Published in final edited form as:

Biotechnol Prog. 2009 November ; 25(6): 1810–1818. doi:10.1002/btpr.208.

Expression of a Recombinant Elastin-Like Protein in *Pichia pastoris*

Rory E. Sallach^{1,2}, Vincent P. Conticello³, and Elliot L. Chaikof^{*,1,2,4}

¹Departments of Biomedical Engineering, Emory University/Georgia Institute of Technology, Atlanta, GA 30332

²Surgery, Emory University, Atlanta, GA 30332

³Department of Chemistry, Emory University, Atlanta, GA 30332

⁴School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30322

Abstract

The translation of highly repetitive gene sequences is often associated with reduced levels of protein expression and may be prone to mutational events. In this report, we describe a modified concatemerization strategy to construct a gene with enhanced sequence diversity that encodes a highly repetitive elastin-like protein polymer for expression in *Pichia pastoris*. Specifically, degenerate oligonucleotides were used to create a monomer library, which after concatemerization yielded a genetically nonrepetitive DNA sequence that encoded identical pentapeptide repeat sequences. By limiting genetic repetition, the risk of genetic deletions, rearrangements, or premature termination errors during protein synthesis is minimized.

Keywords

Recombinant elastin; genetic engineering; monomer library; protein expression; *Pichia pastoris*

INTRODUCTION

Over the past decade a number of reports have described the design of synthetic genes, which encode elastin-like proteins (ELP) for bacterial expression in *Escherichia coli*. Although advantages exist, significant limitations associated with *E. coli* expression systems have been noted. The lack of eukaryotic post-translational systems, the insolubility of over-expressed mammalian proteins by sequestering into inclusion bodies, difficult purification from a pool of cytoplasmic proteins and cellular contaminants, and endotoxin contamination have encouraged the use of other expression systems including yeast, plant, insect, and mammalian cells. Endotoxin has been a specific concern for ELP expression as it becomes associated with the protein product upon cell lysis. Endotoxin, an amphiphilic lipopolysaccharide, is a toxic constituent found in the outer cell wall of gram-negative bacteria and known to induce pyrogenic pathologies¹. A number of approaches have been utilized to reduce endotoxin contamination including sodium hydroxide digestion^{2,3}, centrifugal ultrafiltration, phase separations with detergents^{4–7} or solvents⁸, neutralizing agents⁹, and endotoxin selective absorber matrices¹⁰ and membranes¹¹. Removal of endotoxin from ELPs has often required

Address correspondence to: Elliot L. Chaikof, M.D., Ph.D., 101 Woodruff Circle, Room 5105, Emory University, Atlanta, GA 30322, Phone (404) 727-8413, Fax: (404) 727-3660, E-mail: echaiko@emory.edu.

one of these secondary purification treatments in addition to traditional purification through temperature induced precipitation, thereby, reducing overall protein yields.

Recently, yeast and plant expression systems have been explored for the expression of ELPs and related matrix proteins. For example, recombinant silk-elastin proteins have been successfully expressed in tobacco and potato plants ¹². Additionally, the expression of a variety of target proteins in transgenic tobacco has been enhanced by an order of magnitude when fused to elastin-like polypeptides ^{13–15}. Nonetheless, while transgenic plants offer the potential for scalability, reduced costs ¹⁶, and inherent biosafety through a reduced risk of viral or prion contamination ¹⁴, high yields have largely been limited to selected antibodies, enzymes, and vaccines ¹⁷. The majority of recombinant proteins accumulate in only small amounts.

As an alternate approach, yeast expression systems have become an increasingly attractive host for the expression of heterologous proteins ^{18, 19}, due to their capacity to be incorporated into industrial-scale fermentation schemes characterized by high cell densities in relatively inexpensive media. In addition, heterologous proteins have been efficiently secreted into the expression medium, resulting in low-cost recovery of the protein. Significantly, endotoxin is not produced by yeast, thereby simplifying purification and sterilization strategies. However, overall protein expression is influenced by two variables, culture cell density and the amount of recombinant protein per cell ²⁰ and as a consequence of the more complex process of protein production in a eukaryotic organism, yields are often low as compared to *E. coli* expression systems. Nonetheless, tropoelastin, collagen, and silk-like proteins have all been expressed in yeast with varying degrees of success ^{21–24}.

In this report, a novel strategy was devised to construct a gene with enhanced sequence diversity that encodes a highly repetitive elastin-like protein polymer for expression in *P. pastoris*. Traditionally, large repetitive genes that comprise most protein polymers have been created using a concatemerization strategy where a pentapeptide repeat cassette (monomer repeat unit) is self-ligated in a head-to-tail fashion ^{25, 26}. While this strategy has proven suitable for expression of elastin-like proteins in *E. coli*, the translation of repetitive gene sequence, especially in other host systems, is often associated with reduced levels of protein expression. Indeed, this effort was motivated, in part, by an initial failure to observe protein expression for a designed elastin-mimetic protein in which all DNA monomers, though genetically diverse within the repeat sequence, were otherwise identical. Moreover, repetitive sequences are often prone to mutational events. Given these potential limitations, we designed a modified concatemerization strategy in which seven dissimilar monomer repeat units, encoding identical pentapeptide repeat sequences, served as a monomer library for the concatemerization reaction. In this manner, a protein polymer gene was generated through random incorporation of distinct monomer repeat units. DNA monomers encoding identical amino acid sequences were synthesized in a manner that accounted for the preferred codon usage of *P. pastoris*, but in which the third nucleotide for proline, glycine and valine codons was degenerate. Thus, concatemerization of the monomer library produced a genetically nonrepetitive DNA sequence for the pentapeptide repeat [(VPGVG)₂VPGE(VPGVG)₂]. By limiting genetic repetition, the risk of genetic deletions, rearrangements, or premature termination errors during protein synthesis was minimized ^{19, 27, 28,29}.

MATERIALS and METHODS

Synthetic gene construction of Yeast ELP monomer library

A collection of distinct single-stranded oligonucleotides corresponding to a monomer repeat unit was chemically synthesized (Sigma Genosys, Inc). This was accomplished through use of degenerate bases incorporated into the design of the **Yeast ELP** coding sequence. Specifically, W encoded for A or T in proline and glycine and H for A, T, or C in valine. When chemically

synthesized, equal molar amounts of the assigned nucleotides were used when synthesizing the degenerate position affording random incorporation of bases at those designated positions (Table 1B). The lyophilized sequences were resuspended in elution buffer (10mM tris-HCl, pH 8.5) to a final concentration of 0.5 µg/µL. DNA Polymerase I Klenow fragment (New England Biolabs) was utilized in a primed extension of the oligonucleotide template for the second strand synthesis yielding the double stranded cassette of the monomer repeat unit. An aliquot of the reaction mixture was analyzed via gel electrophoresis (4% GTG NuSieve, 1 X TBE buffer) to verify a single band corresponding to the size of the monomer repeat unit (~75 bp). Subsequently, a preparative gel was used to isolate DNA and the corresponding band was purified via Amicon Ultrafree Centrifugal Filter Units (Millipore) and isolated via ethanol precipitation. A total of 20 µg of the DNA cassette was digested with *Bam*H I (10 U/µg) and *Hind* III (10 U/µg) restriction enzymes, extracted with phenol/chloroform, and isolated via ethanol precipitation.

The pZErO-1 (Invitrogen) acceptor plasmid (1 µg) was prepared via *Bam*H I and *Hind* III double digestion, followed by heat inactivation of the enzymes at 65 °C and a dilution of the digested plasmid to 10 ng/µL. **Yeast ELP** monomers were designed with *Bam*H I and *Hind* III overhangs to enable cloning into pZErO-1 at these restriction sites.

The DNA cassettes and respective acceptor plasmids were ligated together in the presence of T4 DNA ligase (New England Biolabs) at 16 °C for 30 minutes. A 2 µL aliquot of the ligation reaction mixture was used to transform 40 µL of electrocompetent TOP10F' *E. coli* cells (Invitrogen). A total of 100 µL of the transformation mixture was spread onto low salt Luria Broth (LSLB) agar (5 g tryptone, 2.5 g yeast extract, 2.5 g NaCl, 7.5 g agar, 200 mL ddH₂O, pH 7.5) supplemented with Zeocin (50 µg/µL). The plates were incubated for 12 hours at 37° C. Twenty transformants were selected from each plate to inoculate individual 7 mL cultures of LSLB/Zeoicin media. Cultures were rotary incubated for 12 hours at 37°C. Plasmid DNA was isolated following a Qiagen Spin Miniprep protocol (Qiagen, Inc.). Miniprep DNA was initially screened by a *Bam*H I and *Hind* III double digestion. Positive transformants were verified by agarose gel electrophoresis (4% GTG NuSieve agarose, 1X TBE buffer). Automated DNA sequencing utilizing the M13 forward and M13 reverse primers confirmed correct and unique DNA products for seven separate monomer repeat unit cassettes, yielding the **Yeast ELP** monomer library (Table 2).

Concatamerization of the Yeast ELP monomer library

Concatamerization reactions utilized a total of 3.2 µg (0.4 µg of each monomer) of the *Bbs* I / *Bsm*B I digested DNA. Monomers were then ligated randomly end-to-end via T4 DNA ligase. The multimerization mixture was separated by size using agarose gel electrophoresis (1% agarose, 1X TBE buffer) (Figure 1A). Concatemers were excised in blocks, <500 bp, 500–1000 bp, 1000–3000 bp and purified using Zymoclean Gel DNA Recovery protocol (Zymo Research, Inc). Concatamers of 1000–3000 bp in size were ligated into the acceptor plasmid at the *Bbs* I site at 16 °C for 16 hours. The acceptor plasmid was prepared from the pZErO-1 plasmid containing a single monomer repeat unit, digested with *Bbs* I, and dephosphorylated with SAP (Shrimp Alkaline Phosphatase, Roache) to prevent self ligation. Ligation mixtures were used to transform competent TOP10F' cells and 100 µL of the transformation mixture was plated on LSLB/Zeoicin agar plates. DNA from positive clones were isolated via MacConnell automated miniprep and screened through double digestion using *Bam*H I and *Hind* III restriction enzymes. Clones of predetermined size (approximately 1.5 kB) were isolated and sequences confirmed with automated DNA sequencing.

Construction of the modified pPICZ α -A expression plasmid

Single stranded oligonucleotides encoding the sense and anti-sense strands of the **Yeast ELP** adaptor were chemically synthesized (Sigma Genosys, Inc.) (Table 3). The **Yeast ELP** adaptor is a ~50 bp DNA cassette designed to contain restriction enzyme cut sites midway through the cassette to allow for insertion of the **Yeast ELP** concatamer and allow for facile cloning into the pET24-a expression vector within the multiple cloning region. This ensures correct insertion of the gene in frame with the N-terminal α -factor secretion signal sequence and the C-terminal polyhistidine tag and *c-myc* epitope. The DNA was suspended in 10mM Tris buffer (pH 8) to a final concentration of 0.5 $\mu\text{g}/\mu\text{L}$. A solution of 10 μg of each corresponding oligonucleotide, 4 μL 5M NaCl, 4 μL 1M MgCl_2 , 152 μL of sterile ddH₂O was subjected to an annealing procedure initiated at a reaction temperature of 99 °C with incremental temperature decrements of 1 °C every 5 minutes to a final reaction temperature of 30 °C. The resultant double stranded DNA cassette was analyzed by agarose gel electrophoresis (4% GTG NuSieve agarose, 1X TBE buffer).

Double stranded synthetic DNA was phosphorylated through a 2 hour incubation with T4 polynucleotide kinase (New England Biolabs) in the presence of T4 DNA ligase buffer with 10mM ATPs (New England Biolabs). The enzymes were removed with phenol/chloroform/isoamyl alcohol (25:24:1) and the dsDNA was recovered through an ethanol precipitation.

The pPICZ α -A plasmid (1 μg , Invitrogen) was doubly digested with *Xho* I and *Xba* I, followed by heat inactivation of the enzymes at 65 °C, and dilution to 10 ng/ μL . The adaptor was designed with *Xho* I and *Xba* I overhangs to enable its cloning into pPICZ α -A. Adaptor and plasmid were ligated together in the presence of T4 DNA ligase at 16 °C for 30 minutes. A 2 μL aliquot of the ligation reaction mixture was used to transform 40 μL of electrocompetent TOP10F' *E. coli* cells. A total of 100 μL of the transformation mixture was spread onto LSLB agar supplemented with Zeocin (50 $\mu\text{g}/\mu\text{L}$). The plates were incubated for 12 hours at 37 °C. Five transformants were selected from each plate to inoculate individual 7 mL cultures of LSLB/Zeocin media. Cultures were rotary incubated for 12 hours at 37 °C. Plasmid DNA was isolated following a Qiagen Spin Miniprep protocol. DNA was initially screened by a *Xho* I and *Xba* I double digestion. Positive transformants were verified by agarose gel electrophoresis (4% GTG NuSieve agarose, 1X TBE buffer). Automated DNA sequencing utilizing the AOX1 5' and AOX1 3' primers confirmed the correct DNA product.

Cloning the Yeast ELP gene into the modified pPICZ α -A expression plasmid

The modified pPICZ α -A plasmid was digested with restriction enzyme *BsmB* I and SAP dephosphorylated. The **Yeast ELP** gene was excised from the pZER0-1 plasmid via sequential digestion using restriction enzymes *Bbs* I and *BsmB* I and purified from an agarose gel. A ligation reaction was performed to relocate the **Yeast ELP** gene from pZER0-1 to pPICZ α -A. The ligation mixture was transformed into electrocompetent TOP10F' cells and plated on LSLB media under Zeocin antibiotic selection. Isolated DNA from transformants was screened via agarose gel electrophoresis using *Xho* I and *Xba* I double digestion. A product of 1575bp was observed (Figure 2B). Automated DNA sequence analysis using AOX1 5' and AOX1 3' primers confirmed correct insertion of the **Yeast ELP** gene in frame with the N-terminal α -factor secretion signal and the C-terminal peptide containing the *c-myc* epitope and the polyhistidine tag.

Preparation of *P. pastoris* for chemical transformation

The preparation of chemically competent *X-33 P. pastoris* cells was performed following the EasyComp Transformation (Invitrogen) protocol. Briefly, a YPD (Yeast Extract Peptone Dextrose) agar plate was streaked with the *X-33* strain of *P. pastoris* such that isolated individual colonies grew after an incubation at 30 °C for 2 days. A total of 10 mL of YPD

media was inoculated with a single colony and grown overnight at 30 °C in a shaking incubator (250 rpm). The cells were diluted from the overnight culture to an OD₆₀₀ of 0.1 in 10 mL of YPD. Cells were pelleted by centrifugation at 500g for 5 minutes at 25 °C and resuspended in 1 mL Solution 1 (EasyComp), yielding competent X-33 cells. A total of 50 µL of competent cells were aliquoted into 1.5 mL sterile screw-cap tubes and stored at –80 °C.

Transformation of the Yeast ELP gene into *P. pastoris*

E. coli was utilized to propagate the plasmid containing the **Yeast ELP** gene. A total of 5 to 10 µg of plasmid carrying the **Yeast ELP** gene was isolated and was linearized within the 5' AOX1 region through digestion with restriction enzyme *Pme* I to promote integration into the *P. pastoris* host. A vector linearized within the 5' AOX1 region will integrate by gene insertion into the host's 5' AOX1 region. The linearized plasmid was isolated via preparative gel electrophoresis (1% agarose, 0.5X TAE) and purified using Zymoclean Gel Recovery.

Transformation was performed following the EasyComp transformation protocol. One tube of competent X-33 cells was thawed at room temperature. A total of 3 µg of linearized **Yeast ELP** DNA was added to the thawed cells. A total of 1 mL of Solution II (EasyComp) was added to the DNA-cell mixture, vortexed, and incubated for 1 hour at 30 °C in a water bath. The mixture was mixed every 15 minutes to increase transformation efficiency. Following the incubation, cells were heat shocked at 42 °C for 10 minutes. Cells were split into two centrifuge tubes with 1mL of YPD medium added to each tube and incubated at 30 °C for 1 hour to allow for expression of Zeocin resistance. Cells were pelleted by centrifugation at 3000g for 5 minutes at 25 °C and the pellet resuspended in 150 µL of Solution III (EasyComp). The entire transformation was spread on YPDS (Yeast Extract Peptone Dextrose with Sorbitol)/Zeocin agar plates and incubated at 30 °C for 4 days.

PCR screening for correct gene insertion into the *P. pastoris* host

All polymerase chain reaction (PCR) screening was performed using the Qiagen Taq PCR, which includes the Q-solution to assist in the amplification of difficult to amplify G-C rich DNA. According to the manufacturer's instruction, a 50 µL reaction mixture using the AOX1 5' and AOX1 3' primers (5 µM) and Q-solution was subjected to a PCR cycle that employed 25 cycles of a one minute fifteen second denaturation at 94 °C, one minute fifteen second primer anneal at 57 °C, and a three minute primer extension at 72 °C.

Expression of Yeast ELP in *P. pastoris*

A single colony of the transformants was used to inoculate 25 mL of BMGY, a buffered glycerol complex medium, in a 250 mL baffled flask, grown at 30 °C with shaking until and OD₆₀₀ of 2 was reached. Cells were harvested by centrifugation at 3000g for 5 minutes at 25 °C and resuspended in 100 mL BMMY, a buffered methanol complex medium, for induction of expression in a 1 L flask. To maintain induction, 100% methanol was added to a final concentration of 0.5% methanol every 24 hours for 4 days. After 4 days, cells were harvested and supernatant collected. Both the cellular fraction and supernatant was analyzed for protein expression.

Purification of Yeast ELP from *P. pastoris*

Prior to protein purification, the supernatant was concentrated using an Amicon Ultra centrifugal filter with a molecular weight cut off of 30,000 kDa. Metal affinity chromatography was used for purification of **Yeast ELP**, which isolated the protein by the polyhistidine tag, according to manufacturer instructions. Briefly, a cobalt based TALON metal affinity resin (Clontech) was equilibrated with equilibration buffer (50mM sodium phosphate, 300mM NaCl, pH 7.0). The concentrated supernatant was run through the column by gravity flow

followed by extensive washes with equilibration buffer. The bound protein was eluted with Elution buffer (50 mM sodium phosphate, 300mM NaCl, 250mM imidazole, pH 7.0) and desalted using PD-10 desalting columns (GE Healthcare Life Sciences). Lyophilization afforded protein **Yeast ELP** as fibrous solid in an isolated yield of 2.5 mg/L. Efforts to optimize yields are underway ³⁰.

Additionally, as observed with other structural protein expression, some protein was isolated from both the cytosolic and membrane fractions, though most of the protein was found to be secreted into the media. It has been noted that some higher eukaryotic proteins are not compatible with the yeast secretory apparatus and these proteins remain trapped at some point along the secretory pathway ^{30, 31}. To isolate this protein, the harvested cells were resuspended in breaking buffer (20mM Tris-Cl, pH 7.5, 1mM EDTA, 5% glycerol) and then disrupted through vortexing cycles. A low speed spin to pellet out whole, unbroken cells was used to assess breaking efficiency. Cell lysates were centrifuged at 40,000g for 40 minutes at 4 °C and the cytosolic fraction was collected and His-purified. The membrane fraction was detergent extracted with 0.5% Triton X-100 overnight at 4 °C. The lysate was centrifuged at 40,000g for 40 minutes at 4 °C and the soluble membrane fraction collected and His-purified. Isolated yields of less than 1 µg were obtained from both cytosolic and membrane fractions, which confirmed that the ELP was not trapped in the secretory pathway.

Identification of the recombinant elastin protein, Yeast ELP

Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis was performed. Samples were separated on a 4–20% gradient acrylamide-SDS gel and total protein was visualized with Coomassie G250 (BioRad). Western blot analysis was performed by transfer to an Immobilon PSQ (Polyvinylidene fluoride) (PVDF) membrane (Millipore) and probed with either mouse His-tag monoclonal antibodies (Novagen) or anti-*c-myc* mouse monoclonal antibodies followed by a goat-anti-mouse HRP (horseradish peroxidase) secondary antibody. Bands were visualized by using the ECL Western blotting detection kit (Amersham Biosciences). Both antibodies stained a band at 65 kDa, corresponding to the **Yeast ELP** protein.

Edman degradation was used to identify the repeat unit of **Yeast ELP** to confirm expression. Briefly, the protein was electroblotted on a PVDF membrane and stained with Amido Black (BioRad). The **Yeast ELP** protein negatively stained using this stain. Approximately 0.5 to 2.0 pmol of protein obtained from the membrane was used for N-terminal sequence analysis using automated sequencers at the Microchemical and Proteomics Facility at Emory University.

RESULTS and DISCUSSION

Conceptual framework for the design of Yeast-expressed ELP

The elastin-like protein, **Yeast ELP**, was employed as a test substrate for assessing the efficacy of monomer library concatamerization in the design of a nonrepetitive ELP gene for expression in yeast. The **Yeast ELP** protein comprises a concatenated series of pentapeptide repeat units yielding 21 repeats of the pentapeptide monomer sequence [(VPGVG)₂VPGE(VPGVG)₂]. Our group has investigated this repeat sequence as a constituent of multiblock protein copolymers expressed by *E.coli*. As a consequence of its high transition temperature (>> 37° C), this sequence affords a hydrophilic, conformationally flexible protein segment under physiologically relevant conditions that is highly water soluble ^{26, 32–34}.

In order to express a highly repetitive elastin-like protein in *P. pastoris*, a novel strategy was employed to reduce primary DNA sequence repetition without altering the recurring 21-mer peptide sequence. This was accomplished through the design of a DNA monomer library.

Protein engineering has frequently used protein libraries that include defined amino acid mixtures at certain positions of interest^{35, 36}. Two of the most commonly used methods for library generation are random mutagenesis and cassette mutagenesis. Random mutagenesis introduces random point mutations throughout the entire protein. The most common method for creating such a library is PCR. This method has been utilized in studies, for example, aimed to identify amino acids integral to enzymatic function and protein-protein interactions³⁷. Cassette mutagenesis is a method of library creation where a region of interest is targeted for mutagenesis through the use of degenerate oligonucleotides. We have used degenerate oligonucleotides in the design of our DNA monomer repeat unit to create a library of monomers, which were subsequently ligated together to create our target ELP gene.

The genetic code is degenerate in that the protein biosynthetic machinery utilizes 61 sense codons to encode the 20 amino acids. Due to this degeneracy, different nucleotide sequences code for the same amino acid. These coding differences are restricted to usually one position in the codon triplet and allow for multiple nucleotides to encode the same amino acid, thereby increasing variability of the primary DNA sequence of the protein. For example, the third position of the glycine codons (GGA, GGG, GGC, GGT) is a fourfold degenerate site because all nucleotide substitutions at this site are synonymous, in that the coded amino acid is unchanged. Degenerate bases allow for the incorporation of multiple nucleotides into the specified site within a codon. As detailed in Table 1B, the monomer repeat unit for the **Yeast ELP** gene was designed through chemical synthesis based on the degenerate bases, W, encoding for either A or T, and H, encoding for A, T or C. While the primary genetic sequence for the **Yeast ELP** was based on the midblock sequence of a previously reported triblock protein expressed in *E. coli* (Table 1A)^{26, 32}, the sequence was modified in accordance with the preferred codon usage of *P. pastoris*. The codon triplet encoding proline residues was determined to be critically important in the successful expression of this protein. CCG, a high usage codon in *E. coli* and used extensively for encoding proline in the pentapeptide repeat is not adequately translated in *P. pastoris*. Therefore, in the design of the **Yeast ELP** gene, only CCA and CCT codons were employed to code for proline residues (Table 1).

Design of a library of nonrepetitive DNA monomers

Chemically synthesized, single-stranded oligonucleotides designed with degenerate bases at the specified locations, were obtained from SigmaGenosys. At each degenerative base location, equal molar amounts of each base were introduced. For example, at a W site, equal amounts of A and T were available during synthesis. Through primed extension, double stranded cassettes of the monomer repeat units were generated. The cassettes were digested with restriction enzymes *Bam*H I and *Hind* III to enable cloning into the pZErO-1 cloning plasmid at the complementary sites within the multiple cloning region. Multiple transformants were screened and automated DNA sequencing confirmed correct and unique DNA products for seven separate monomer cassettes, comprising the **Yeast ELP** monomer library (Table 2A).

The literature reports that gene locations enriched in repetitive sequences are 'hot spots' for mutational events such as insertions, deletions, and frame shifts^{38, 39}. DNA sequences with GC and CA/GT repeats have been reported to undergo spontaneous deletions^{40, 41}, while regions rich in AT have been reported as 'hot spots' for frame shifts and deletions⁴². Additionally, long repetitive sequences have been found to be less stable than shorter ones⁴³. Specifically, in *P. pastoris*, the expression of foreign genes with high A and T content can be affected by premature transcription termination^{44, 45}, therefore, care was taken in the genetic design to limit such 'hot spot' motifs.

Concatamerization and identification of the Yeast ELP gene

Since automated DNA synthesis is currently limited to the production of oligonucleotides of lengths corresponding to approximately 100 bases, sequences encoding larger proteins cannot be directly synthesized. DNA cassette concatamerization is a commonly employed method for assembling genes encoding large, repetitive proteins (Scheme 1)^{25, 26}. The genes composing the monomer library were designed with non-palindromic cohesive ends. This was accomplished through the use of type II restriction enzymes, *Bbs* I and *Bsm*B I, which recognize and cleave non-palindromic sequences as detailed in Table 2B, creating the digested cassette in Table 2C. Consequently, random ligation of the seven DNA monomers proceeds in a head-to-tail fashion to generate concatamers of varying lengths by increments of a monomer repeat unit. Concatamers were separated by size, or degree of concatamerization, on an agarose gel (Figure 1A) and concatamers of 1000 to 3000bp in size were isolated and ligated back into the acceptor plasmid. A clone of 1575bp was identified and was denoted as **Yeast ELP** (Figure 1B).

Expression system employed for Yeast ELP

Recombinant elastin proteins have traditionally been expressed through microbial expression in *E. coli*⁴⁶. Nonetheless, recognized drawbacks exist. In particular, purification of the expressed protein products is labor intensive as isolation from a pool of cytoplasmic proteins and cellular contaminants is required. Additionally, endotoxin, an amphiphilic lipopolysaccharide found in the wall of gram-negative bacteria, is released upon cell lysis and can become associated with the target protein¹. As an alternative, the methylotrophic yeast, *P. pastoris* was investigated for elastin expression and secretion. Notably, *P. pastoris* secretes very low levels of native proteins, which simplifies purification protocols^{32, 47}. Moreover, endotoxin is not present as a potential contaminant.

Recent reports have described the expression of structural proteins in yeast. Spider dragline silk-like proteins have been expressed in both *E. coli* and *P. pastoris* at high levels^{22, 48}. Likewise, full length tropoelastin has been expressed in *Saccharomyces cerevisiae* using a fusion peptide for targeted secretion into the endoplasmic reticulum with enhanced biostability²¹. *P. pastoris* has also been employed for co-expression of Type-I, -II or -III collagen with prolyl 4-hydroxylase yielding collagen fibrils that display D-periodic banding^{23, 24}. However, issues related to gene rearrangement have been observed with these highly repetitive proteins. For example, a 101 amino acid spider dragline silk-like protein, derived from the major dragline protein of *Nephila clavipes*⁴⁹, produced multiple sized protein products, each an integral number of repeats of the 101 amino acid sequence encoded by the gene. It was speculated that different sized products were the result of expansion or deletion of the DNA repeat sequence²². Additionally, the importance of genetic sequence diversity is noted among patients with trinucleotide repeat associated disease where sequence divergence between repeated sequences reduce the severity of the disease and likelihood of inheritance^{50, 51}. In fact, studies examining spontaneous deletions demonstrate that alteration of a single base pair within homology regions reduces the deletion incidence by an order of magnitude⁵².

In this report, *P. pastoris* was employed as the host for expression of the recombinant protein **Yeast ELP**. The methylotrophic yeast strain X-33 was selected. Without glucose present, *P. pastoris* uses methanol as a carbon source. The alcohol oxidase promoter (AOX 1) controls the expression of alcohol oxidase, which is integral to the initial step in methanol metabolism^{30, 53}. The expression vector pPICZ α -A, takes advantage of the AOX 1 promoter and uses methanol to induce protein expression of the **Yeast ELP** protein.

The 5' AOX1 gene within the pPICZ α -A vector allows for targeted integration of the expression construct through homologous recombination into the *P. pastoris* genome, creating

a stable host able to generate high levels of protein expression (Figure 2). The pPICZ α -A vector also contains a Zeocin resistance gene that allows for antibiotic screening of transformed cells. Additionally, this vector includes the C-terminal fusion tags, *c-myc* epitope and polyhistidine (6x His) sequences, that facilitate purification and analysis of expressed proteins. Finally, pPICZ α -A contains an α -factor secretion signal, which targets the protein product for secretion into the growth media.

The 1575 bp **Yeast ELP** concatamer was inserted into a specially designed adaptor at the double *BsmB* I sites (Table 3) for cloning into the pPICZ α -A expression vector at *Xho* I and *Xba* I restriction sites within the multiple cloning region. This allowed for correct in frame insertion of the gene with the N-terminal α -factor secretion signal and the C-terminal *c-myc* epitope and polyhistidine tag.

Confirming Integration by PCR

Polymerase Chain Reaction (PCR) was used to analyze *P. pastoris* integrants to determine successful introduction of the expression cassette into the yeast chromosome. The **Yeast ELP** gene was amplified using the AOX1 5' (5'-GACTGGTTCCAATTGACAAGC-3') primer paired with the AOX1 3' (5'-GCAAATGGCATT-CTGACATCC-3') primer. Using this method, two bands were expected in positive transformants, one corresponding to the **Yeast ELP** gene (1575 bp gene + 588 bp from the pPICZ α -A parent plasmid) and the other to the AOX1 gene within the chromosome (approximately 2.2 kb). All PCR screening was performed using the Qiagen Taq PCR, which includes Q-solution to assist in the amplification of G-C rich DNA. The PCR product obtained was suggestive of incomplete amplification as multiple product bands were observed, most likely as a result of both its repetitive nature and its G-C rich DNA. PCR products were run on a preparative agarose gel. Amplification of the positive control, **Yeast ELP** miniprep DNA from *E. coli* propagation, appeared as a smear and ladder product. Empty *X-33* cells were analyzed as a negative control and amplification of only the AOX1 gene from the yeast chromosome at 2.2 kb was evident. A screen of colonies indicated a possible transformant where a smear and ladder product was observed.

Confirmation of **Yeast ELP** gene integration was accomplished through amplification using **Yeast ELP** 3'-2 (5'-CTCCGACTCCTGGAACAC-3') primer paired with the AOX1 5' (5'-GACTGGTTCCAATTGACAAGC-3') primer. The **Yeast ELP** 3'-2 primer was designed to amplify only a 400 bp product between the regions of the AOX1 5' priming site within the pPICZ α -A vector and into the 5' terminus of the **Yeast ELP** gene. This product was present in the positive control and the putative transformant, but not observed in the negative control. DNA sequencing verified the identity of the 400 bp PCR product as the expected region from the AOX1 5' priming site into the 5' segment of the **Yeast ELP** gene; confirming insertion of **Yeast ELP** into the *P. pastoris* chromosome.

Western blot analysis

Yeast ELP was isolated and affinity purified from the growth media with non-optimized yields of 2.5 mg/L. Small amounts of the protein were identified in the membrane and soluble cytosolic fractions, but in low amounts and were not utilized in protein analysis. The expected molecular mass for the **Yeast ELP** protein was approximately 56 kDa. As the expressed protein product did not stain with Coomassie, Western blot analysis was utilized to confirm the identity. An anti-His primary antibody revealed the protein band migrating at approximately 65 kDa and was confirmed as the band corresponding to **Yeast ELP** using the anti-*myc* antibody (Figure 3 A and B). Elastin proteins tend to migrate approximately 20% higher than the theoretical molecular weight in a PAGE gel^{54, 55}. To confirm the product was not migrating at a higher mass due to glycosylation, the purified protein was treated with N-glycosidase F (Prozyme), an enzyme which catalyzes the release of all N-linked oligosaccharides added by

yeast during secretion. A 65 kDa product was observed before and after treatment with the enzyme ^{31, 56} (Figure 3C).

CONCLUSIONS

In this report, we have demonstrated the expression of a 56 kDa elastin-like protein from *P. pastoris*. We have employed a new strategy, monomer library multimerization, in designing non-repetitive ELP genes for highly repetitive protein sequences. This was accomplished through the synthesis of seven unique monomer cassettes utilizing degenerate bases. The monomer cassettes were randomly ligated in a concatamerization reaction, thereby creating the target gene with varying genetic sequences throughout, which, nonetheless, encode identical repetitive amino acid sequences. We anticipate that this strategy will be useful for creating large, repetitive genes for a variety of expression systems; in order to more closely approach the genetic diversity inherent to native DNA sequences. Additionally, using a yeast expression system the potential exists to generate glycosylated ELPs through incorporating appropriate glycosylation sites. Indeed, several approaches including genomics, combinatorial libraries, and synthetic chemistry have been employed for sugar chain remodeling of human proteins in *P. pastoris* and other expression systems ^{57–59}.

Acknowledgments

This work was supported by NIH grants HL60464, HL71336, and HL083867 (E.L.C. and V.P.C.). We thank Dr. Jan Pohl of the Microchemical and Proteomics Facility at Emory University for assistance with protein sequencing.

REFERENCES

1. Wandrey C, Vidal DS. Purification of polymeric biomaterials. *Ann N Y Acad Sci* 2001;944:187–198. [PubMed: 11797669]
2. Lee SY, Choi J, Han K, Song JY. Removal of endotoxin during purification of poly(3-hydroxybutyrate) from gram-negative bacteria. *Appl Environ Microbiol* 1999;65(6):2762–2764. [PubMed: 10347075]
3. Choi J, Lee SY. Efficient and economical recovery of poly(3-hydroxybutyrate) from recombinant *Escherichia coli* by simple digestion with chemicals. *Biotechnol Bioeng* 1999;62(5):546–553. [PubMed: 10099563]
4. Aida Y, Pabst MJ. Removal of endotoxin from protein solutions by phase separation using Triton X-114. *J Immunol Methods* 1990;132(2):191–195. [PubMed: 2170533]
5. Fiske MJ, Fredenburg RA, VanDerMeid KR, McMichael JC, Arumugham R. Method for reducing endotoxin in *Moraxella catarrhalis* UspA2 protein preparations. *J Chromatogr B Biomed Sci Appl* 2001;753(2):269–278. [PubMed: 11334340]
6. Salek-Ardakani S, Stuart AD, Arrand JE, Lyons S, Arrand JR, Mackett M. High level expression and purification of the Epstein-Barr virus encoded cytokine viral interleukin 10: Efficient removal of endotoxin. *Cytokine* 2002;17(1):1–13. [PubMed: 11886166]
7. Fernando GJ, Murray B, Zhou J, Frazer IH. Expression, purification and immunological characterization of the transforming protein E7, from cervical cancer-associated human papillomavirus type 16. *Clin Exp Immunol* 1999;115(3):397–403. [PubMed: 10193409]
8. Kim C, Ryu H. Immunologic reactivity of a lipopolysaccharide-protein complex of type A *Pasteurella multocida* in mice. *J Vet Sci* 2000;1(2):87–95. [PubMed: 14614303]
9. Fletcher MA, McKenna TM, Quance JL, Wainwright NR, Williams TJ. Lipopolysaccharide detoxification by endotoxin neutralizing protein. *J Surg Res* 1993;55(2):147–154. [PubMed: 8412093]
10. Adam O, Vercellone A, Paul F, Monsan PF, Puzo G. A nondegradative route for the removal of endotoxin from exopolysaccharides. *Anal Biochem* 1995;225(2):321–327. [PubMed: 7762798]
11. Gui W, Xue L, Jing L, Ying L, Zhen S. Endotoxin removal in some medicines and human serum albumin solution by affinity membranes. *Chromatography* 2002;20(2):108–114.

12. Scheller J, Henggeler D, Viviani A, Conrad U. Purification of spider silk-elastin from transgenic plants and application for human chondrocyte proliferation. *Transgenic Res* 2004;13(1):51–57. [PubMed: 15070075]
13. Lin M, Rose-John S, Grotzinger J, Conrad U, Scheller J. Functional expression of a biologically active fragment of soluble gp130 as an ELP-fusion protein in transgenic plants: Purification via inverse transition cycling. *Biochem J* 2006;398(3):577–583. [PubMed: 16716147]
14. Daniell H, Streatfield SJ, Wycoff K. Medical molecular farming: Production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci* 2001;6(5):219–226. [PubMed: 11335175]
15. Scheller J, Leps M, Conrad U. Forcing single-chain variable fragment production in tobacco seeds by fusion to elastin-like polypeptides. *Plant Biotechnol J* 2006;4(2):243–249. [PubMed: 17177800]
16. Scheller J, Guhrs KH, Grosse F, Conrad U. Production of spider silk proteins in tobacco and potato. *Nat Biotechnol* 2001;19(6):573–577. [PubMed: 11385464]
17. Larrick JW, Thomas DW. Producing proteins in transgenic plants and animals. *Curr Opin Biotechnol* 2001;12(4):411–418. [PubMed: 11551472]
18. Romanos MA, Scorer CA, Clare JJ. Foreign gene expression in yeast: A review. *Yeast* 1992;8(6):423–488. [PubMed: 1502852]
19. Cregg JM, Vedvick TS, Raschke WC. Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology (N Y)* 1993;11(8):905–910. [PubMed: 7763913]
20. Srinivasan S, Barnard GC, Gerngross TU. A novel high-cell-density protein expression system based on *Ralstonia eutropha*. *Appl Environ Microbiol* 2002;68(12):5925–5932. [PubMed: 12450812]
21. Toonkool P, Weiss AS. Expression of recombinant human tropoelastin in *Saccharomyces cerevisiae* containing synthetic gene with a high codon adaptation index coupled to the SUC2 invertase signal sequence. *Acta Biotechnol* 2001;2:189–193.
22. Fahnestock SR, Bedzyk LA. Production of synthetic spider dragline silk protein in *Pichia pastoris*. *Appl Microbiol Biotechnol* 1997;47(1):33–39. [PubMed: 9035408]
23. Myllyharju J, Nokelainen M, Vuorela A, Kivirikko KI. Expression of recombinant human type I–III collagens in the yeast *Pichia pastoris*. *Biochem Soc Trans* 2000;28(4):353–357. [PubMed: 10961918]
24. Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA, Toman PD. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J Biol Chem* 2001;276(26):24038–24043. [PubMed: 11279215]
25. McMillan RA, Lee TAT, Conticello VP. Rapid assembly of synthetic genes encoding protein polymers. *Macromolecules* 1999;32:3643–3648.
26. Wright ER, Conticello VP. Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. *Adv Drug Deliv Rev* 2002;54:1057–1073. [PubMed: 12384307]
27. Daly R, Hearn MT. Expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production. *J Mol Recognit* 2005;18(2):119–138. [PubMed: 15565717]
28. Cregg JM, Barringer KJ, Hessler AY, Madden KR. *Pichia pastoris* as a host system for transformations. *Mol Cell Biol* 1985;5(12):3376–3385. [PubMed: 3915774]
29. Arad O, Goodman M. Depsipeptide analogues of elastin repeating sequences: Conformational analysis. *Biopolymers* 1990;29(12–13):1652–1668. [PubMed: 2386811]
30. Cregg JM. Gene expression systems: Using nature for the art of expression. Academic Press; 1999.
31. Higgins, DR.; Cregg, JM. Methods in molecular biology: *Pichia* protocols. Vol. 103. Humana Press, Inc; 1998.
32. Nagapudi K, Brinkman WT, Thomas BS, Park JO, Srinivasarao M, Wright E, Conticello VP, Chaikof EL. Viscoelastic and mechanical behavior of recombinant protein elastomers. *Biomaterials* 2005;26(23):4695–4706. [PubMed: 15763249]
33. Nagapudi K, Brinkman WT, Thomas BS, Wright ER, Conticello VP, Chaikof EL. Protein-based thermoplastic elastomers. *Macromolecules* 2005;38:345–354.

34. Wright ER, McMillan RA, Cooper A, Apkarian RP, Conticello VP. Thermoplastic elastomer hydrogels via self-assembly of an elastin-mimetic triblock polypeptide. *Advanced Functional Materials* 2002;12(2):1–6.
35. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* 2002;99(12):7877–7882. [PubMed: 12060735]
36. Nguyen AW, Daugherty PS. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* 2005;23(3):355–360. [PubMed: 15696158]
37. Kim SJ, Park H, Kim JK, Lee JY, Ahn K, Choe M, Choi YJ, Kim J. Random changes of amino acid residues with expected frequency by saturated point mutagenesis. *Mol Cells* 2000;10(2):232–235. [PubMed: 10850667]
38. Bzymek M, Lovett ST. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* 2001;98(15):8319–8325. [PubMed: 11459970]
39. Morag AS, Saveson CJ, Lovett ST. Expansion of DNA repeats in *Escherichia coli*: Effects of recombination and replication functions. *J Mol Biol* 1999;289(1):21–27. [PubMed: 10339402]
40. Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, Bertheas MF, Mandel JL. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 1991;252(5010):1097–1102.
41. Kraemer KH, Seidman MM. Use of supF, an *Escherichia coli* tyrosine suppressor tRNA gene, as a mutagenic target in shuttle-vector plasmids. *Mutat Res* 1989;220(2–3):61–72. [PubMed: 2494447]
42. Ogiwara Y, Ohsawa T. Molecular analysis of the complete set of length mutations found in the plastomes of *Triticum-Aegilops* species. *Genome* 2002;45(5):956–962. [PubMed: 12416629]
43. Sinden RR, Zheng GX, Brankamp RG, Allen KN. On the deletion of inverted repeated DNA in *Escherichia coli*: Effects of length, thermal stability, and cruciform formation in vivo. *Genetics* 1991;129(4):991–1005. [PubMed: 1783300]
44. Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA. High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Biotechnology (N Y)* 1991;9(5):455–460. [PubMed: 1367310]
45. Scorer CA, Buckholz RG, Clare JJ, Romanos MA. The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast *Pichia pastoris*. *Gene* 1993;136(1–2):111–119. [PubMed: 8293993]
46. Daniell H, Guda C, McPherson DT, Zhang X, Xu J, Urry DW. Hyperexpression of a synthetic protein-based polymer gene. *Methods Mol Biol* 1997;63:359–371. [PubMed: 9113662]
47. Barr KA, Hopkins SA, Sreekrishna K. Protocol for efficient secretion of HSA developed from *Pichia pastoris*. *Pharm Eng* 1992;12:48–51.
48. Wong PoFoo C, Kaplan DL. Genetic engineering of fibrous proteins: Spider dragline silk and collagen. *Adv Drug Deliv Rev* 2002;54(8):1131–1143. [PubMed: 12384311]
49. Fahnestock SR, Irwin SL. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl Microbiol Biotechnol* 1997;47(1):23–32. [PubMed: 9035407]
50. Feschenko VV, Rajman LA, Lovett ST. Stabilization of perfect and imperfect tandem repeats by single-strand DNA exonucleases. *Proc Natl Acad Sci U S A* 2003;100(3):1134–1139. [PubMed: 12538867]
51. Petit MA, Dimpfl J, Radman M, Echols H. Control of large chromosomal duplications in *Escherichia coli* by the mismatch repair system. *Genetics* 1991;129(2):327–332. [PubMed: 1743481]
52. Albertini AM, Hofer M, Calos MP, Miller JH. On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell* 1982;29(2):319–328. [PubMed: 6288254]
53. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 1997;186(1):37–44. [PubMed: 9047342]
54. Meyer DE, Chilkoti A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system. *Biomacromolecules* 2002;3(2):357–367. [PubMed: 11888323]

55. Trabbic-Carlson K, Setton LA, Chilkoti A. Swelling and mechanical behaviors of chemically cross-linked hydrogels of elastin-like polypeptides. *Biomacromolecules* 2003;4:572–580. [PubMed: 12741772]
56. Goochee CF, Gramer MJ, Andersen DC, Bahr JB, Rasmussen JR. The oligosaccharides of glycoproteins: Bioprocess factors affecting oligosaccharide structure and their effect on glycoprotein properties. *Biotechnology (N Y)* 1991;9(12):1347–1355. [PubMed: 1367768]
57. Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischnewski H, Wildt S, Gerngross TU. Production of complex human glycoproteins in yeast. *Science* 2003;301(5637):1244–1246. [PubMed: 12947202]
58. Chiba Y, Jigami Y. Production of humanized glycoproteins in bacteria and yeasts. *Curr Opin Chem Biol* 2007;11(6):670–676. [PubMed: 17936668]
59. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* 2006;313(5792):1441–1443. [PubMed: 16960007]

Figure 1.

A. 1% TAE (Tris-acetate-EDTA) agarose gel. Ladder of concatamerization products, each band indicates different sized concatamers, differing by 75 bp (addition of one monomer cassette). Seven unique monomer units were pooled in equal amounts in the concatamerization reaction for random incorporation of the monomer units in the concatamerization of the **Yeast ELP** gene. **B.** 1% TAE (Tris-acetate-EDTA) agarose gel. *Xho* I / *Xba* I digested **Yeast ELP** (1575 bp), pPICZα-A (3.6 kb). DNA standard used was a 1Kb DNA ladder (NEB).

Figure 2.

Plasmid integration in *P. pastoris*. pPICZ α -A vector containing the **Yeast ELP** gene of interest was linearized within the 5' AOX1 region through digestion with restriction enzyme *Pme* I. Linearized plasmid was integrated by gene insertion into the host AOX1 region.

Figure 3.

Western blotting of elastin-mimetic protein **Yeast ELP** run on a 4–20% SDS-PAGE gel. **A.** c-myc primary antibody, goat-anti-mouse secondary antibody. **B.** anti-His primary antibody, goat-anti-mouse secondary antibody. **C.** Deglycosylation of **Yeast ELP**. Western blotting of elastin-mimetic protein **Yeast ELP** run on a 4–20% SDS-PAGE gel, anti-His primary antibody, goat-anti-mouse secondary antibody. Marker lane: Precision Plus Protein Kaleidoscope (Bio-Rad). Lane 1: Purified **Yeast ELP**, Lane 2: Purified and deglycosylated **Yeast ELP**. Analysis indicates no deglycosylation of the **Yeast ELP** protein.

Scheme 1.

Cloning strategy in the construction of the **Yeast ELP** gene. The seven unique double stranded DNA monomer cassettes were ligated together to form concatamer of random sequence. A target multimer of 1575 bp was selected, cloned into the recombinant plasmid, pPICZα-A and integrated into the *P. pastoris* host DNA.

Table 1

Amino acid sequence and related nucleic acid coding sequence for the **Yeast ELP** gene in *E. coli* (**A**) and *P. pastoris* (**B**) expression systems.

A											
Val	Pro	Gly	Val	Gly	Val	Pro	Gly	Val	Pro	Gly	Gly
GTA	CCT	GGT	GTT	GGC	GTT	CCG	GGT	GTA	CCA	Gly	GGC
Glu	Gly	Val	Pro	Gly	Val	Gly	Val	Val	Gly	Gly	Gly
GAA	GGT	GTA	CCG	GGT	GTT	GGC	GTA	GTA	GGC	Gly	GGC
B											
Val	Pro	Gly	Val	Gly	Val	Pro	Gly	Val	Pro	Gly	Gly
GTT	CCA	GGW	GTH	GGW	GTH	CCW	GGW	GTH	CCW	Gly	GGW
Glu	Gly	Val	Pro	Gly	Val	Gly	Val	Val	Gly	Gly	Gly
GAA	GGW	GTH	CCW	GGW	GTH	GGW	GTH	GTH	GGW	GGW	GGW

^A CCG is a high usage codon encoding proline in *E. coli* (24.5%), but extremely low in *P. pastoris* (3.9%)

^B Degenerate base **W** encodes for A or T while **H** encodes for A, T, or C (Sigma Genosys)

Table 2

Monomer Library

A		Val	Pro	Gly	Val	Gly	Val	Pro	Gly	Val	Gly	Val	Pro	Gly	Glu	Gly	Val	Pro	Gly	Val	Gly	Val	Pro	Gly
	M1:	GTT	CCA	GGA	GTC	GGA	GTT	CCT	GGT	GTT	GGA	GTA	CCA	GGT	GAA	GGT	GTT	CCT	GGT	GTA	GGA	GTC	CCT	GGT
	M2:	GTT	CCA	GGT	GTC	GGA	GTA	CCA	GGT	GTT	GGA	GTC	CCT	GGA	GAA	GGT	GTA	CCT	GGT	GTT	GGT	GTT	CCT	GGT
	M3:	GTT	CCA	GGT	GTC	GGT	GTA	CCT	GGT	GTA	GGT	GTT	CCT	GGT	GAA	GGT	GTA	CCA	GGT	GTC	GGT	GTA	CCT	GGT
	M4:	GTT	CCA	GGT	GTC	GGA	GTA	CCT	GGA	GTT	GGT	GTC	CCT	GGT	GAA	GGT	GTT	CCA	GGA	GTT	GGA	GTC	CCA	GGT
	M5:	GTT	CCA	GGT	GTT	GGA	GTT	CCT	GGA	GTT	GGT	GTC	CCT	GGA	GAA	GGA	GTT	CCT	GGT	GTT	GGA	GTA	CCT	GGT
	M6:	GTT	CCA	GGT	GTT	GGT	GTT	CCT	GGT	GTT	GGT	GTT	CCC	GGA	GAA	GGA	GTC	CCT	GGA	GTC	GGA	GTT	CCT	GGT
	M7:	GTT	CCA	GGT	GTT	GGA	GTT	CCT	GGA	GTA	GGT	GTT	CCT	GGA	GAA	GGA	GTA	CCT	GGT	GTT	GGT	GTA	CCA	GGT
B																								
	<div><div>Hind III</div><div>Bbs I</div><div>5'GCGCAAGCTTGAAGACGTTCCAGGAGTCGGAGTTCCTGGTGTGGAGTACCAGGTGAAGGTGTTCTGGTGTAGGAGTCCCTGGTGTAGGTGTTCCAAGAGACGGGATCC3'</div><div>BsmB I</div><div>BamH I</div></div>																							
C																								
	<div>5'-TCCAGGAGTCGGAGTTCCTGGTGTGGAGTACCAGGTGAAGGTGTTCTGGTGTAGGAGTCCCTGGTGTAGGTGTCTCAGCCTCAAGGACCACAACCTCATGGTCCACTCCACAAGGACCACATCCTCAGGGACCACATCCACAAGGT-3'</div>																							

A Nucleic acid sequence of the seven unique **Yeast ELP** monomer units created using the wobble base system and employed in multimerization of the **Yeast ELP** gene.

B Recognition (underlined) and cleavage sites (arrows) for restriction enzymes within the representative monomer, **M1**, repeat unit.

C Monomer repeat unit, **M1**, with cohesive ends generated from a *Bbs* I / *BsmB* I sequential digestion and used in the multimerization reaction. Note that *Bbs* I / *BsmB* I are type II restriction enzymes and cleave the DNA downstream of their recognition sites.

Table 3
Nucleic acid sequence of **Yeast ELP** adaptor with related amino acid coding sequence and restriction enzyme cut sites.

	Val	Pro	Ala	Val	Gly	Val	Pro		Pro	Ala	Val	Gly	
	CTCGAGAAAAGAGAGGCTGAAGCT	GTT	CCA	GCT	GTT	GGT	GTT	CCA	<u>AGAGAC</u> GGTACCC <u>GTCTCT</u>	CCA	GCT	GTT	GGT
	<i>Xho</i> I								<i>Bsm</i> BI		<i>Bsm</i> BI		

Adaptor was inserted at *Xho*I and *Xba*I sites within the pPICZ α -A expression vector. The **Yeast ELP** gene was inserted within the adaptor at the double *Bsm*B I sites.