals for 3 min. The lysate was centrifuged at 10 000*g* and 4 °C for 30 min to pellet the cell debris, and the supernatant was loaded onto Ni-chelating resin (Probond, Invitrogen). The target proteins were selectively eluted at low pH (pH 4.0) after the endogenous host proteins had been washed off from the resin extensively with high-pH buffers (pH 7.8, pH 6.0, and pH 5.3 buffers). The eluents were analyzed by 12% discontinuous SDS-PAGE. Purified proteins were dialyzed against deionized water (MWCO = 10 000) (Pierce) for 3 days and lyophilized to a dry powder.

Protein Analysis and Characterization. The purified proteins were dissolved in water and analyzed by gradient reversed-phase HPLC on C18 packing (Vydac, 5 μm, 300 Å, 2.1 × 250 mm). A linear gradient of 0−95% B in A was run over 50 min at a flow rate of 0.1 mL/min (solvent A = 0.1% TFA in water, solvent B = 0.1% TFA in acetonitrile) at 60 °C; peaks were detected by UV absorbance at 220 nm. A Voyager-DE PRO mass spectrometer (Analytical Services Laboratory, Northwestern University) was used for MALDI−TOF analysis, and sinapinic acid was used as the matrix. Amino acid compositional analyses[5] were obtained from the Analytical Services Laboratory at Northwestern University. Circular dichroism (CD) spectra were recorded on a Jasco J-715 spectropolarimeter (Keck Biophysics Facility, Northwestern University). Protein samples were dissolved in sterile water at a concentration of 1 mg/mL. Spectra were obtained from 260 to 190 nm as a scanning range. The temperature-dependent phase behavior of the repetitive polypeptides was monitored at 254 nm as a function of temperature on a Cary 500-UV/vis/NIR spectrometer (Keck Biophysics Facility, Northwestern University). Spectra were recorded between 20 and 95 °C with a temperature ramping rate of 10 °C/min.

Results and Discussion

The longer-term goal of this work is to create long, repetitive water-soluble polypeptides for diverse biomaterials and tissue engineering applications. A key feature desired in these protein polymers is the presence of reactive amino acid side chains at regular intervals along the backbone, which can be subsequently derivatized with bioactive factors such as short amino acid sequences.[20−22] Alternatively, these reactive groups could be utilized in a biomaterial cross-linking reaction, as has been shown by Conticello's group[23] for elastin−mimetic polypeptides. However, unlike elastin-like protein polymers, we desired polypeptides that would remain strictly water-soluble over a wide range of temperatures and pH. Hence, we wanted no LCST-like behavior and no secondary structure (e.g., β-sheet) that would lead to aggregation and precipitation in water. Moreover, we desired polypeptides of a truly polymeric nature, longer than 500 monomers. Such polymers could serve as a generally useful scaffold for biomaterials and tissue engineering application.

Several attempts were made to clone a long DNA concatemer gene successfully. A description of these experiments, which included use of the standard seamless cloning method, seamless cloning after fractionating the DNA to isolate long concatemer genes, and seamless cloning with pre-multimerized inserts, are presented below in detail. The failure of these experiments to produce the desired gene provided us new insight into the requirements of the directed DNA polymerization process, which we have applied to the development of a new method that works reliably toward this end. Our modified, controlled cloning strategy is described and illustrated with the successful expression of a particular protein polymer sequence fitting the requirements described above.

Gene Construction and Multimerization. A 104-base synthetic oligonucleotide (ssDNA), which encodes three tandem repeats of GKGSAQA, was designed using the most favorable codons in typical *E. coli* strain codon usage[24] and is shown in Scheme 1. This synthetic template was PCR-amplified with two primers to create a large amount of dsDNA (gene **1**). The coding sequence

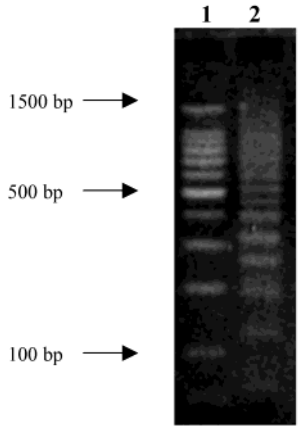


Figure 1. Self-ligation products obtained by concatemerization of gene **1** monomer: lane 1, 100 bp DNA ladder; lane 2, self-ligation products obtained with gene **1** monomer (63 bp).

of **1** was flanked by two *Eam*1104 I recognition sites, the cleavage sites of which are indicated by arrows in Scheme 1.

After PCR amplification, gene **1** (104 bp) was digested with *Eam*1104 I restriction enzyme to generate a 63-bp DNA "monomer". The 63-bp DNA monomer for gene **1** was isolated from the reaction mixture by agarose gel electrophoresis onto a (diethylamino)ethyl (DEAE)− cellulose membrane. The isolated monomers were then multimerized through enzymatic ligation with T4 DNA ligase. Analysis of the ligation products via 1.5% agarose gel electrophoresis showed that self-ligation of the DNA monomer afforded a distribution of multimers with degree of concatemerization extending to more than 20 (see Figure 1, lane 2). That is to say, species existed in the mixture that contained up to 20 ligated repeats of gene **1**. The self-ligated DNA mixture was then purified from the PCR reaction constituents with standard phenol/chloroform extraction and ethanol precipitation.

Standard Seamless Cloning Method. To clone the concatemer gene for subsequent protein expression, a modified recipient expression plasmid, pGGT, was generated using PCR in the presence of 5-methyldeoxycytosine. The procedure which created recipient plasmid pGGT was performed on the basis of the seamless cloning technique developed by McMillan et al.[17] (see Figure 2A).

Purified multimeric fragments and pGGT were ligated by T4 DNA ligase treatment, and *E. coli* strain Top10 was then transformed with the resulting ligated plasmids. As a result of plasmid isolation and enzyme digestion (*Nde*I/*Bam*HI), transformants containing multimeric DNAs ranging from 1 to 6 repeats of gene **1** were identified from among several hundred colonies. The longest concatemer that was cloned into a protein expression plasmid under varying conditions contained only 6 repeats, even though DNA concatemer fragments of up to 20 repeats were known to have been present in substantial concentration in the products of the ligation reaction.

It seemed that, in order to clone long DNA concatemers into pGGT successfully, such long concatemers might need to be fractionated from the shorter ones first. We reasoned that the probability of incorporating short DNA concatemers may be much higher than the insertion probability of long ones. To fractionate the long concatemers, DNA fragments containing fewer than 10 repeats were removed from the self-ligation mixture by

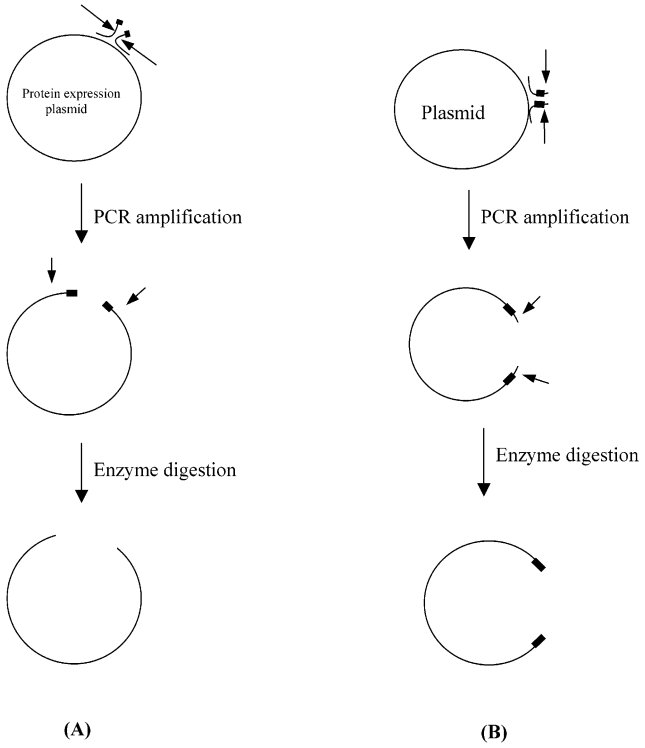


Figure 2. Schematic representation of recipient expression plasmid (A) and a plasmid for liberating concatemers freely (B). Rectangular boxes and arrows represent recognition sites and cleavage sites of the restriction enzyme, respectively.

electroelution from an agarose gel. The fractionated long concatemers were then ligated with pGGT and transformed following the same protocol described above. After repeated attempts, *no* plasmids containing concatemerized gene **1** inserts were found to have been incorporated into *E. coli*. The reason for this was determined in a subsequent experiment, described below.

Seamless Cloning with Pre-multimerized Inserts. It appeared that, with the cloning method described above, the longest concatemer that can be generated contains ~20 repeats, but the relative concentration of the long concatemers is low. Under these circumstances, the probability of successfully cloning long concatemers would also be extremely low. We hypothesized that multimerization with "pre-multimerized" gene **1** constructs would be a prospective method to produce a high concentration of long concatemers (comprising >10 repeats). If a short concatemer (i.e., 6 repeats) could be used for self-ligation as a long monomer, a large amount of long concatemers would be produced, under the assumption that the self-ligation pattern of a long DNA monomer (378 bp) would be the same as that of a short one (63 bp). To accomplish this, a plasmid that allows liberation of the concatemerized gene **1** insert for a subsequent, second multimerization step was needed. In the case of the seamless cloning method mentioned above, the PCR-amplified pET-19b plasmid contains *Eam*1104 I recognition sites that are derived from the primers. These recognition sites are then used to create cohesive ends for the ligation with the concatemerized inserts, thus preventing the use of *Eam*1104 I to release the concatemerized gene **1** inserts from the protein expression plasmid for subsequent multimerization (Figure 2A). To get around this limitation, a shuttle vector was created that allows for the

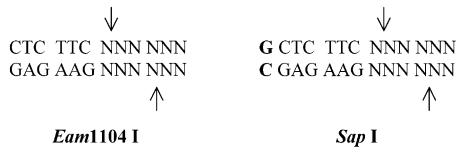


Figure 3. Recognition and cleavage sequences of restriction endonuclease *Eam*1104 I and *Sap* I.

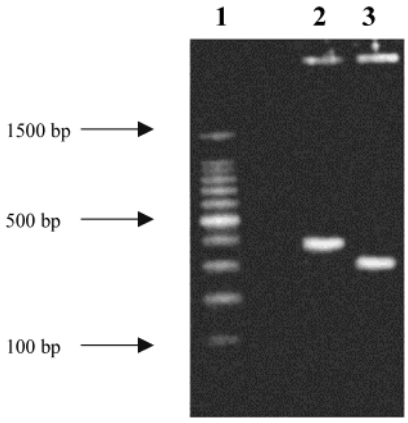


Figure 4. Self-ligation product of pre-multimerized monomer: lane 1, 100 bp DNA ladder; lane 2, pre-multimerized monomer before self-ligation (378 bp, 6 repeats).

liberation of concatemerized gene **1** inserts that are suitable for a second multimerization step, using only "monomers" containing 6 repeats of gene **1**.

To create the shuttle vector, *Sap* I restriction digestion was used instead of *Eam*1104 I digestion, as the two enzymes have similar recognition sites, but *Sap* I has one more recognition base (Figure 3). Because of the extra base in *Sap* I's recognition site, only one cleavage site exists in pUC18, whereas *Eam*1104 I has three cleavage sites in that vector.[25] Utilizing synthetic "dangling" primers (Scheme 2A), which anneal to the 3′ and 5′ termini of the linear plasmid, the shuttle vector was amplified by PCR. Because, by design, both primers encode the *Sap* I recognition site, which lies downstream of the cleavage sites, both recognition sites remain at the ends of the linear vector after *Sap* I digestion (Figure 2B and Scheme 2B). Ligation of this pUC18 recipient vector with the DNA concatemer mixture and transformation of this ligation mixture were performed following the same protocol described above. Again, when several hundred colonies were assayed, transformed bacteria containing plasmids with DNA inserts ranging only from 1 to 6 repeats were obtained. The longest concatemer gene (i.e., 6 repeats, 378 bp) was liberated from the recombinant plasmid by *Sap* I digestion. The isolated 6-repeat DNA "monomer" was then self-ligated again as a long monomer with T4 DNA ligase. The size distribution of the concatemers was analyzed by 1.5% agarose gel electrophoresis. However, the self-ligation pattern of the long monomer (378 bp, 6 repeats) was quite different from that of short one (63 bp, monomer) (Figure 4, lanes 2 and 3). Results from the self-ligation of pre-multimerized gene **1** showed a single predominant band of a higher mobility than the pre-multimerized gene **1** before self-ligation; no "ladder" of DNA concatemers was observed. This implies that the linearized long monomer was cyclized during the self-ligation reaction instead of multimerizing. We realized that this strong bias for cyclization over multimerization must be due to the longer length of the pre-multimerized "monomer", which allows for facile cyclization.[26] This observation also explains why the longest concatemer was only 6 repeats with the previous method, although a wide range of concatemer sizes (from monomer up to ~20 repeats) was initially seen on the gel; most concatemers longer than 6 repeats would rapidly cyclize during the ligation reaction, leaving only a small amount of linear concatemers longer than 6 repeats.

Controlled Cloning Method. To prevent this intramolecular cyclization, which was apparently disallowing the expression of long protein polymers, a novel cloning method (which we call the "controlled cloning" method) was developed that uses both *Sap* I and *Eam*1104 I endonucleases. Because *Sap* I has one more recognition base than *Eam*1104 I, all *Sap* I restriction sites can be digested with *Eam*1104 I endonuclease, but not vice versa (Figure 3). The controlled cloning method requires two different primers to create a suitable PCR product from a concatemer gene. Forward and reverse primers are designed to anneal to both termini of a linearized pUC18 recipient vector, which has *Sap* I recognition sites at both termini. While the forward primer is completely complementary to the pUC18 recipient vector [5′-TTAATGAATCGGCCAACGCGC-3′], the reverse primer has a single mismatched base, to change the *Sap* I recognition site into an *Eam*1104 I recognition site [5′-TGAGCGAGGA**A**CTCTTCAGGT-3′; bold letter represents mismatched base]. Thus, applying PCR to this recombinant plasmid with these two primers succeeds in amplifying a concatemer gene which has one *Eam*1104 I recognition site and one *Sap* I recognition site, each located at one terminus. This PCR product was split into two different tubes. One was digested with *Eam*1104 I endonuclease, and the other was digested with *Sap* I endonuclease (see Figure 5). Since the PCR product has an *Eam*1104 I recognition site at one terminus and a *Sap* I recognition site at the other, incubation with *Eam*1104 I yields cohesive ends on both termini, whereas incubation with *Sap* I produces one blunt-ended terminus and one cohesive-ended terminus. The *Eam*1104 I-digested product was dephosphorylated with calf intestinal alkaline phosphatase (CIP) to minimize intramolecular cyclization. The two differently digested products were then mixed together and ligated with T4 DNA ligase. *Eam*1104 I-digested products will not cyclize because they are dephosphorylated, and *Sap* I-digested products will not also cyclize because one end is blunt and the other end is cohesive. Thus, ligation of these two products results in a controlled and predictable 2× increase in the concatemer length inserted into the pUC18 recipient plasmid.

As is shown in Figure 6, a 12-repeat DNA concatemer was successfully produced from a 6-repeat DNA concatemer by this method. A conjugated product (i.e., a 12-repeat concatemer, lane 4) was generated by the ligation reaction between *Eam*1104I-digested 6-repeat concatemer (lane 2) and *Sap* I-digested 6-repeat concatemer (lane 3). The 12-repeat concatemer (756 bp) was then isolated from the reaction mixture by gel electrophoresis with a DEAE−cellulose membrane and digested again with *Eam*1104 I to make cohesive ends at both termini. This was followed by incubation with T4 nucleotide kinase for phosphorylation and ligation with the pUC18 recipient vector. The entire procedure is diagrammed in Figure 5.

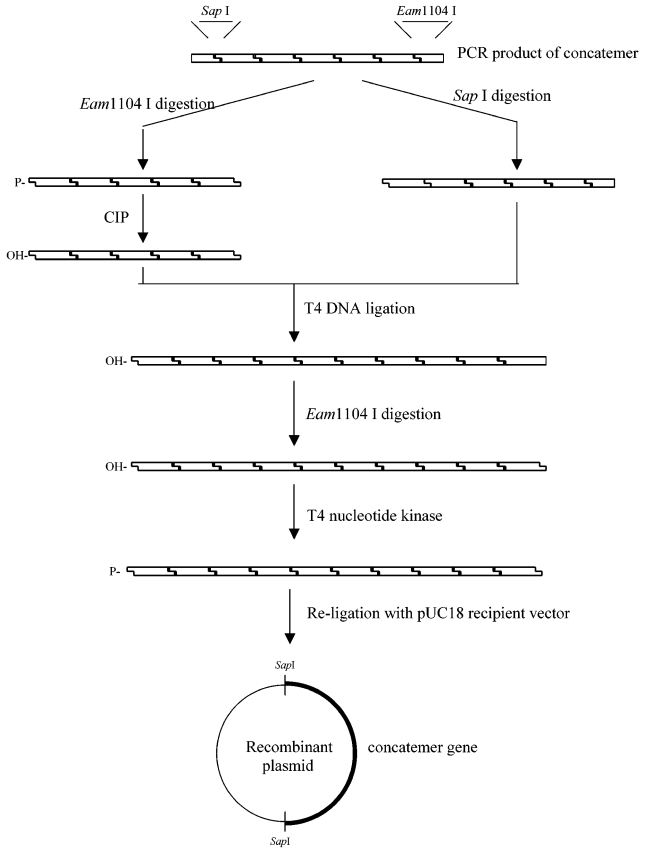


Figure 5. Schematic diagram of the controlled cloning method.

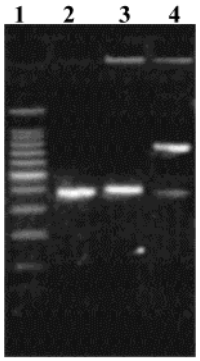


Figure 6. Ligation of two differently digested products: lane 1, 100 bp DNA ladder; lane 2, *Eam*1104I-digested 6-repeat concatemer; lane 3, *Sap* I-digested 6-repeat concatemer; lane 4, a conjugated product from lane 2 and lane 3 (a 12-repeat concatemer).

Although the concatemer insert is still able to cyclize during the ligation reaction with the pUC18 recipient vector, this cyclization is not a significant problem for the following two reasons. First, a high concentration of the desired, pre-multimerized gene insert can be produced by the controlled cloning method, as opposed to the mixture of differently sized gene inserts produced in previously described methods. Second, the gene insert produced is not a cyclized concatemer, but a linearized one. If one uses previously reported methods,[9−14,17] most of the long concatemers are cyclized during the self-ligation reaction, and hence only a small concentration of linear concatemers remains for ligation with a recipient plasmid. However, if we clone a concatemer gene using our method, a long, linearized concatemer gene is reproducibly obtained immediately before the ligation

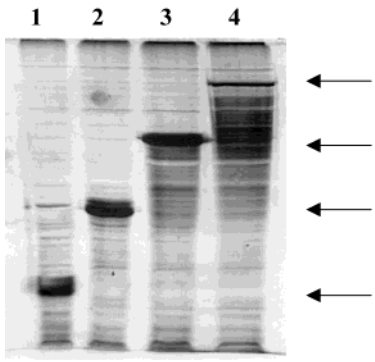


Figure 7. Expression of a "ladder" of proteins with four different lengths (12% polyacrylamide gel electrophoresis, visualized by silver staining): lane 1, 6 repeats (151 amino acids); lane 2, 12 repeats (277 amino acids); lane 3, 24 repeats (529 amino acids); lane 4, 48 repeats (1033 amino acids).

reaction between the concatemer gene and the pUC18 recipient vector. Thus, although cyclization can still occur, its functional effect in gene insert cloning is mitigated due to the high concentration of the desired insert in the ligation reaction.

Using the controlled cloning method, a 24-repeat concatemer (1512 bp) was produced from one with 12 repeats, and a 48-repeat (3024 bp) gene was produced from one with 24 repeats. The forward and reverse sequences of each recombinant plasmid were verified through 3 repeats in both sense and antisense directions using automated DNA sequence analysis by capillary array electrophoresis (data not shown).

Protein Expression and Purification. To express the repetitive proteins, the concatemer genes that were cloned into the pUC18 plasmid (i.e. 6 repeats, 12 repeats, 24 repeats, and 48 repeats) were liberated with *Sap* I digestion and ligated with the pET-19b recipient plasmid (pGGT). *E. coli* strain Top10 was transformed with this mixture and cultured on LB solid media under carbenicillin selection. Plasmids isolated from these transformants were screened by *Nde*I/*Bam*HI double digestion, and the recombinant plasmids containing concatemer inserts were selected. Finally, these recombinant plasmids were introduced into *E. coli* expression strain BLR(DE3), which encodes T7 RNA polymerase under *lac* control on a dormant λ prophage localized on the bacterial chromosome and which allows for IPTG induction of protein expression. A small-scale culture (working volume: 5 mL) of the protein expression strain was performed in LB medium under carbenicillin (50 μg/mL) and tetracycline (30 μg/mL) selection. The cells were incubated at 37 °C until midlog growth was achieved, and an expression of the target polypeptide was induced by addition of IPTG to a final concentration of 1 mM in the media. The proteins of interest were produced from the pET-19b expression system, which encodes the target proteins as a translational fusion to a 24-mer polypeptide that includes an N-terminal decahistidine sequence. The 10 consecutive histidine residues of the fusion proteins allow their purification by immobilized metal affinity chromatography. SDS-PAGE analysis of the whole-cell lysates from small-scale culture (Figure 7) indicated that the four differently sized proteins accumulated well under inducible expression.

Two of the fusion polypeptides (12 and 24 repeats) were subsequently expressed under large-scale conditions (working volume: 1 L), and the cellular extracts

were suspended with lysis buffer. Isolation of fusion proteins was accomplished under denaturing conditions by immobilized metal affinity chromatography. The purified proteins were dialyzed for 3 days to remove salts and then lyophilized. The final fusion polypeptides afforded unoptimized yields of ~15 mg/L. The purified polypeptides were analyzed by HPLC to ensure that the proteins were pure (data not shown). The final products were identified with MALDI−TOF mass spectrum analysis. The observed masses of 24.59 and 46.25 kDa (12 and 24 repeats, respectively) correlated well with the theoretical values of 24.56 and 46.15 kDa. The identity of the protein polymer with 12 repeats was confirmed by amino acid compositional analysis, where the measured amino acid composition was commensurate with the theoretical values.

The secondary structure of the largest of the protein polymers (with 24 repeats) in aqueous solution was studied by far-ultraviolet circular dichroism (CD) spectroscopy. The CD spectrum, exhibiting a minimum signal at 196 nm and a maximum signal at 215 nm (data not shown), indicates that the protein adopts a random coil conformation in solution. Hence, as we had hoped, we were able to avoid the formation of β-sheet structure, which can negatively affect protein solubility. Additionally, the temperature-dependent phase behavior of the same polypeptide in aqueous solution was studied by temperature-dependent turbimetry. The polymer solution (1 mg/mL) does not coacervate over a wide range of temperature (20−95 °C), which indicates that this peptide is strictly water-soluble and does not show LCST-induced precipitation from solution. This protein polymer, which presents evenly spaced lysine groups for chemical or enzymatic modification, should serve as a generally useful scaffold for biomaterial and tissue engineering applications.

Conclusions

We have presented a novel cloning method that is useful for the production of long repetitive polypeptides. This approach not only enables precise control of the size of a concatemerized gene by providing higher-order multimers through the reconcatemerization of premultimerized genes but also eliminates a specific sequence requirement of the DNA monomer for restriction endonuclease cleavage. Additionally, this approach minimizes the effect of intramolecular cyclization of the concatemerized genes which can happen during the self-ligation reaction. Four differently sized DNA concatemers (i.e., 378 bp, 756 bp, 1512 bp, and 3024 bp) were successfully cloned using this technique, and the corresponding repetitive polypeptides were expressed in an *E. coli* expression strain. The size of the polypeptides was confirmed by SDS-PAGE and MALDI−TOF mass spectrum analysis. The consecutive histidine residues of the fusion proteins allow purification of the proteins by immobilized metal affinity chromatography, where the N-terminal leader sequences can be removed from the fusion proteins by cyanogen bromide or other appropriate endoproteinases. This controlled cloning technique should be applicable to the production of any repetitive polypeptide, so long as the expressed protein polymer is not extremely toxic to the host cells.

Acknowledgment. We thank professor Vincent Conticello (Emory University), Dr. Kent Kirshenbaum (California Institute of Technology), and Dr. Wyatt Vreeland (Northwestern University) for helpful discussions. We acknowledge support from the U.S. Department of Energy, Office of Biological and Environmental Research (Grant DE-FG02-99ER62789). We acknowledge the use of the Keck Biophysics Facility at Northwestern University for CD and UV studies.

References and Notes

(1) Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1999**, *32*, 1701−1703.
(2) Welsh, E. R.; Tirrell, D. A. *Biomacromolecules* **2000**, *1*, 23−30.
(3) Fahnestock, S. R.; Irwin, S. L. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23−32.
(4) Winkler, S.; Szela, S.; Avtges, P.; Valluzzi, R.; Kirschner, D. A.; Kaplan, D. *Int. J. Biol. Macromol.* **1999**, *24*, 265−270.
(5) Zhou, Y.; Wu, S.; Conticello, V. P. *Biomacromolecules* **2001**, *2*, 111−125.
(6) Yu, S. M.; Conticello, V.; Kayser, C.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Nature (London)* **1997**, *389*, 167−170.
(7) Yu, S. M.; Soto, C. M.; Tirrell, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 6552−6559.
(8) McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809−4821.
(9) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M. F. F. *Biotechnol. Prog.* **1990**, *6*, 198−202.
(10) McGrath, K. P.; Tirrell, D. A.; Kawai, M.; Mason, T. L.; Fournier, M. J. *Biotechnol. Prog.* **1990**, *6*, 188−192.
(11) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727−733.
(12) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471−5475.
(13) McPherson, D. T.; Xu, J.; Urry, D. W. *Protein Expression Purif.* **1996**, *7*, 51−57.
(14) Fukushima, Y. *Biopolymers* **1998**, *45*, 269−279.
(15) Prince, J. T.; Mcgrath, K. P.; Digirolamo, C. M.; Kaplan, D. L. *Biochemistry* **1995**, *34*, 10879−10885.
(16) Winkler, S.; Wilson, D.; Kaplan, D. L. *Biochemistry* **2000**, *39*, 12739−12746.
(17) McMillan, R. A.; Lee, T. A. T.; Conticello, V. P. *Macromolecules* **1999**, *32*, 3643−3648.
(18) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2002**, *3*, 357−367.
(19) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1992.
(20) Nagy, I. B.; Majer, Z.; Hudecz, F. *Biopolymers* **2001**, *58*, 152−164.
(21) Brinkley, M. *Bioconjugate Chem.* **1992**, *3*, 2−13.
(22) Yu, M.; Nowak, A. P.; Deming, T. J. *J. Am. Chem. Soc.* **1999**, *121*, 12210−12211.
(23) McMillan, R. A.; Conticello, V. P. *Macromolecules* **2000**, *33*, 4809−4821.
(24) Moriyama, E. N.; Powell, J. R. *Nucleic Acids Res.* **1998**, *26*, 3188−3193.
(25) Goeden-Wood, N. L.; Conticello, V. P.; Muller, S. J.; Keasling, J. D. *Biomacromolecules* **2002**, *3*, 874−879.
(26) Podtelezhnikov, A. A.; Mao, C.; Seeman, N. C.; Vologodskii, A. *Biophys. J.* **2000**, *79*, 2692−2704.

MA020892I