# Appendix A: Manuscript Submitted to *Biomacromolecules*[1]

## A.1 Manuscript Title: *"A new cloning strategy for producing high molecular weight brush-forming poly(amino acids)"*

Douglas B. Henderson[a], Richey M. Davis[a], William A. Ducker[b] and Kevin E. Van Cott[c].

[a]Department of Chemical Engineering, Virginia Tech, Blacksburg, VA 24061
[b]Department of Chemistry, Virginia Tech, Blacksburg, VA 24061
[c]Department of Chemical Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588

### *A.1.1* Abstract

A simple and versatile strategy has been developed for producing de novo-designed, high molecular weight poly(amino acids) using recombinant DNA technology. This method is based on the assembly of modules of DNA that encode for blocks of poly(amino acids) into a commercially available expression vector; there is no need for custom-modified vectors, and no need for intermediate cloning vectors. Additionally, because the design of new protein-like biopolymers can be an iterative process, our method enables sequential modification of a poly(amino acid) product. That is, if a poly(amino acid) is produced and it is determined that the molecular weight needs to be increased, the expression vector can be re-opened and additional DNA modules can be inserted. For example, the polymer composition could be sequentially modified by adding to either its amino- or carboxy-terminus without having to re-start the gene assembly process from the beginning. The feasibility and simplicity of this method is shown in the production of a poly(amino acid) that was designed as a model for polymers that form brush layers on alumina surfaces. The success and flexibility of this method indicates that it can be applied for production of *de novo*-designed polypeptides in general. It is hoped that this method will contribute towards the rapid development of bio-inspired protein-based polymers for a variety of applications.

**Keywords:** biopolymer, poly(amino acid), block copolymer, polymer brush

### *A.1.2* Introduction

Brush-forming polymers have been used in a variety of applications, including biomolecule separations,[1-4] diagnostic and sensor devices,[5-8] biomaterials and tissue engineering,[9-11] specialty coatings and adhesives,[12-14] and microelectronics processing and nanotechnology.[15,16] These polymers have a well defined anchor block domain that adsorbs preferentially to the surface, and a non-adsorbing, soluble tail block domain that extends into the solvent. While the use of synthetic organic block co-polymers has long

---

[1] Henderson DB, Davis RM, Ducker WA and Van Cott KE. "A new cloning strategy for producing high molecular weight brush-forming poly(amino acids)". *Biomacromolecules* (submitted 12/2004).

been established, harnessing biotechnology to produce brush-forming biopolymers has the potential to increase the number of available applications. Nature uses polymers of 20 amino acids to create a remarkable array of complex polypeptides that perform functions such as electron transport, water and ion transport, ligand binding, and catalysis. Through recombinant DNA technology, there exists the potential to harness a cell's protein synthesis machinery to produce a polymer of exactly specified amino acid sequence, therefore controlling the polymer's composition at a level unequaled by conventional organic polymer synthesis, and thus enabling the incorporation of unique structures and functions in these brush layers.

In spite of these possibilities, there is a gap that prevents efficient production of synthetic poly(amino acids) (PAA). Solid phase peptide synthesis is limited in usefulness because of the inherent limit to the size of the polypeptide product. Groundbreaking work from several research groups has demonstrated that recombinant DNA technology can be used to produce unique synthetic protein polymers for tissue engineering and material science applications.[17] However, in the design and production of new polymers with optimal brush-forming properties it is desirable to be able to use an *iterative* approach, where the molecular weight of each block is varied to produce optimal surface coverage and brush extension. The recombinant DNA strategies developed by Cappello *et al*.,[18] McMillan *et al*.,[19] Meyer and Chilkoti,[20] and Won and Barron[21] for structural protein polymers are a good starting point, but these methods are not easily adapted for an iterative process. We set out to advance the discovery and design of novel PAA brush-forming polymers by developing a new cloning strategy that simplifies the process of synthetic gene-assembly and subsequent PAA production in genetically engineered organisms.

For the task of producing high molecular weight, surface-active PAA's we developed a cloning strategy with the goal of incorporating the following features:

1) **Modular** – DNA molecules can be synthesized and inserted directly into an expression vector in modules; an intermediate cloning vector is not needed.
2) **Sequentially-modifiable** – DNA modules can be *sequentially* inserted directly into the expression vector. That is, if a PAA is produced and it is determined that the molecular weight needs to be increased, the vector can be re-opened and

additional DNA modules can be inserted. For example, the new insertion points can be either at the 5' or 3' end of the previously inserted DNA modules. Thus, the PAA composition can be sequentially modified by adding to either its amino- or carboxy-terminus without having to re-start the gene assembly process from the beginning.

3) **Minimal Amino Acid Requirement** – The composition of the resulting PAA should be minimally impacted by the restriction enzymes and cloning methods used.
4) **Versatility and Transferability** - The cloning strategy should be simple, using only commercially available expression vectors and molecular biology reagents, and should be easily transferable to other laboratories that have basic molecular biology and protein production expertise.

In this work, we report on the development of a method that incorporates these features, and the application of this strategy to the production of a block co-polymer that was designed to function as a brush-forming polymer on metal oxide surfaces such as alumina and titania. We present data showing the feasibility and simplicity of this strategy, and show that it can be applied for facile production of *de novo*-designed protein polymers in general, not just for brush-forming PAA's.

### *A.1.3* Experimental

**Materials**. pBAD/TOPO® ThioFusion™ Expression Kit, One Shot® TOP10 Chemically Competent *E. coli* strain, OneShot® MAX Efficiency® DH5 $T_1^R$ chemically competent *E. coli* strain, NuPAGE® Novex 10% Bis-Tris polyacrylamide gels, LB Broth Base, UltraPure™ agarose, UltraPure™ ethidium bromide solution, SeeBlue® Plus2 Pre-Stained Standard, Colloidal Blue Staining Kit, and Trx Forward/pBAD Reverse sequencing primers were purchased from Invitrogen Corp. (Carlsbad, CA). Sequenase™ Version 2.0 DNA polymerase was obtained from USB Corporation (Cleveland, OH). *Taq* PCR Core Kit, PAGE-purified synthetic oligonucleotides, QIAprep® Spin Miniprep, QIAquick® Gel Extraction/PCR Purification and Ni-NTA Spin Kits were purchased from Qiagen, Inc. (Valencia, CA). *PfuTurbo*® DNA polymerase was purchased from Stratagene (La Jolla, CA). LB Agar, CelLytic™ B Bacterial Cell Lysis Extraction Reagent, and L-(+)-arabinose were obtained from Sigma (St. Louis, MO). Ampicillin and 0.2 μm pre-sterilized syringe filters were purchased from VWR International (So. Plainfield, NJ). Slide-A-Lyzer® Dialysis Cassettes (3.5 kDa molecular weight cutoff)

were purchased from Pierce (Rockford, IL). All restriction enzymes, Calf Intestinal Alkaline Phosphatase (CIP), T4 Polynucleotide Kinase, T4 DNA Ligase, 100 bp DNA and 1 kbp DNA Ladders were purchased from New England Biolabs (Beverly, MA). DNA primers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA).

**General Methods**. The procedures for manipulation of DNA, transformation, cell growth, product expression and analysis were adapted from published literature[22,23] or from instructions provided by product manufacturers. All reagents for the manipulation of DNA were sterile and DNAse/RNAse free. Enzymatic manipulations of DNA were conducted in reagent buffers supplied by the manufacturer. PCR amplification and DNA extension reactions were performed in an Omnigene thermal cycler from Hybaid (United Kingdom). Automated DNA sequence analysis was performed on a Perkin-Elmer ABI Prism model 377 DNA Sequencer at Cleveland Genomics, Inc. (Cleveland, OH). Cells were lysed using a Tekmar Sonic Disruptor with a microtip sonicator (Mason, OH). Product analysis was conducted by SDS-PAGE using pre-cast NuPAGE Novex 10% Bis-Tris polyacrylamide gels on an Xcell Surelock Mini-Cell apparatus from Invitrogen Corp. The gels were visualized using coomassie blue staining. Agarose and polyacrylamide gel images were captured using a Polaroid GelCam and Microtek ScanMaker X6EL scanner, respectively. Polypeptide product and DNA concentrations were calculated from the optical density (OD) obtained using a Milton Roy Spectronic 1201 UV spectrophotometer (Ivyland, PA).

**Cloning Strategy.** The general strategy that we present in this work is illustrated in Figure A.1, and can be summarized as follows:

**Step 1.** The first recombinant DNA molecule (DNA Module #1) is inserted into an unmodified expression vector. DNA Module #1 contains a unique *Sfo*I restriction enzyme site (5'-*ggc*|*gcc*-3'). This *Sfo*I site can be located anywhere within DNA Module #1. Since this is where the subsequent blocks will be inserted, it will likely be located at either the 5' or 3' end, so that the resulting polypeptide can be modified at either the carboxyl or amino terminus. In this example, we have positioned the *Sfo*I site at the 3' end so that the resulting polypeptide is sequentially modified at its carboxyl terminus. The DNA sequence 5'-*ggc gcc*-3' encodes for Gly-Ala.

**Step 2.** The expression vector containing DNA Module #1 is then digested with *Sfo*I, and DNA Module #2 is inserted *via* blunt-end ligation. An appropriate double restriction digest or PCR can be used after each insertion to confirm insert size and orientation. The only requirements for Module #2 are that there are no internal *Sfo*I sites, and in this example the DNA encoding Module #2 must have a *ggc* codon (Gly) at the 3' end so that the unique *Sfo*I site is regenerated for subsequent insertions. The recombinant protein can be expressed, or it can be modified again. Step 2 is repeated until the gene encoding for the desired PAA is complete.

As with any cloning method, the DNA should be sequenced before proceeding to expression. If Gly-Ala sequences are required to occur within any of the polypeptide

blocks, the degeneracy of the genetic code can be used to ensure that these do not contain an *Sfo*I recognition site.

**Choosing an Expression Vector**. Our goal was to use only commercially available expression vectors, so we required that the expression vectors contain no internal *Sfo*I sites. Also, for convenience in our initial development work, it was desired that the vector express the PAA with an N-terminal fusion protein tag that would facilitate purification and detection during development. In this work, we used the pBAD/Thio-TOPO® expression vector from Invitrogen.

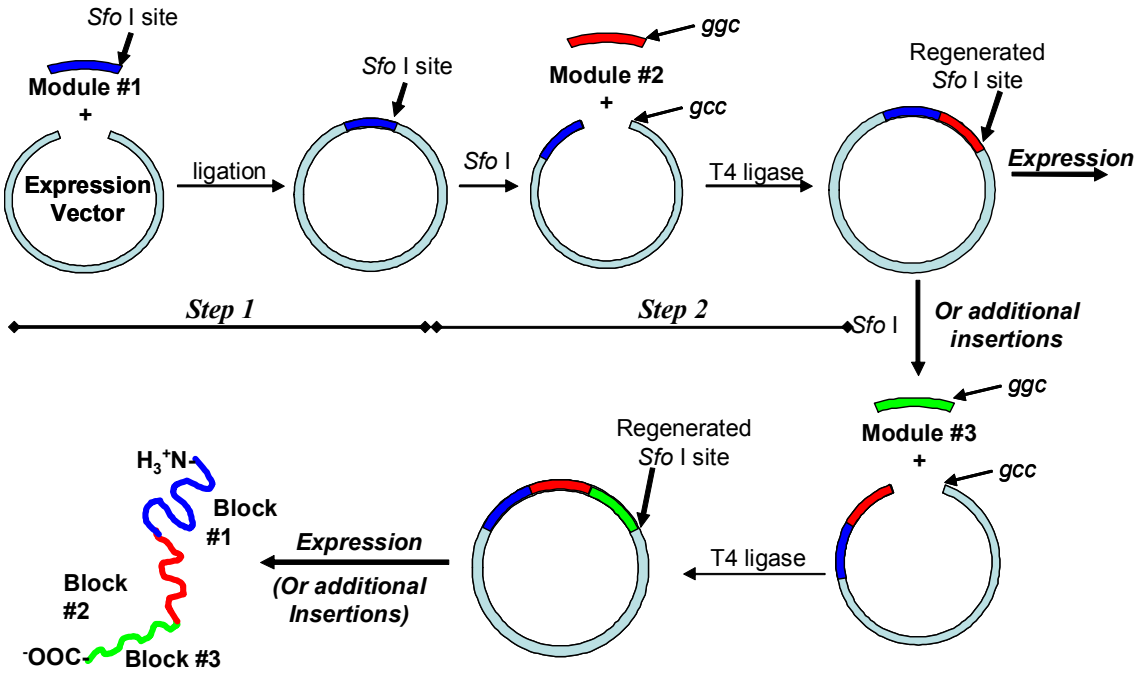



**Figure A.1 Illustration of the method used in this work to generate high molecular weight PAA's.** DNA molecules can be sequentially inserted into the expression plasmid until the desired polypeptide composition is obtained. The position of the *Sfo*I restriction enzyme site in the first module determines how the polypeptide is modified. After each insertion, the protein can be expressed, or it can be further modified.

**DNA Modules.** In general, modules for the anchor and tail blocks were made by annealing two overlapping single-stranded, synthetic oligonucleotides. The sense and antisense oligonucleotides were mixed and then heated to 95°C. The mixture was cooled down slowly to 5°C below the estimated melting temperature ($T_m$) of the complimentary oligonucleotides and held for 5 minutes to anneal the complimentary bases. A solution of dNTPs was added, followed by the addition of Sequenase™ to begin the 'fill-in' reaction. The double-stranded products were amplified by PCR as needed.

The sequence of the first DNA module inserted into the expression vector is given in Table A.1, along with the encoded amino acid sequence. The unique *Sfo*I recognition site (5' *ggc*|*gcc* 3') was placed towards the 3' end. This allowed for subsequent insertions of DNA modules downstream of the first module. The base pairs encoding for a hydroxylamine chemical cleavage site (Asn-Gly) were included at the 5' end of the

module to allow for future removal of N-terminal fusion tag after PAA expression. A stop codon (*tga*) was included at the 3' end of the module. For the first DNA module to be inserted, 3' A-overhangs were added to facilitate TOPO® cloning by adding 1U of *Taq* DNA polymerase and incubating the mixture at 72°C (15 minutes).

**Table A.1 DNA sequences and the corresponding amino acid sequences.** Underlined base pairs denote the *Sfo*I recognition site in DNA Module #1 and the *Bsu*36I site in DNA Module #4.

| | |
|---|---|
| **DNA Module #1 (87 bp)** Tail Block #1 | 5'- aac ggc aag agc gaa ggc gag aag cgc agc gat aaa gag ggt aaa gag ggc ggt gac cgc aaa gag agc gat ggt cgc ggc gcc tga -3' |
| | NGKSEGEKRSDKEGKEGGDRKESDGRGA* |
| **DNA Module #2 (156 bp)** Tail Blocks # 2&3 | 5'-ggc aaa gag ggc gag aag ggg gag ggc cgt ggc cgc gag cgc ggc gat ggc gat cgt ggt aag gaa aaa gac ggt ggc ggc gac cgt ggc aaa gac agc agc aag gat cgt gac agc tct gac cgt ggt gaa aaa agc gac aag gac aag ggc ggc-3' |
| | GKEGEKGEGRGRERGDGDRGKEKDGGGDRGKDSSKDR DSSDRGEKSDKDKGG |
| **DNA Module #3 (78 bp)** Tail Block #4 | 5'- ggc gaa tct cgc gac aag ggc ggt gaa aaa cgt gat tct gat cgc tct gaa cgt agc gat cgc agc agc gaa cgc ggc -3' |
| | GESRDKGGEKRDSDRSERSDRSSERG |
| **DNA Module #4 (48 bp)** Anchor Block | 5'-gcc ggc gga ggc ggt gaa gag gag gag gaa gag gaa gaa gag gaa gcc tga ggc tga -3' |
| | AGGGGEEEEEEEEEA*G* |
| **Amino acid sequence of finished PAA** (including N-terminal HP-Thioredoxin fusion tag) | MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCK MIAPILDGIADGYQGKLTVAKLNIDHNPGTAPKYGIRGIP TLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGD DDDKLALNGKSEGEKRSDKEGKEGGDRKESDGRGGKE GEKGEGRGRERGDGDRGKEKDGGGDRGKDSSKDRDSS DRGEKSDKDKGGGESRDKGGEKRDSDRSERSDRSSERG AGGGGEEEEEEEEEA*G*A* |

A longer recombinant DNA molecule was used for the second module (Table A.1). By using the methods described in this work, we had previously made an expression plasmid with two concatenated tail block modules (78 bp each), here denoted

as Tail DNA Modules #2 and #3. We PCR-amplified the concatenated sequence of Tail Modules #2 and #3 from this existing recombinant vector (see *Supporting Information*). The PCR was conducted using *PfuTurbo*® DNA polymerase (leaving blunt ends) and the corresponding amplification primers. PCR was performed with 35 cycles at 95°C for 30 seconds (denaturing), 60°C for 30 seconds (annealing), and 72°C for 50 seconds (elongation). PCR products were pooled together and the DNA was purified using QIAquick® PCR purification kit and dried down using a Savant SpeedVac. DNA was then re-suspended in T4 DNA Ligase Buffer and the 5' ends were phosphorylated with T4 polynucleotide kinase. DNA was purified by agarose electrophoresis and extracted using the QIAquick® Gel Extraction kit into water.

Overlapping, single-stranded, synthetic oligonucleotides were used to construct DNA Module #3 and DNA Module #4 (Table A.1). The products were then used as templates for PCR amplification to make larger amounts of each DNA molecule, and the 5' ends were phosphorylated as described above. DNA Module #4 was designed so that the *Sfo*I site would be regenerated at the junction between DNA Module #3 and #4. The 3' end of DNA Module #4 (Anchor Block) included a *Bsu*36I site (5'-cc|tgagg-3'). This was included to add more flexibility for the sequential modifications, allowing for more additions to the 3' end of this module (see *Supporting Information*).

The order of insertion of the Tail and Anchor modules for this work was Tail Block #1 (87 bp), Tail Block #2&3 (156 bp), Tail Block #4 (78 bp), and finally the Anchor Block (48 bp). A schematic of the structure of the PAA is given in Figure A.2.

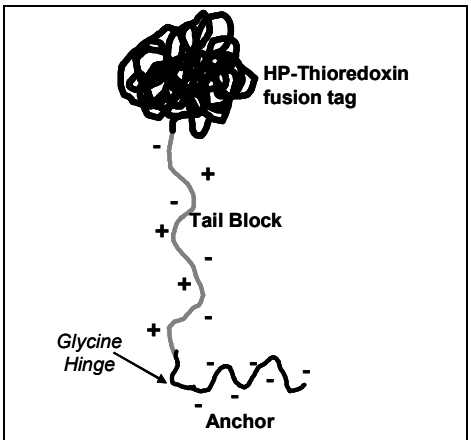



**Figure A.2 Schematic of the general structure of the designed synthetic polypeptide.**

**Insertion of the First DNA Module – Tail Block #1.** All cloning steps were conducted within the *E. coli* strain DH5 $T_1^R$ to minimize possible recombination events and loss of the recombinant gene. The first DNA module, Tail Block #1, was inserted into the expression vector pBAD/Thio-TOPO® by TOPO® cloning per the manufacturer's instructions. Insertion of PCR products with 3' A overhangs by TOPO® cloning is quick and convenient, but the orientation of the insert must be checked. The cloning reaction was then transformed into One Shot® MAX Efficiency® DH5 $T_1^R$ Chemically Competent *E. coli* and selected on LB Agar containing 100 μg/mL ampicillin and

incubated overnight (16 hours) at 37°C. Plasmids of positive clones were analyzed by colony PCR to determine insert size and orientation (described below). Colonies were chosen and used to inoculate 5 mL cultures of LB Broth containing 100 μg/mL ampicillin. Cultures were incubated with shaking (~200 rpm) overnight (~16 hours) at 37°C, and 3 mL of each culture was used to isolate plasmid using the QIAprep® Spin Miniprep kit. Plasmid was eluted from the spin columns into 40 μl of sterile water to a concentration of ~100 ng/μL. Stocks of positive clones were made in 15% glycerol and stored at -80°C. This plasmid is denoted as $pTrxT_1$.

**Insertion of Subsequent DNA Modules.** We sequentially inserted DNA modules encoding for the remaining tail blocks and the anchor block *via* blunt end ligation after linearization of the expression vector by digestion with *Sfo*I, which cuts the plasmid DNA at the unique (5' *ggc*|*gcc* 3') site.

The modified vector was re-suspended in 10X NEB Buffer 2, *Sfo*I (1 unit per 1μg DNA) and sterile water to a concentration of 0.5 μg DNA/10 μL and incubated overnight (~16 hours) at 37°C. The restriction enzyme was deactivated by heating to 65°C for 20 minutes. The 5'ends of the linearized vector were dephosphorylated to prevent self-ligation by adding 0.5 U CIP per μg DNA. The reaction was incubated at 37°C for 4 hours and then terminated by heating to 75°C for 10 minutes. The reactions were loaded onto an agarose gel for purification. The linear, dephosphorylated vector was extracted into sterile water using the QIAquick® Gel Extraction kit.

The DNA modules and the linearized vector were ligated together using T4 DNA Ligase. Blunt-end ligations were conducted at an insert:vector molar ratio of 3:1 in a 20 μL reaction mixture. We found this insert:vector ratio to be optimal for our work. This mixture consisted of 10X ligase buffer, 40-60 fmol of linearized vector, 120-180 fmol of DNA module, 1 Weiss unit of T4 DNA Ligase, and sterile water. The ligation mixture was incubated for ~24 hours at 14°C. The total ligation reaction was then transformed into One Shot® MAX Efficiency® DH5 $T_1^R$ Chemically Competent *E. coli* and selected on LB Agar containing 100μg/mL ampicillin after overnight incubation (~16 hours) at 37°C. The expression vectors we generated were named as follows: $pTrxT_3$ (3 tail block modules), $pTrxT_4$ (4 tail block modules), and $pTrxT_4A$ (4 tail block modules and 1 anchor block module).

**Analysis of Insert Size – Colony PCR.** A minimum of 20 colonies from each transformation were chosen and used to inoculate 50 μl of LB Broth containing 100 μg/mL ampicillin. Cultures were incubated for 3 hours at 37°C, and then 1 μL of each culture was directly added to a 20 μL PCR reaction mixture containing 10x PCR buffer, dNTPs, primers (Table A.2), sterile water, and *Taq* DNA polymerase. These primers were designed to anneal to sequences in the vector flanking the inserted DNA modules. PCR was performed with 20 cycles at 95°C for 45 seconds (denaturing), 50°C for 30 seconds (annealing), and 72°C for 30 seconds (elongation). The samples from each PCR reaction were analyzed using agarose gel electrophoresis for the correct size insert.

| Table A.2 DNA primers used to analyze insertion size. | |
|---|---|
| **Trx Forward** | 5'-ttcctcgacgctaacctg-3' |
| **pBADRev2 Reverse** | 5'-ctcatccgccaaaacagcc-3' |

**Analysis of Insert Orientation – Colony PCR.** Colonies containing the correct size insert were also analyzed by colony PCR to check for insert orientation. One microliter of each culture was directly added to a 20 µL PCR reaction mixture containing 10x PCR buffer, dNTPs, primers, sterile water and *Taq* DNA polymerase. PCR was performed with 20 cycles at 95°C for 45 seconds (denaturing), 53°C for 30 seconds (annealing), and 72°C for 30 seconds (elongation). The primers used are given in Table A.3. For each insertion, Thio Forward was combined with the reverse primer corresponding to the module inserted. For example, after inserting DNA Module #3, Thio Forward and Tail3 Reverse were used. The samples from each PCR reaction were analyzed using agarose gel electrophoresis for the correct orientation.

| Table A.3 DNA primers used to analyze insertion orientation. | |
|---|---|
| Thio Forward | 5'-tctgataaaattattcatctgact-3' |
| Tail1 Reverse | 5'-tcaggcgccgcgaccatcgctctc-3’ |
| Tail3 Reverse | 5’-gccgcccttgtccttgtcgc-3’ |
| Tail4 Reverse | 5’-gccgcgttcgctgctgcga-3’ |
| Anchor Reverse | 5’-tcagcctcaggcttcctc-3’ |

**Small Scale Product Expression**. After each DNA insertion step, the recombinant plasmids were transformed into the expression strain, One Shot® TOP10 Chemically Competent *E. coli*, according to the manufacturer’s instructions, and analyzed for protein expression. The transformation mixture was spread on LB medium agar plates containing 100 µg/mL ampicillin and incubated overnight at 37°C. A single colony was then used to inoculate 5 mL of LB medium containing 100 µg/mL ampicillin and incubated overnight at 37°C with shaking (225-250 rpm). The next day, 10 mL of LB medium containing 100 µg/mL ampicillin were inoculated with 500 µL of the overnight culture and incubated at 37°C with shaking (225-250 rpm) to midlog phase ($OD_{600}$ = ~0.5). Expression was induced with the addition of L-(+)-arabinose to a final concentration of 0.02%. A 500 µL aliquot was removed, and the cells were pelleted by centrifugation at maximum speed in a microcentrifuge tube for 30 seconds. The supernatant was removed and the pellet was stored at -80°C. This is the zero time point. Expression was continued for 5-6 hours and the cells were harvested in the same manner. Each pellet was re-suspended in 125 µL CelLytic™B Bacterial Cell Lysis Extraction Reagent and the cells were lysed repeating a

freeze (-80°C) / thaw (42°C) / sonicate (10 seconds) cycle four times. The cell debris was pelleted by centrifugation at maximum speed for 15 minutes and the soluble cell extract was removed. The extracts were analyzed by SDS-PAGE and visualized using coomassie blue staining.

**Larger Scale Expression.** In a 250 mL culture flask, 50 mL of LB Broth containing 100 µg/mL ampicillin were inoculated with stock TOP10 bacterial culture containing the recombinant vector pTrxT$_4$A. The culture was incubated overnight at 37°C with shaking at 200 rpm. The culture was pelleted in a 50 mL centrifuge tube at 3000 X *g* for 10 minutes at 4°C. The supernatant was removed and the pellet was re-suspended in 10 mL of fresh LB media. Four 1 L culture flasks with 250 mL of fresh LB media containing 100 µg/mL ampicillin were inoculated with 2 mL of this cell suspension. The culture was incubated at 37°C with shaking at 200 rpm until mid-log phase ($OD_{600}$ ~ 0.5). Expression of the product was induced by the addition of L-arabinose to a final concentration of 0.02%, and the cultures were incubated for an additional 5-6 hours at 32°C. The cells were harvested by centrifuging at 3000 X *g* at 4°C, and the pellets were stored at -80°C.

**TrxT$_4$A Purification – Ni-NTA Spin Columns.** Approximately 1 gram cell pellet was re-suspended in 7 mL of Lysis Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.8). Cells were lysed repeating a freeze (-80°C) / thaw (42°C) / sonicate (10 seconds) cycle four times. The cell debris was pelleted by centrifugation at maximum speed for 15 minutes. The soluble cell extract was removed, and 600 µL of the extract was loaded onto a Ni-NTA metal affinity spin column (Qiagen). The column was centrifuged at 700 X *g* for 2 minutes. The flow-through was collected. The spin column was washed with 600 µL Wash Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 6.0) and then centrifuged at 700 X *g* for 2 minutes. This was repeated with another 600 µL of Wash Buffer. The TrxT$_4$A product was eluted in two washes with 200 µL Elution Buffer (20 mM sodium phosphate, 500 mM NaCl, pH 4.0). After centrifugation at 700 X *g* for 2 minutes, the eluate was collected and was adjusted to pH~7 with 1 M Tris, pH 8.0. The eluate was then dialyzed against water using Slide-A-Lyzer® Dialysis Cassettes (3.5 kDa molecular weight cutoff) at 4°C for 24 hours. The samples were collected and product concentration was estimated by measuring the OD at 280 nm (assuming an extinction coefficient of 0.548 l/g·cm calculated based on the amino acid sequence of the product.)[24] Samples were stored at -80°C.

**HPLC Analysis of the Product.** Chromatographic analysis of protein samples was obtained on a Waters Alliance® 2695 HPLC system with cooled autosampler chamber and column heater, a Waters 2996 photodiode array detector, and a Waters Fraction Collector III. High performance ion exchange chromatography (HPIEC) was performed on a Vydac 301VHP column (DEAE stationary phase, 5 µm particle size, 900 Å pore size, L = 5 cm, ID = 7.5 mm, column temperature = 24°C) with a Vydac CPF10 column prefilter (P.J. Cobert Associates, St. Louis, MO). The mobile phases used were A: 7.5mM Tris-HCl, pH 8.0 in 25% acetonitrile, and B: 500 mM NaCl in A. Gradient programs are given in individual chromatograms.

**Mass Spectrometry.** Samples were analyzed by MALDI-TOF Mass Spectrometry using an Applied Biosystems Voyager DE Pro at the Nebraska Center for Mass Spectrometry at

the University of Nebraska-Lincoln. Spectra were acquired manually in linear positive ion detection mode, with an accelerating voltage of 25 kV and a 300 nsec extraction delay time. A minimum of 400 shots per sample were used to obtain the final spectrum. Spectra are presented without any smoothing. Before analysis, samples were mixed 1:1 with sinapic acid (Fluka) matrix solution (10 mg/mL sinapic acid in 50% acetonitrile/0.05% TFA). 1 µL of sample/matrix was spotted onto a 100-well stainless steel sample plate (ABI, Framingham, MA) and air dried. The instrument was calibrated using the following protein standards (Sigma, MS-CAL1): bovine insulin, equine cytochrome C, equine apomyoglobin and rabbit muscle aldolase.

### *A.1.4* Results and Discussion

The objective of this work was to demonstrate a simple and versatile method for designing and producing synthetic protein-like block co-polymers that can be used to form brush layers on surfaces. For the production of high molecular weight polypeptides with a defined amino acid sequence and composition, the protein synthesis machinery of living organisms can be exploited. McGrath *et al.* compared the use of recombinant DNA methods to chemical synthesis methods, and showed that the biosynthetic approach to producing block copolypeptides offers better efficiency and a higher level of control over monomer sequence and molecular weight.[25] This and other recent advances in molecular bioengineering and polymer chemistry indicate that the area of "bio-inspired polymeric materials" is rapidly growing.[26]

The size of the resulting *de novo*-designed polypeptides made using conventional recombinant DNA techniques is limited by the difficulty inherent to assembling large synthetic DNA molecules in an expression vector. Several groups have reported new strategies for making high molecular weight protein polymers, which are based on working with modules of DNA that encode for portions of the polypeptide product.[18-21] In particular, the work reported by Won and Barron[21] was of great interest to us as we began the process of making protein-like block copolymers. They assembled and concatenated modules of synthetic DNA molecules and were able to produce high molecular weight, monodisperse polymers. However, the drawbacks of this method in our hands were that it required a custom-modified intermediate cloning vector, and that the DNA encoding the protein product in the expression vector could not be easily modified without going back to the intermediate cloning vector.

While it is certainly still possible to use the method of Won and Barron to produce brush forming PAA's, we designed a method that is suited for an iterative

process of making block copolymers of varying anchor and tail block sizes and compositions. First, we wanted a method that used only commercially available molecular biology reagents. Our rationale was that a method that required custom-modification of plasmids would not likely be easily adapted by other labs with little experience in molecular biology. Secondly, once a PAA was made, we wanted to have the option of incrementally changing the composition and size by adding DNA modules to the expression vector, without having to start the whole cloning process over again or go back to an intermediate cloning vector. And thirdly, we wanted to have a method that had little to no restrictions on the amino acid sequence, so that a variety of synthetic PAA's could be produced.

The strategy we devised to address these concerns is outlined in Figure A.1. Borrowing from previously mentioned published reports, this strategy is based on assembling modules of synthetic DNA molecules that were synthesized by conventional DNA synthesis and molecular biology techniques. Our method does not have any special requirements for how the module is generated; in this work, we assembled the modules by purchasing sense and antisense oligonucleotides, annealing, and then polymerizing with DNA polymerase to form the double-stranded molecule. The key feature of our method is that we are using an unmodified, commercially available expression vector that contains no *Sfo*I restriction enzyme sites. The *Sfo*I recognition sequence is 5'-ggc|gcc-3', and digestion results in blunt ends. Several features made the use of *Sfo*I a desirable restriction site to use in our method:

1. The *Sfo*I site is relatively uncommon in prokaryotic expression vectors;
2. The *Sfo*I site is 6 bases, and digestion and subsequent insertion events will not result in any frame-shifts for the encoded polypeptide;
3. The *Sfo*I bases encode for the amino acid sequence Gly-Ala – the two naturally occurring amino acids with the smallest side chains (–H and $–CH_3$, respectively).

The presence of Gly and Ala residues in a synthetic polypeptide will have a minimal impact on the polypeptide structure. In the case of brush-forming polypeptides, the

addition of Gly residues between modules can contribute to the conformational flexibility, which is advantageous.

A search of suppliers' catalogs found several bacterial expression vectors that contained no internal *Sfo*I sites. In this work, we used the pBAD/Thio-TOPO® expression vector, which results in a low basal level expression but high levels of expression when the inducer (L-(+)-arabinose) is present.[27,28] This vector expresses recombinant proteins as fusions containing an N-terminal HP-Thioredoxin (Trx) tag in *E. coli*. The HP-Thioredoxin fusion tag increases the solubility of the recombinant protein, and it helps simplify purification and detection.[29] We have shown separately that the pCR®T7/NT-TOPO® vector (N-terminal poly-His fusion tag) works just as well (see *Supporting Information*). Other prokaryotic expression vectors from Invitrogen are also compatible with our strategy, such as the pBAD/TOPO® TA, and the pCR®T7/CT-TOPO®, which have C-terminal fusion tags. We also note that the following commercially available expression vectors used with the methylotrophic yeast *Pichia pastoris* do not contain *Sfo*I sites and could be conceivably be used to produce sequentially modified proteins in yeast: pPICZ, pPICZα, pPIC6, pPIC6α, pPICZ-E, pPICZα-E, pGAPZ, and pGAPZα.

Insertion of the first DNA Module incorporates a unique *Sfo*I site into the expression vector. The placement of the *Sfo*I site and how it is regenerated can be used to control how the polypeptide composition is incrementally changed. In this work, as outlined in Figure A.1, the *Sfo*I site was located at the 3' end of DNA Module #1. After insertion, the vector was digested with *Sfo*I and DNA Modules #2 and #3 were inserted, which both had a *ggc* Gly codon at the 3' end. The correctly oriented insertion process regenerated the unique *Sfo*I site at the 3' end of the DNA construct, and the resulting polymer is "grown" by additions to the C-terminus. Thus, in this version of our strategy, a Gly residue is required to exist at the C-terminus of the blocks of polypeptides that are encoded by DNA Modules #2 and #3. For this model PAA, we regenerated the *Sfo*I site after addition DNA Module #4 (Anchor Block) by placing a *gcc* codon (Ala) at the 5' end of the module. This was done because our design for this model PAA included only one anchor block module. We wanted the flexibility of using this tail block structure with different anchor blocks in the future and having the ability to increase the size of the tail

domain if necessary. We also designed for potential sequential addition of more DNA modules to the 3' end of the Anchor Block DNA module to make a PAA with the structure of a tri-block copolymer. We included a unique *Bsu*36I site (5'-cc|tgagg-3') at the 3' end of DNA Module #4. However, use of this *Bsu*36I to grow a PAA is limited to C-terminus additions in our case, since *tga* is a stop codon (see *Supporting Information*).

There is flexibility in this strategy in using the location of the initial *Sfo*I site and how that site is regenerated to grow the PAA. In Figure A.3, the N-terminal and C-terminal blocks of the PAA are set, and a number of "spacer" blocks can be inserted to vary the spacing between the terminal blocks. The *Sfo*I site in the first DNA module is located at the 5' end of DNA Module #1, and the *Sfo*I site is regenerated by insertion of subsequent modules having a *ggc* codon at the 3' end. Thus DNA modules #1 and #2 comprise the C-terminal and N-terminal blocks, respectively, and all subsequent DNA modules encode the "spacer" blocks. We have demonstrated the feasibility of this variation in our laboratory (see *Supporting Information*).

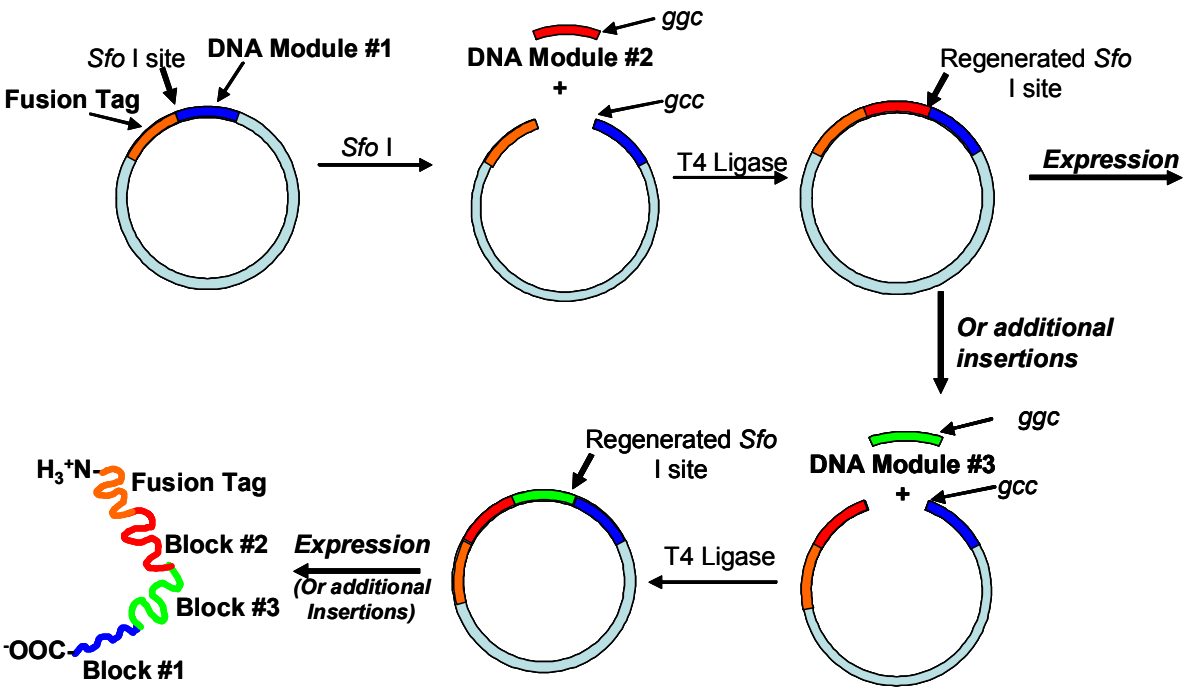



**Figure A.3 A variation on the theme of sequentially modifying the polypeptide.** In this case, the first DNA module encodes for the C-terminus of the polypeptide and the second DNA module encodes for the N-terminus. All subsequent modules that are inserted encode "spacer" blocks between the terminal blocks.

The polypeptide sequence that we set out to produce is given in Table A.1. This polypeptide was designed as a model for making brush-forming polymers for alumina

surfaces. We have previously shown that synthetic polypeptides with a poly-glutamate anchor block and a proline-rich tail block adsorbs to alumina and can be used to control the surface forces.[30] In this work, we used a C-terminal anchor block consisting of 10 Glu residues with an Ala-$Gly_4$ "linker" region between the anchor and tail blocks (Figure A.2). Glycine-rich regions maximize the conformational freedom of a polypeptide backbone, and four glycines plus an alanine were added at the amino-terminus of the anchor block to act as a flexible "hinge" between the anchor and tail blocks. Our tail block domain was designed to be very hydrophilic and zwitterionic in nature, consisting of Gly, Ser, Asp, Glu, Lys, and Arg residues, with a zero net charge at neutral pH. Our strategy required inclusion of Gly residues in between each block, and this fit naturally into the composition design of the tail blocks; an Ala residue was at the C-terminus of the final PAA product. An Asn-Gly dipeptide was included at the junction of the HP-Thioredoxin fusion tag and the N-terminus of the tail block region. This was designed to allow for future chemical cleavage of the fusion tag from the PAA by hydroxylamine.[31]

Our method is based on digestion with *Sfo*I, an enzyme that produces blunt ends, and insertion of blunt-end DNA modules into the linearized vector. Blunt-end ligation is non-directional, and less efficient than ligation of DNA with cohesive ends. Several steps were taken to optimize this process. First, to minimize recircularization of the vector, the linearized vector was treated with phosphatase to remove the 5' phosphate groups. Additionally, we determined the optimal insert:vector molar ratio that resulted in the most single-module insertions in the correct orientation.

Several methods can be used to monitor the blunt-end ligation products. Since multiple copies of the insert can be ligated, the size of the insert must be checked. This can be easily accomplished by colony PCR after transformation, using primers (Table A.2) designed to hybridize upstream and downstream of the insert (Figure A.4A). The orientation of the insert must also be confirmed, and this can be done by restriction enzyme digest or colony PCR. Double restriction digest with *Sfo*I and an endonuclease that cuts at an upstream site in the vector (*Nco*I in our case) can be used, as illustrated in Figure A.4B to check the size and orientation of the insert. Colony PCR is a more facile way to confirm the size and orientation, however. The primers given in Table A.3 were used to confirm the insert size and orientation, as illustrated in Figure A.4C. The colony PCR results of the insert size confirmation for each step of the process in production of the pTrxT$_4$A expression vector is given in Figure A.5.

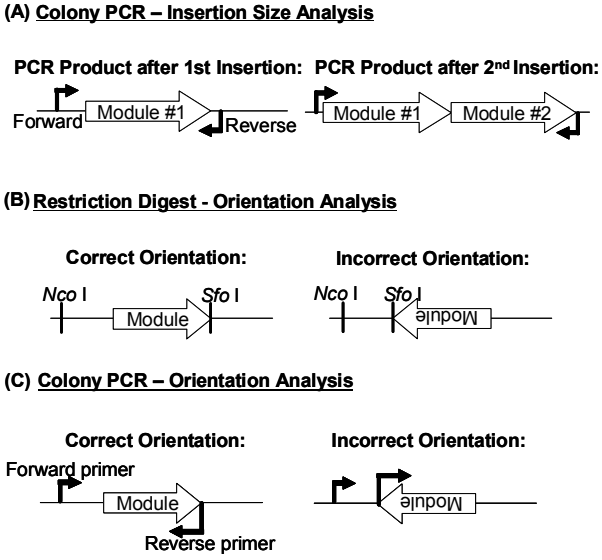
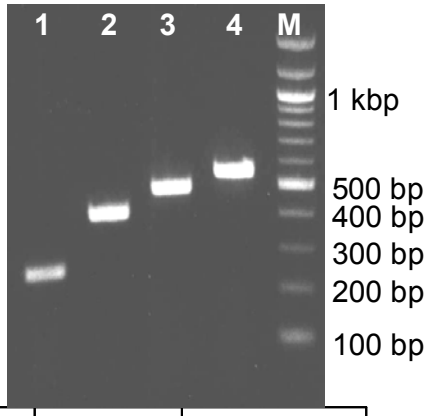



**Figure A.4 Methods to analyze insertions.** (A) colony PCR to confirm the size of the blunt-end ligation insert, (B) double restriction digest to confirm the orientation of the insert, and (C) colony PCR to check the orientation of the blunt-end ligation insert.

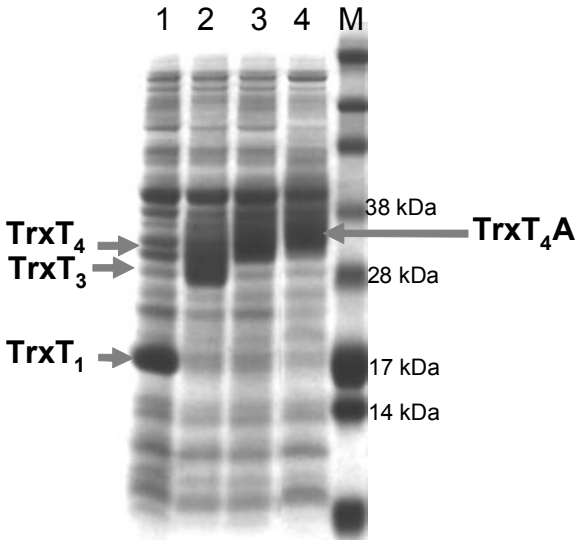
We defined the efficiency of each insertion step as the percentage of colonies obtained with both the correct insertion size and the correct orientation (out of >20 colonies analyzed for each step). For the insertion of DNA modules #2-4, we found that a 3:1 insert:vector molar ratio resulted in 20-30% efficiencies. This was equivalent or greater than the 20% efficiency we achieved in the TOPO® cloning step we used to insert DNA module #1.




| Lane | Vector | PCR Product Size (bp) |
|---|---|---|
| 1 | $pTrxT_1$ | 273 |
| 2 | $pTrxT_3$ | 429 |
| 3 | $pTrxT_4$ | 507 |
| 4 | $pTrxT_4A$ | 564 |

**Figure A.5 Insertion analysis using PCR.** Agarose electrophoresis gel showing colony PCR screening of the insert size for the expression vectors generated in this work.

After each ligation step, the DNA was also sequenced to confirm there were no recombination events or deletions, and the recombinant expression vector was transformed into a bacterial expression strain – One Shot® TOP10 Chemically Competent *E. coli.* The fusion protein product at each stage was soluble, and not partitioned into inclusion bodies. Thus after lysing the cells, the supernatant could be collected for analysis. Figure A.6 shows a coomassie blue-stained SDS PAGE of the products from the $pTrxT_1$, $pTrxT_3$, $pTrxT_4$, and $pTrxT_4A$ vectors. The products from these vectors are denoted as $TrxT_1$, $TrxT_3$, $TrxT_4$, and $TrxT_4A$, respectively. As we have shown previously[30], the synthetic polypeptides we are producing do not migrate in SDS PAGE as most natural proteins do. The PAA products in Figure A.6 are diffuse bands that migrate with an apparent higher molecular weight than natural proteins, which is likely due to reduced efficiency of SDS binding to the polypeptide backbone.




| Lane | Product | Apparent MW (kDa) | Expected MW (kDa) |
|---|---|---|---|
| 1 | $TrxT_1$ | ~17 | 16 |
| 2 | $TrxT_3$ | ~27 | 21 |
| 3 | TrxT4 | ~30 | 24 |
| 4 | $TrxT_4A$ | ~33 | 26 |

**Figure A.6 Coomassie-stained SDS PAGE of the polypeptide products (with the HP-Thioredoxin fusion tag) after each ligation step.** Arrows indicate each polypeptide product. M: prestained protein standards.

Protein expression of the $TrxT_4A$ product was carried out in shake flasks. After extraction of the cell lysis supernatant, immobilized metal affinity chromatography (IMAC) was used to purify the product, based on the affinity of the HP-Thioredoxin fusion tag affinity for chelated nickel ions. An SDS PAGE of the column products is shown in Figure A.7. Weakly bound proteins were washed off the column with a pH 6 buffer. A highly pure $TrxT_4A$ product was eluted off the column with a pH 4 buffer, shown in Lane 4. The PAA product migrates as a diffuse band at a higher apparent molecular weight.

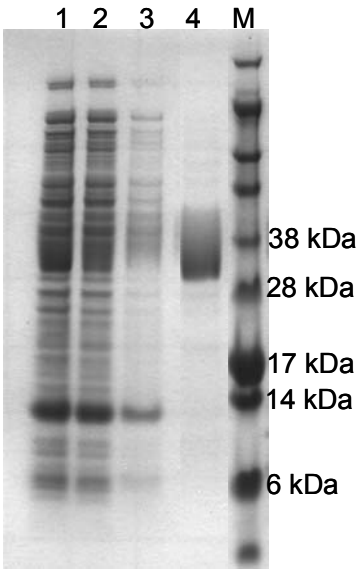



**Figure A.7 Coomassie blue-stained SDS PAGE of the products from IMAC purification of $TrxT_4A$.** Lane 1: Column feed - cell extract supernatant; Lane 2: Column flow through; Lane 3: pH 6 Wash; Lane 4: pH 4 Elution; M: Protein Standards

The purity and identity of the $TrxT_4A$ product was confirmed by high performance ion exchange chromatography (HPIEC) and MALDI-TOF mass spectrometry. The PAA species in the diffuse band in the SDS PAGE of Figure A.7 elutes as a sharp peak at 13 minutes from the HPIEC column (Figure A.8), and the measured m/z was 25,891.35 (Figure A.9), which is in good agreement with the calculated $(M+H)^+$ m/z of 25,889.91 ($MW_{avg}$).

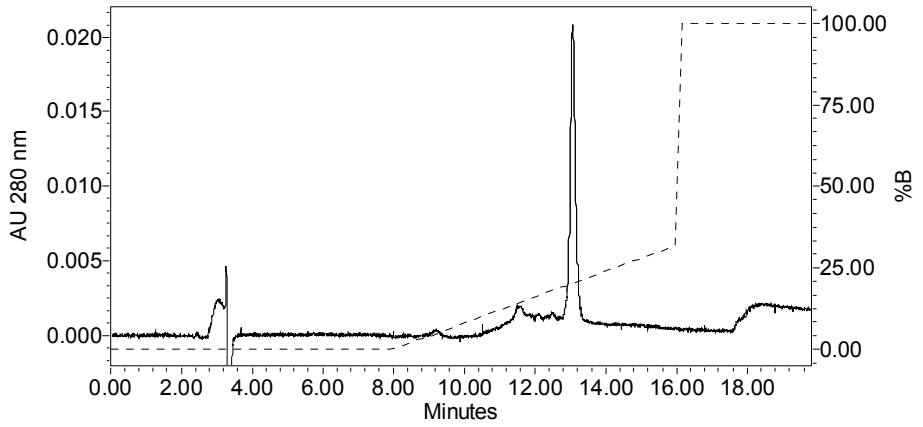



**Figure A.8 HPIEC analysis of the TrxT$_4$A product purified by IMAC.** Injection volume: 100 µL. Column: Vydac DEAE 301VHP with Vydac CPF10 prefilter, 7.5 mm ID x 50 mm L; 5 µm particle size, 900 Å pore size. Column temperature: 24°C. Mobile phase: A = 7.5 mM Tris-HCl in 25% acetonitrile, pH 8.0. B = 500 mM NaCl in Buffer A. Gradient program: 0-8 minutes: 0.5 ml/min, 0% B; 8-16 minutes: 1 ml/min, 0 to 32% B; 16-20 minutes: 100% B.

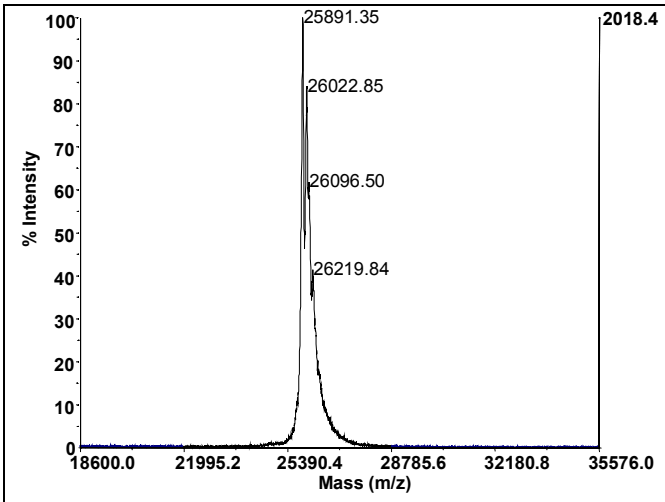



**Figure A.9 MALDI-TOF MS analysis of the TrxT$_4$A IMAC purification product.**

### *A.1.5* Conclusions

The objective of this work was to develop a method that can be used to produce high molecular weight PAA's especially suited to making brush forming polymers. The key difference between this method and those used to make other protein polymers is that in our method, the DNA construct can be sequentially modified as it exists in the expression vector. There is no need for intermediate cloning vectors, and a polypeptide product can produced and tested, and then modified if necessary. Thus if a modification

needs to be made, the DNA encoding the PAA does not need to be re-assembled from the beginning stages. There are several different ways that a PAA can be "grown" using this strategy, depending on the location of a unique *Sfo*I site in the first DNA module inserted, and how that *Sfo*I site is regenerated. In this work, we showed how a PAA can be grown by sequential additions to the C-terminus in a commercially available prokaryotic expression vector. Modules of DNA ranging from 48 bp to 156 bp were sequentially inserted, the size and orientation of the inserts were easily confirmed by colony PCR, and the PAA product was produced at each stage. Using this strategy, the only amino acid requirements were a Gly residue (-H side chain) at the C-terminus of each block, and an Ala residue (-$CH_3$ side chain) at the C-terminus of the finished PAA. These moderate amino acid requirements are fully compatible with our goals of making brush-forming polypeptides, as Gly and Ala residues have maximal steric flexibility in polypeptide structures, and do not have large bulky hydrophobic or ionic side chains. Gly and Ala residues are also common in the junctions between block of synthetic proteins that have been produced in the past.[17] We note that potential products from this method could include polymers with di-block or tri-block copolymer structure, or synthetic proteins such as elastin analogues, spider silk, and other biomimetic polypeptides. Additionally, yeast expression vectors used in producing proteins in Pichia pastoris are compatible with our method. This method is designed to use only commercially available reagents, and thus our goal is to lower the barrier to using biotechnology to produce new materials.

### *A.1.6* Acknowledgments

This work was supported by NSF Grant BES-0086876. We thank Dr. Ron Cerny of the University of Nebraska Center for Mass Spectrometry for the use of the Voyager DE Pro MALDI-TOF MS and technical assistance. We acknowledge conversations with the late Dr. Christopher G. Russell in the very beginning stages of this project.

## A.2 Manuscript: "*Supporting Information*"

The feasibility of variations in our method is shown in *Supporting Information*.

### *A.2.1* Addition of a *Bsu*36I site

A *Bsu*36I site was added to the 3' ends of the DNA module for the anchor blocks in this work (DNA Module #4). *Bsu*36I recognizes 5'-cc|tnagg-3', where *n* is any base. If *n=g*, then *Bsu*36I will recognize the sequence 5'-gcc-tga-ggc-3', which is not present in the pBAD/Thio-TOPO® vector and codes for Ala-Stop-Gly. Digestion of the vector with *Bsu*36I followed by treatment with mung bean nuclease (to make blunt ends) and subsequent insertion of new DNA modules by blunt-end ligation will result in in-frame additions to the C-terminus of the anchor blocks. This could be used to either extend the length of the anchor block domain, or to add new tail blocks to make a tri-block copolymer. This process could be repeated if necessary, and is illustrated in Figure A.10 below.

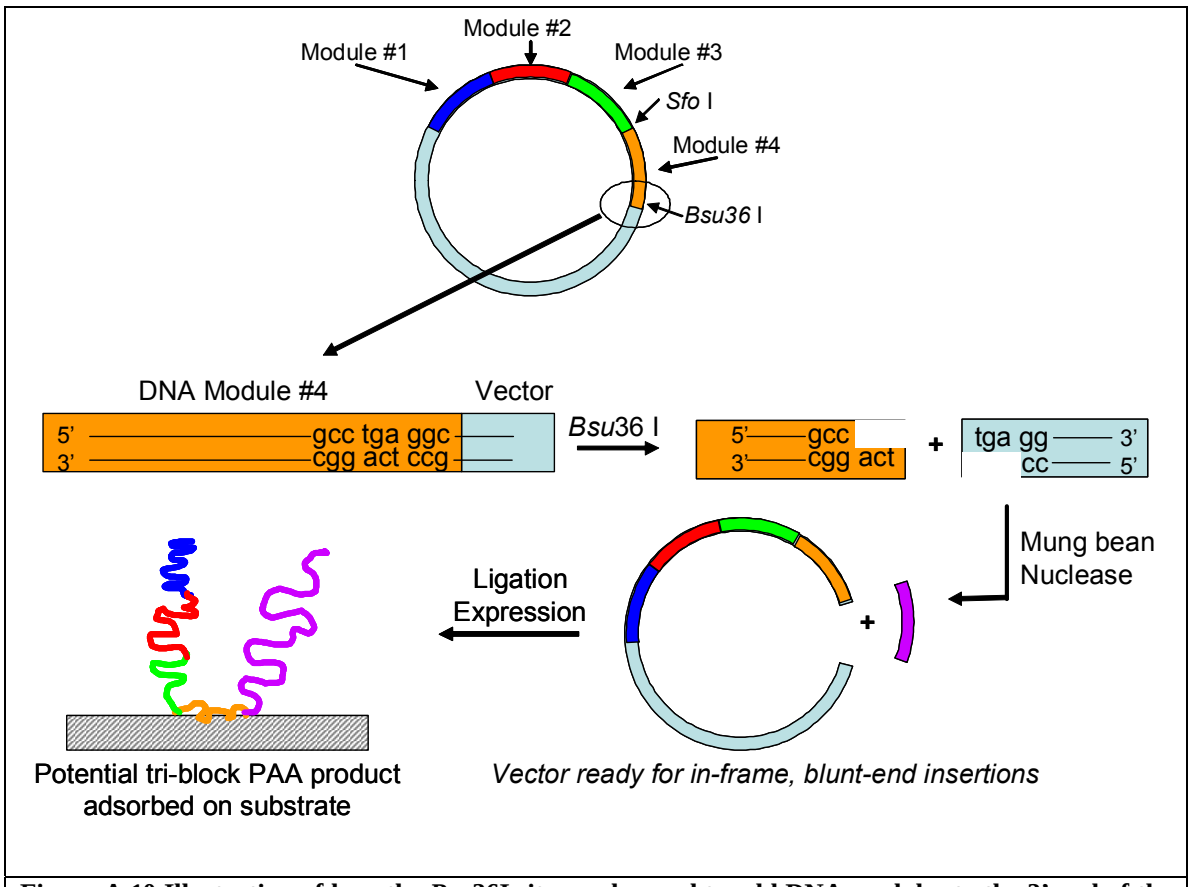



**Figure A.10 Illustration of how the *Bsu*36I site can be used to add DNA modules to the 3' end of the PAA-coding sequence (C-terminal anchor block in this example) to make a tri-block copolymer.**

### *A.2.2* Example Variation of the Strategy

In this section, we show that the variation of our method outlined in Figure A.1 of the main text (expanded below in Figure A.11) is also feasible for making genes that express high molecular weight PAA's. In this example, we used Invitrogen's pCR®T7/NT-TOPO® expression vector, which met our criteria of having no *Sfo*I site. This vector expresses recombinant proteins as fusions containing an N-terminal polyhistidine ($His_6$), an Xpress™ epitope, and an enterokinase cleavage site. The PAA sequences and composition were analogous to those we presented in the main text and are given here in Table A.4. The primary difference is that, as outlined below in Figure A.11, the C-terminal Anchor Block module was inserted first, followed the N-terminal Tail Block Module. All subsequent DNA modules were then inserted in between these two modules.

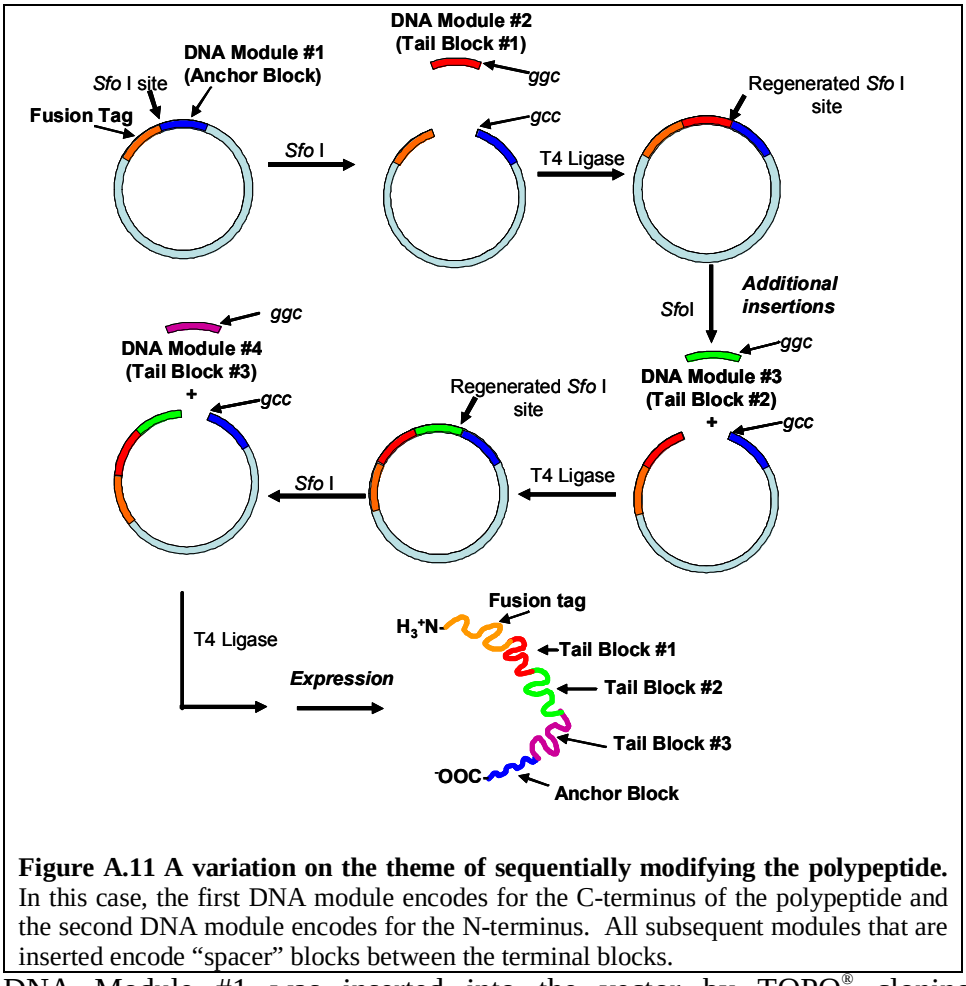


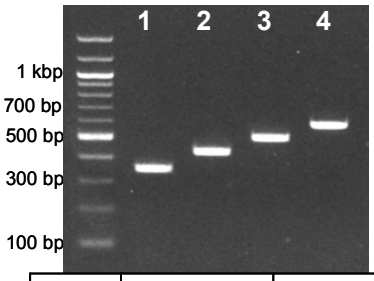

**Figure A.11 A variation on the theme of sequentially modifying the polypeptide.** In this case, the first DNA module encodes for the C-terminus of the polypeptide and the second DNA module encodes for the N-terminus. All subsequent modules that are inserted encode "spacer" blocks between the terminal blocks.

DNA Module #1 was inserted into the vector by TOPO® cloning. This polypeptide block was designed as an acidic anchor block consisting of a random sequence of glutamate and aspartate residues (Table A.4). The unique *Sfo*I recognition site (5'-*ggc*|*gcc*-3') was placed towards the 5' end. This allowed for subsequent insertions of DNA modules *upstream* of this module. The base pairs encoding for a hydroxylamine

chemical cleavage site (Asn-Gly) were included at the 5' end of the module to allow for future removal of N-terminal fusion tag after PAA expression. Finally, a stop codon was added at the 3' end of the module. The recombinant vector produced was designated $pHis_6$-A. Three different DNA modules, encoding for 3 different hydrophilic Tail Blocks, were used. The three tail DNA modules were inserted in the order of Tail #1, Tail #2, then Tail #3, and resulted in the recombinant vectors: $pHis_6$-TA, $pHis_6$-$T_2$A and $pHis_6$-$T_3$A.

**Table A.4 DNA sequences and the corresponding amino acid sequences.** Underlined base pairs in DNA Module #1 denote the *Sfo*I recognition site (ggc-gcc) and a *Bsu*36I site (cctgagg).

| | |
|---|---|
| **DNA Module #1 (75 bp)**<br>Anchor Block | 5'-aac ggc gcc ggc gga ggc ggt gat gac gat gag gat gaa gat gag gaa gat gac gag gac gac gag gat gcc tga ggc tga-3' |
| | NGAGGGGDDDEDEDEEDDEDDEDA*G* |
| **DNA Module #2 (78 bp)**<br>Tail Block # 1 | 5'-ggc aaa gag ggc gag aag ggg gag ggc cgt ggc cgc gag cgc ggc gat ggc gat cgt ggt aag gaa aaa gac ggt ggc-3' |
| | GKEGEKGEGRGRERGDGDRGKEKDGG |
| **DNA Module #3 (78 bp)**<br>Tail Block #2 | 5'-ggc gac cgt ggc aaa gac agc agc aag gat cgt gac agc tct gac cgt ggt gaa aaa agc gac aag gac aag ggc ggc-3' |
| | GDRGKDSSKDRDSSDRGEKSDKDKGG |
| **DNA Module #4 (78 bp)**<br>Tail Block #3 | 5'-ggc aag agc gaa ggc gag aag cgc agc gat aaa gag ggt aaa gag ggc ggt gac cgc aaa gag agc gat ggt cgc ggc-3' |
| | GKSEGEKRSDKEGKEGGDRKESDGRG |
| **Amino acid sequence of finished PAA**<br>(not including N-terminal fusion tag) | …NGGKEGEKGEGRGRERGDGDRGKEKDGGGDRGKDSSKDRDSSDRGEKSDKDKGGGKSEGEKRSDKEGKEGGDRKESDGRGAGGGGDDDEDEDEEDDEDDEDA*G* |




| Lane | Vector | PCR Product Size (bp) |
|---|---|---|
| 1 | $pHis_6A$ | 350 |
| 2 | $pHis_6T_1A$ | 428 |
| 3 | $pHis_6T_2A$ | 506 |
| 4 | $pHis_6T_3A$ | 584 |

**Figure A.12 Colony PCR results confirming the sequential insertion of the DNA modules given in Table A.4 into the pCR®T7/NT-TOPO® vector.**

The size of the insert was monitored by PCR with primers (T7 Forward: 5'-taatacgactcactataggg-3'; pRSET Reverse: 5'-tagttattgctcagcggtgg-3') that annealed to sequences that flanked the insertion site. Shown in Figure A.12 is an agarose gel of PCR products conducted after each sequential insertion of the four DNA modules. The orientation of the inserts was analyzed by both colony PCR and by double restriction digest with *Xba*I and *Sfo*I. The restriction enzyme *Xba*I cuts within the original vector sequence (in the $His_6$ coding region). The location of the *Sfo*I site depends on the orientation of the inserted module (Figure A.13). If inserted correctly, the restriction digest product size will be proportional to the size of the module inserted. If inserted incorrectly, the *Sfo*I site will be closer to the *Xba*I site, resulting in a smaller sized restriction fragment that does not include the module sequence. The combination of the PCR and restriction enzyme digest analyses was used to identify the clones that contained properly inserted DNA modules.

Our method is based on digestion with *Sfo*I, an enzyme that produces blunt ends, and insertion of blunt-end DNA modules into the linearized vector. Blunt-end ligation is non-directional, and less efficient than ligation of DNA with cohesive ends. We defined the efficiency of each insertion step as the percentage of colonies obtained with *both* the correct insertion size and the correct orientation (out of at least 20 colonies analyzed for each insertion step). For the insertion of DNA Modules #2-4, we found that a 3:1 insert:vector molar ratio resulted in 20-30% efficiencies (Table A.5). This was equivalent or greater than the 20% efficiency we achieved in the TOPO®-TA cloning step we used to insert DNA Module #1. DNA sequencing confirmed the fidelity of the cloning process.

In summary, using this variation of the cloning strategy, we were able to successively make four sequential insertions directly into an unmodified, commercially

available expression vector. DNA sequencing confirmed the construction of a PAA gene enoding for the composition: $NH_3^+$-(Fusion Tag)-(Tail Block #1)-(Tail Block #2)-(Tail Block #3)-(Anchor Block)-$COO^-$.

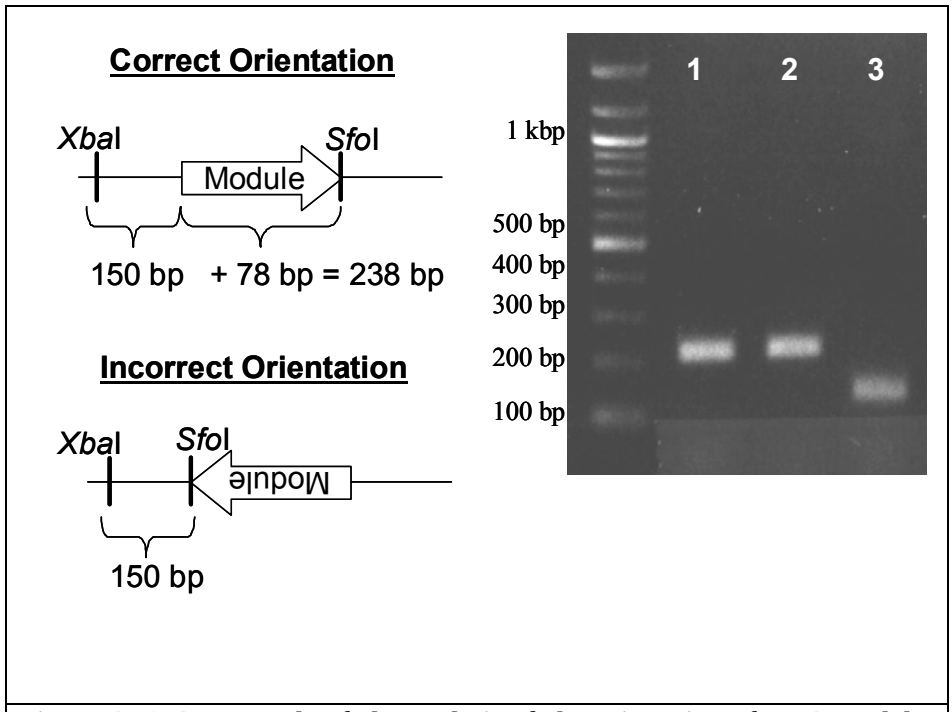



**Figure A.13 An example of the analysis of the orientation of DNA modules inserted into the pCR®T7/NT-TOPO® vector by restriction digest with *Xba*I and *Sfo*I.** In this case the transformed colonies containing the expression vector with the correctly oriented insert resulted in a digest product of 238 bp (Lanes 1&2); a reverse-oriented insert resulted in a digest product of 150 bp (Lane 3).

**Table A.5 Efficiencies of the sequential insertion steps made during the feasibility experiments using TOPO® cloning and blunt-end ligation.** Efficiency calculated by the number of isolated colonies having the correct insert size and orientation divided by the number of transformant colonies analyzed (>20).

| Step | Insertion Method | Efficiency |
|---|---|---|
| Vector + DNA Module #1 | TOPO® cloning | 20% |
| + DNA Module #2 | Blunt-end ligation | 25% |
| + DNA Module #3 | Blunt-end ligation | 30% |
| + DNA Module #4 | Blunt-end ligation | 20% |

## A.3 References


(1) Wittemann, A.; Haupt, B.; Ballauff, M. Adsorption of proteins on spherical polyelectrolyte brushes in aqueous solution. *Phys. Chem. Chem. Phys.* **2003**; *5* (8), 1671-1677.

(2) Kawai, T.; Saito, K.; Lee, W. Protein binding to polymer brush, based on ion-exchange, hydrophobic, and affinity interactions. *J. Chromatogr. B.* **2003**; *790* (1-2), 131-142.

(3) Saito, K. Charged polymer brush grafted onto porous hollow-fiber membrane improves separation and reaction in biotechnology. *Separ. Sci. Technol.* **2002**; *37* (3), 535-554.

(4) Ito, H.; Nakamura, M.; Saito, K.; Sugita, K.; Sugo, T. Comparison of L-tryptophan binding capacity of BSA captured by a polymer brush with that of BSA adsorbed onto a gel network. *J. Chromatogr. A.* **2001**; *925* (1-2), 41-47.

(5) Bures, P.; Huany, Y.B.; Oral, E.; Peppas, N.A. Surface modifications and molecular imprinting of polymer in medical and pharmaceutical applications. *J. Control Release.* **2001**; *72* (1-3), 25-33.

(6) Spangler, B.D.; Tyler, B.J. Capture agents for a quartz crystal microbalance-continuous flow biosensor: functionalized self-assembled monolayers on gold. *Anal. Chim. Acta.* **1999**; *399* (1-2), 51-62.

(7) Lee, K.B.; Park, S.J.; Mirkin, C.A.; Smith, J.C.; Mrksich, M. Protein nanoarrays generated by dip-pen nanolithography. *Science.* **2002**; *295*, 1702-1705.

(8) Kusnezow, W.; Hoheisel, J.D.; Solid support for microarray immunoassays. *J Mol. Recognit.* **2003**; *16* (4), 165-176.

(9) Otsuka, H.; Nagasaki, Y.; Kataoka, K. Self-assembly of poly(ethylene glycol)-based block copolymers for biomedical applications. *Curr. Opin. Colloid In.* **2001**; *6* (1), 3-10.

(10) Rovira-Bru, M.; Giralt, F.; Cohen, Y. Protein adsorption onto zirconia modified with terminally grafted polyvinylpyrrolidone. *J. Colloid Interf. Sci.* **2001**; *235* (1), 70-79.

(11) Lieb, E.; Tessmar, J.; Hacker, M.; Fischbach, C.; Rose, D.; Blunk, T.; Mikos, A.G.; Gopferich, A.; Schulz, M.B. Poly(D,L-lactic acid)-poly(ethylene glycol)-monomethyl ether diblock copolymers control adhesion and osteoblastic differentiation of marrow stromal cells. *Tissue Eng.* **2003**; *9* (1), 71-84.

(12) Huang, K.; Lee, B.P.; Ingram, D.R.; Messersmith P.B. Synthesis and characterization of self-assembling block copolymers containing bioadhesive end groups. *Biomacromolecules.* **2002**; *3* (2), 397-406.

(13) Pinaud, F.; King, D.; Moore, H-P.; Weiss, S. Bioactivation and cell targeting of semiconductor CdSe/ZnS nanocrystals with phytochelatin-related peptides. *J. Am. Chem. Soc.* **2004**; *126*, 6115-6123.

(14) Winblade, N.D.; Schmokel, H.; Baumann, M.; Hoffman, A.S.; Hubbell, J.A. Sterically blocking adhesion of cells to biological surfaces with a surface-active copolymer containing poly(ethylene glycol) and phenylboronic acid. *J. Biomed. Mater. Res.* **2002**; *59* (4): 618-631.

(15) Orth, J.; Meyer, W.H.; Bellman, C.; Wegner, G. Stabilization of aqueous alpha-$Al_2O_3$ suspensions with block copolymers. *Acta Polym.* **1997**; *48* (11), 490-501.

(16) Creutz, S.; Jerome, R. Effectiveness of block copolymers as stabilizers for aqueous titanium dioxide dispersions of a high solid content. *Prog. Org. Coat*. **2000**; *40,* 21-29.

(17) van Hest, J.C.M.; Tirrell, D.A. Protein-based materials, toward a new level of structural control. *Chem. Comm.* **2001**; 1897-1904.

(18) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. Genetic engineering of structural protein polymers. *Biotechnol. Prog.* **1990**; *6,* 198-202.

(19) McMillan, R.A.; Lee, T.A.T.; Conticello, V.P. Rapid assembly of synthetic genes encoding protein polymers. *Macromolecules* **1999**; *32,* 3643-3648.

(20) Meyer, D.E. 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,* 357-367.

(21) Won, J-I.; Barron, A.E. A new cloning method for the preparation of long repetitive polypeptides without a sequence requirement. *Macromolecules*. **2002**; *35,* 8281-8287.

(22) Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. **1992**.

(23) Ausubel, F.M. *Current Protocols in Molecular Biology*. J. Wiley: New York, N.Y. **1997**.

(24) Gill, S.C.; Von Hippel, P.H. Calculation of protein extinction coefficient from amino acid sequence data. *Anal. Biochem.* **1989**; *182* (2), 319-326.

(25) McGrath, K.P.; Tirrell, D.A.; Kawai, M.; Mason, T.L.; Fournier, M.J. Chemical and biosynthetic approaches to the production of novel polypeptide materials. *Biotechnol. Prog.* **1990**; *6,* 188-192.

(26) Barron, A.E.; Zuckermann, R.N. Bioinspired polymeric materials: in-between proteins and plastics. *Curr. Opin. Chem. Biol*. **1990**; *3,* 681-687.

(27) Miyada, C.G.; Stoltzfus, L.; Wilcox, G. Regulation of the *ara*C gene of *Escherichia coli*: catabolite repression, autoregulation and effect on *ara*BAD expression, *Proc. Natl. Acad. Sci. U.S.A.* **1984**; *81,* 4120-4124.

(28) Lee, N.; Francklyn, C.; Hamilton, E.P. Arabinose-induced binding of AraC Protein to *ara*$I_2$ activates the *ara*BAD operon promoter. *Proc. Natl. Acad. Sci. U.S.A.* **1987**; *84,* 8814-8818.

(29) Lavallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio-Technology*. **1993**; *11,* 187-193.

(30) Tulpar, A.; Henderson, D.B.; Mao, M.; Caba, B.; Davis, R.M.; Van Cott, K.E.; Ducker, W.A. Unnatural proteins for the control of surface forces. *Langmuir*; ***In Press***

(31) Canova-Davis, E.; Eng, M.; Mukku, V.; Reifsnyder, D.H.; Olson, C.V.; Ling, V.T. Chemical heterogeneity as a result of hydroxylamine cleavage of a fusion protein of human insulin-like growth factor I. *Biochem. J.* **1992**; *278*: 207-213.