Completely Monodisperse, Highly Repetitive Proteins for Bioconjugate Capillary Electrophoresis: Development and Characterization

Jennifer S. Lin,[†] Jennifer Coyne Albrecht,[‡] Robert J. Meagher,[§] Xiaoxiao Wang,[†] and Annelise E. Barron*,[†]

[†]Department of Bioengineering and [‡]Department of Chemical Engineering, Stanford University, Stanford, California, United States
[§]Sandia National Laboratories, P.O. Box 969, MS 9292, Livermore, California, United States

S *Supporting Information*

ABSTRACT: Protein-based polymers are increasingly being used in biomaterial applications because of their ease of customization and potential monodispersity. These advantages make protein polymers excellent candidates for bioanalytical applications. Here we describe improved methods for producing drag-tags for free-solution conjugate electrophoresis (FSCE). FSCE utilizes a pure, monodisperse recombinant protein, tethered end-on to a ssDNA molecule, to enable DNA size separation in aqueous buffer. FSCE also provides a highly sensitive method to evaluate the polydispersity of a protein drag-tag and thus its suitability for bioanalytical uses. This method is

able to detect slight differences in drag-tag charge or mass. We have devised an improved cloning, expression, and purification strategy that enables us to generate, for the first time, a truly monodisperse 20 kDa protein polymer and a nearly monodisperse 38 kDa protein. These newly produced proteins can be used as drag-tags to enable longer read DNA sequencing by free-solution microchannel electrophoresis.
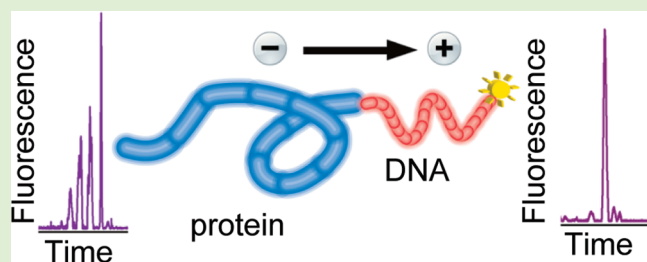
■ INTRODUCTION

Free-solution conjugate electrophoresis (FSCE), or alternatively "end-labeled free-solution electrophoresis" (ELFSE), uses a pure, monodisperse "drag-tag" tethered end-on to a DNA molecule to enable DNA size separation by free-solution electrophoresis.[1−3] FSCE enables a novel method of DNA sequencing and genotyping that is ideal for implementation on microfluidic devices because it obviates the need for viscous polymer solutions to separate DNA. Elimination of the viscous polymer solution will save time, reduce costs, and avoid challenges associated with loading and replacing the polymer matrix for electrophoresis in microfabricated devices ("microchips"). An aqueous buffer could simply be loaded into a microchip for the analysis, an aspect that will facilitate automation. Free-solution electrophoretic separation of DNA can also be easily integrated into "lab-on-a-chip" devices.

In FSCE, a monodisperse perturbing entity, which has a different charge-to-friction ratio than DNA, is attached to DNA to break the symmetry between charge and friction that prevents DNA separation in free solution. The presence of this "drag-tag" alters the electrophoretic mobility of each DNA molecule in a size-dependent manner, allowing the separation of these bioconjugates to occur in free solution. DNA sequencing can be performed by separating Sanger sequencing fragments[4] in free-solution electrophoresis with single base resolution. The larger the hydrodynamic drag provided by the drag-tag (i.e., the larger the size of the drag-tag), the greater the size of the sequencing fragments that can be resolved and, consequently, the longer the read length obtainable by FSCE sequencing.

The ideal drag-tag is completely monodisperse, water-soluble, uncharged or nearly so, has minimal adsorption to or nonspecific interaction with microchannel walls, and can be uniquely and stably attached to DNA, as previously described.[5] One of the most important properties for a drag tag is complete monodispersity where every tag is identical in charge and drag. If a polydisperse molecule is used as a drag-tag for FSCE, then the resulting peak pattern can be ambiguous, where each DNA length is represented by multiple peaks in the electropherogram instead of a single peak. These peaks would also overlap with multiple peaks corresponding to other DNA sizes, making accurate DNA sizing nearly impossible. This essential requirement for total monodispersity eliminates all commonly available synthetic polymers and microparticles as useful drag-tag candidates for DNA sequencing.[6] Although solid-phase synthesis techniques can be used to generate monodisperse molecules such as polypeptoids (poly-N substituted glycines) in a controlled manner, these molecules are too small to generate sufficient hydrodynamic drag to separate large DNA fragments for FSCE sequencing.[7] Natural proteins can be much larger in size than chemically synthesized molecules but have several drawbacks of their own that also make them nonideal drag-tag candidates. In aqueous solution, most natural proteins are folded into compact shapes and typically present numerous surface

```
Gene 1:
                         ↓(Gly Ala Gly Thr Gly Ser Ala Gly Ala Gly Thr Gly Ser
ATA TAG AAT TCC TCT TCA GGT GCG GGC ACC GGA AGT GCA GGT GCT GGA ACG GGC AGC

Ala Gly Ala Gly Thr Gly Ser Ala)
GCA GGA GCT GGC ACC GGT TCC GCG GGT AGA AGA GGA ATT CAT ATA
                                  ↑
```
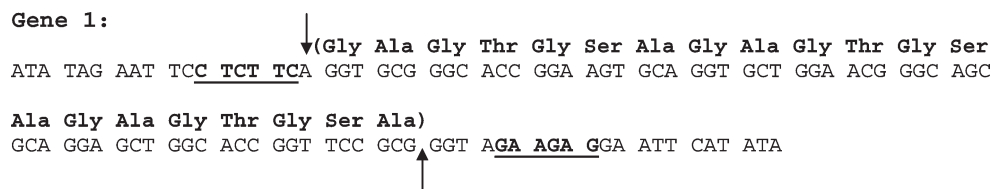
Figure 1. DNA sequence of the macromonomer used initially to generate the repetitive genes. *Ear*I restriction sites are in bold and underlined, and arrows indicate the cleavage site of the enzyme.

charges that are likely to have local interactions with the DNA or microchannel walls. Additionally, natural proteins typically contain multiple reactive groups (e.g., amino, thiol, carboxylic acid) on their surface, making unique attachment to DNA difficult. In contrast, protein polymers can meet the many stringent requirements of a drag-tag through careful design of the repetitive sequence to reduce or eliminate the number of potentially problematic charged and reactive sites. These highly repetitive proteins can be produced in biological systems to a higher degree of control than synthetic materials and are often touted as being truly monodisperse.[8−10] For most applications, this claim is true in comparison with chemically synthesized polymers, which are inherently highly polydisperse. However, for FSCE, a completely monodisperse molecule is essential.

FSCE itself can be used as a highly sensitive, fluorescence-based detection method to investigate the polydispersity of a given protein polymer drag-tag. A drag-tag is conjugated to a monodisperse, fluorescently labeled DNA primer of known length, and the bioconjugate is then analyzed in free-solution electrophoresis. In an ideal case, only two peaks are present in the electropherogram: a free (unconjugated) DNA peak eluting first, followed by a peak corresponding to the drag-tag and DNA conjugate, where the DNA has been slowed down by the attachment of the drag-tag. This strategy has been used to characterize a synthetic (low-polydispersity) poly(ethylene glycol) (PEG) sample,[11] solid-phase synthesis products,[12] and the deamidation of protein polymers containing glutamine residues.[13]

Obtaining completely monodisperse protein polymers that are also suitable as drag-tags for FSCE-based DNA sequencing has been a challenging task. Previous drag-tag designs of various lengths and designs were unexpectedly heterogeneous when assessed by FSCE despite being produced in *E. coli*.[6,13,14] Recently, a small, random-coil, sufficiently monodisperse protein polymer of 127 amino acids was produced and used for DNA sequencing in free solution with an obtainable read length of ∼180 bases.[5] This result demonstrated that protein polymers can be used successfully as drag-tags for sequencing. However, to obtain even longer read lengths (>400 bases), a larger, yet still completely monodisperse, drag-tag needs to be produced.

Here we detail the synthesis and characterization of large (>250 amino acids), monodisperse protein polymers that will be used as drag-tags to enable longer read length FSCE sequencing than has been previously reported. New strategies had to be developed to produce protein polymers that are more than double the size of the original 127-amino acid drag-tag yet are still monodisperse when analyzed by the highly sensitive FSCE method. These new protein drag-tags have the potential to achieve single-base resolution sequencing of 400 bases by FSCE, which is on par with read lengths obtained using next generation technology. These results would facilitate further development of

FSCE as a rapid, selective, and highly accurate sequencing method, characteristics ideal for use in medical diagnostics such as matching organs for transplant,[15] where sequencing a whole genome is not needed. Until now, development of FSCE for sequencing applications has been hindered by the lack of a large, monodisperse drag-tag for obtaining long reads. The repetitive nature of the protein polymer amino acid sequences, conditions used to isolate and purify the material, as well as the method of attachment to DNA are all factors to be considered when producing a completely monodisperse product that is suitable as a drag-tag for DNA sequencing or genotyping using FSCE.

■ MATERIALS AND METHODS

All molecular biology protocols were conducted according to standard protocols or from instructions provided by manufacturers unless otherwise noted. Unless specifically stated, enzymes were all obtained from New England Biolabs (Ipswich, MA). General reagents for cloning, protein expression, and purification were obtained from Fisher Scientific (Pittsburgh, PA) unless otherwise noted.

Creation of Multimer Gene. A 102-bp single-stranded synthetic oligonucleotide was designed to consist of three repeats of the seven amino acid sequence Gly-Ala-Gly-Thr-Gly-Ser-Ala. The gene sequence is shown in Figure 1. The oligonucleotide was purchased from Integrated DNA Technologies (Coralville, IA) and was polymerase chain reaction (PCR)-amplified using high fidelity *Pfu* DNA polymerase (Stratagene, La Jolla, CA). The PCR product was then digested at 37 °C by *Ear*I. The fully cleaved 63-bp fragment was isolated and purified from undigested products by agarose gel electrophoresis and the QIAEX II gel extraction kit (Qiagen, Valencia, CA). Multimers of the gene were generated by self-ligation using T4 DNA ligase. These multimers were inserted into a modified pUC18 cloning vector containing flanking *Sap*I sites in accordance with the controlled cloning method, a novel strategy developed in the Barron group for generating larger genes from multimers without sequence requirements based on the use of two Type IIS endonucleases, *Sap*I and *Ear*I.[16] These restriction enzymes cut downstream of their recognition sites. Note that circularization of sufficiently long DNA multimers limits the size of the multimer gene that can be obtained simply by self-ligation. The plasmids were transformed via heat shock into NovaBlue competent cells (Novagen, Madison, WI). The resulting transformants were screened by DNA sequencing to verify the identity and size of the insert DNA. Sequencing showed that the selected multimer gene, consisting of 18 repeats of the 7 amino acid sequence, had two serine-to-arginine mutations and that the actual sequence is (GAGTGSA)$_4$GAGTGRA(GAGTGSA)$_7$GAG-TGRA(GAGTGSA)$_5$. This sequence is designed to be random coil in structure,[17] providing greater hydrodynamic drag than a similarly sized but more compact, globular protein. This gene was used to produce a protein that demonstrated the feasibility of FSCE sequencing using a protein polymer drag-tag.[5] The gene encoding 18 repeats of the 7 amino acid sequence (with mutations) was doubled twice by controlled cloning to produce genes 36 and 72 repeats long. Note that *Ear*I is an analog of

*Eam*1104 I and either enzyme can be used in conjunction with *Sap*I for this cloning strategy.

Generation of Expression Vector with C-Terminal Affinity Tag. Site-directed mutagenesis (QuikChange Kit, Stratagene, La Jolla, CA) was used to alter the two existing *Sap*I sites of pET-41a (Novagen) into *Ear*I recognition sites. Primer sequences 5′-CTT GAA GAA AAA TAT GAG GAG CAT TTG TAT GAG CGC GAT G-3′ and 5′-GAG GAA GCG GAA GAG AGC CTG ATG CGG-3′ along with their respective reverse complementary sequences (four primers total) were designed according to the manufacturer's guidelines and purchased as PAGE-purified DNA oligomers from IDT. Two rounds of mutagenesis were performed according to the suggested manufacturer's protocol. *Sap*I digestion of the recovered plasmid DNA confirmed that the modifications were successful based on the observation of intact vector on an agarose gel.

Assembly PCR was used to generate a 179-bp oligonucleotide containing a T7 tag (MASMTGGQQMG) for enhanced expression and an octahistidine tag for affinity purification to be inserted into the multiple cloning site of the expression plasmid. (See the Supporting Information.) Six synthetic oligonucleotides were designed with overlapping bases and similar melting temperatures (55 °C) according to the outlined protocol for assembly PCR[18] along with flanking primers. (See the Supporting Information.) The oligonucleotides were purchased from IDT and resuspended at a concentration of 12.5 μg/μL in water. For the first thermal cycling step, 2 μL of each oligonucleotide was combined with 0.25 μL of Go*Taq* polymerase (Promega, Madison, WI), 0.4 μL of 25 mM dNTP, 10 μL of 5X Go*Taq* buffer, and water for a 50 μL reaction. A 7 min initial denaturing step at 94 °C was followed by 2 min of annealing at 54 °C and 3 min at 72 °C. Seven amplification cycles were then carried out with 1.5 min at 94 °C, 2 min at 54 °C, and 3 min at 72 °C, followed by a final extension step at 72 °C for 5 min. This reaction is followed by a standard PCR amplification using the flanking primers. The primers were resuspended in water at 0.25 μg/μL. We combined 1 μL from the first reaction with 0.5 μL of Go*Taq*, 0.8 μL of 25 mM dNTP, 20 μL 5X Go*Taq* buffer, 4 μL of each primer, and water for a 100 μL volume reaction. After an initial 5 min denaturing step at 94 °C, 25 cycles of amplification were carried out for 30 s at 94 °C, 2 min at 54 °C, and 1.5 min at 72 °C, followed by a final 5 min extension step at 72 °C. The desired product band was isolated and purified by agarose gel electrophoresis.

The existing cloning region of the modified pET-41a plasmid was excised via double digestion using *Xba*I and *Xho*I enzymes. This region was replaced with the 179-bp oligonucleotide discussed above that was similarly digested with *Xba*I and *Xho*I to generate cohesive ends for ligation. This modified vector is designated MpET-41a. A fusion protein expressed in this vector would have a T7 tag at the N-terminus and an octahistidine tag at the C-terminus. The recipient vector was prepared by digesting the circular plasmid with *Sap*I at 37 °C for 16 h. This was followed by digestion with *Nde*I enzyme for 1 h to linearize any undigested plasmid and then slab gel purification to isolate the desired vector band. Finally, the vector was reacted with calf intestinal phosphatase (CIP) for 1 h to minimize recircularization of the plasmid in subsequent ligation steps.

Protein Expression and Purification. Desired multimer genes were excised from the pUC18 cloning vector via *Sap*I digestion and were ligated into either the modified N-terminal decahistidine tag expression vector MpET-19b[16] (Novagen) or the MpET-41a recipient vector described above. Sequencing confirmed the identity of the resulting plasmid DNA before transferring the DNA into *E. coli* BLR(DE3) expression cells (Novagen) via heat shock. Protein expression was performed using Terrific Broth (EMD Biosciences, San Diego, CA) media at 37 °C under tetracycline (12.5 μg/mL) and either carbenicillin (50 μg/mL) or kanamycin (30 μg/mL) antibiotic selection for the MpET-19b and the MpET-41a vectors, respectively. We inoculated 1 L

cultures with 25 mL of a culture grown from a single colony in LB media overnight. After the cells reached OD$_{600}$ = 0.6 to 0.8, isopropyl-β-D-thiogalactoside (IPTG, U.S. Biologicals, Swampscott, MA) was added at a final concentration of 0.5 mM to induce protein synthesis. Cells were harvested by centrifugation at 6000g at 4 °C for 12 min after 3 h of additional growth time. The cell pellet was resuspended in denaturing buffer (8 M urea, 50 mM sodium phosphate, 300 mM NaCl at pH 7.8) and frozen overnight at −80 °C. Cells were then lysed by ultrasonication for 2 min. The resulting cell lysate was centrifuged at 20 000 g at 4 °C for 45 min to pellet the cell debris. The clarified lysate was bound to Talon cobalt-chelated resin (Clontech, Mountain View, CA) for 1 h at room temperature prior to column purification. The resin was washed twice with 10 column volumes of the above-mentioned denaturing buffer. The target protein was eluted using buffer containing an additional 150 mM imidazole (three column volumes). Cell lysate, flow through, washes, and elutions were all analyzed on a discontinuous 12% SDS-PAGE gel. All gels were visualized with Coomassie staining. Negative zinc staining did not show better results than Coomassie. Elutions containing the target protein were combined and dialyzed 3 days against deionized water at 4 °C. Finally, the protein was lyophilized into a dry powder. When needed, the proteins were further purified using preparative reversed-phase high-performance liquid chromatography (HPLC) on a Phenomenex Jupiter C18 column (10 μm, 300 Å, 21.2 × 250 mm) using a linear gradient of 5−95% solvent B in solvent A over 35 min at a flow rate of 15 mL/min. Solvent A is 0.1% trifluoroacetic acid (TFA) in water (v/v), and solvent B is 0.1% TFA in acetonitrile (v/v). Collected fractions were lyophilized to a dry powder, resuspended in water, pH adjusted to near neutral, and then lyophilized again.

Removal of N-Terminal Histidine Affinity Tag from the Expressed Fusion Protein. For proteins expressed with the N-terminal His tag, removal of the affinity tag can be accomplished through chemical cleavage at the N-terminal methionine residue (assuming no Met residues in the repetitive sequence) using cyanogen bromide in 70% formic acid for 24−48 h.[19] Proteins were dissolved in the reaction mixture at a final concentration of ∼1 mg/mL. Cyanogen bromide was added at ∼5 mg/mg protein. After nitrogen purging, the entire mixture was covered with aluminum foil and gently mixed for several hours. A rotary evaporator was then used to remove volatiles and dry the solution under vacuum. The product was resuspended in water and lyophilized. A second column chromatography purification step with Talon resin was performed to separate successfully cleaved protein from protein still containing the His tag.

Assay of Reaction Conditions for Enzymatic Removal of C-Terminal Affinity Tag. The C-terminal His tag includes an IEGR recognition site for site-specific cleavage by Factor Xa (Novagen). Factor-Xa-to-target-protein ratios (unit/μg) of 1:100, 1:50, and 1:20 were tested. We digested 10 μg of protein by varying amounts of enzyme (0, 0.1, 0.2, 0.5 units) in a 50 μL reaction at 20 °C. We took 10 μL of sample at 2, 4, 8, and 16 h time intervals and immediately mixed with 10 μL of SDS-containing sample buffer for future SDS-PAGE analysis and to halt the cleavage reaction. We digested 2 μg of the control protein with 0.1 units of enzyme for 16 h. Test cleavage results were all analyzed by Western blot using a penta-His antibody (Qiagen) and antimouse IgG horseradish peroxidase (HRP) antibody (GE Healthcare, Piscataway, NJ). Test digestions using endoproteinase GluC (New England Biolabs) as the protease were carried out at 25 °C in the provided reaction buffer using the protease-to-target-protein (μg/μg) ratios of 1:100, 1:50, and 1:20. These reactions were similarly monitored over the course of 16 h, and the time points were analyzed by Western blot using the penta-His antibody.

General Protein Analysis and Characterization. Purified protein dissolved in water at 1 mg/mL was analyzed by reversed-phase HPLC on a Phenomenex Jupiter C18 column (5 μm, 300 Å, 2 × 250 mm) at a gradient of 5−95% acetonitrile to water with 0.1% TFA. Peaks were

detected at 220 nm. A Voyager DE-PRO mass spectrometer (Analytical Services Laboratory, Northwestern University and Protein and Nucleic Acid Facility, Stanford University) was used for MALDI-TOF analysis of the protein using sinapinic acid as the matrix. Amino acid compositional analysis was performed by the W. M. Keck Facility at Yale University (New Haven, CT). Circular dichroism (CD) spectroscopy was conducted using a J-715 Jasco (Easton, MD) spectrophotometer (Keck Biophysics Facility, Northwestern University). Data were collected between 185 and 280 nm using a 0.02 cm path length cuvette.

Protein Analysis and Characterization Using Free-Solution Conjugate Electrophoresis. Protein polymers were further characterized by FSCE to determine the actual protein purity.[14] First, two oligonucleotides containing a thiol (−SH) functionality on the 5′ terminus were purchased from IDT: a 23-base oligonucleotide (SH-GCA T*GT ATC TAT CAT CCA TCT CT) and a 30-base oligonucleotide (SH-CCT* TTT AGG GTT TTC CCA GTC ACG ACG TTG) were used (where T* indicates the dT-fluorescein). To reduce the DNA, 2 nmol of DNA primer was incubated with a 20:1 molar excess of Tris(2-carboxyethyl)phosphine (TCEP, Pierce Biotechnology, Rockford, IL) at 40 °C for 90 min in 20 μL of 70 mM sodium phosphate buffer, pH 7.2.[5] Protein polymers were activated at the N-terminus with the heterobifunctional cross-linker sulfosuccinimidyl 4(N-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC, Pierce). Sulfo-SMCC was chosen as the cross-linker because of its high reaction specificity for creating a stable covalent bond between the amino terminus of the protein and the thiolated DNA and its high molecular stability, particularly its robustness to rapid temperature cycling up to 96 °C as part of the Sanger sequencing reaction.[6,14,20] Additionally, sulfo-SMCC is water-soluble and thus compatible with all of the steps of drag-tag conjugation and FSCE DNA sequencing.[20] A 10:1 molar excess of sulfo-SMCC was added to 1.2 mg of protein polymer in 80 μL of 100 mM sodium phosphate buffer, pH 7.2, and the mixture was vortexed for 1 h at room temperature. Excess sulfo-SMCC was separated from the activated protein polymer drag-tag by gel filtration with a Centri-Sep column (Princeton Separations, Adelphia, NJ). The activated, purified protein polymer was frozen, lyophilized, resuspended in water at 10 mg/mL concentration, and stored at −80 °C until used.[5] To conjugate the activated protein polymer to the reduced DNA, 90 pmol of DNA was mixed with 2.5 nmol of drag-tag to a final volume and concentration of 10 μL in 25 mM sodium phosphate buffer at pH 7.2. The mixture was then incubated at room temperature for 3−24 h. A large excess of protein to DNA (typically 100-fold) is necessary to ensure nearly complete (>95%) conjugation of drag-tags to each DNA molecule.[5,21,22]

For a sequencing sample, the protein drag-tag was instead conjugated to a thiol-containing M13 sequencing primer (SH-GTT TTC CCA GTC ACG AC from IDT). We mixed 8 μL of BigDye terminator v1.1 cycle sequencing mix (Applied Biosystems, Foster City, CA), 0.16 μg of M13mp18 single-stranded DNA (ssDNA) template, and combined them with 8.4 pmol of sequencing primer conjugated to drag-tag to a total volume of 20 μL. After incubating at 96 °C for 1 min, the sequencing reaction was cycled 36 times (96 °C for 10 s, 50 °C for 5 s, and 60 °C for 30 s to 2 min on an Eppendorf Mastercycler gradient instrument). The sample was purified via Centri-Sep column, denatured at 95 °C for 2 min, and then snap-cooled on ice for 5−10 min prior to analysis.

An ABI 3100 genetic analyzer (ABI) with a 16-capillary array of fused silica capillaries (50 μM inner diameter) and four-color laser-induced fluorescence (LIF) detection using a 488 nm laser was used to analyze the protein polymer−DNA conjugates and sequencing reactions in free solution. Capillary electrophoresis separations were done in denaturing buffer consisting of 89 mM Tris(hydroxymethyl)aminomethane (Tris), 89 mM Tris(hydroxymethyl) methylaminopropanesulfonic acid (TAPS), 2 mM ethylenediaminetetraacetic acid (EDTA), and 7 M urea. A 0.5−3% (v/v) POP-5 ("performance optimized polymer") or POP-6

polymer solution (ABI) was added to the denaturing buffer as a dynamic wall coating agent to suppress electroosmotic flow and prevent adsorption to capillary walls. Capillaries with an effective length from inlet to detector of 36 cm were used for FSCE separations (total length 47 cm). Typical electrophoresis conditions include electrokinetic injection with a potential of 1 to 2 kV applied for 5−30 s and running voltage of 14.7 kV, all at 55 °C.[21−23]

RESULTS AND DISCUSSION

We have developed a novel method of producing completely monodisperse repetitive polypeptides ("protein polymers") whose purity was assessed using a highly sensitive method called FSCE. A short 127-amino acid protein polymer drag-tag, whose repetitive sequence was designed de novo, was previously used successfully to enable DNA sequencing by FSCE with an obtainable read length of ~180 bases.[5] Challenges remained in obtaining a drag-tag with greater hydrodynamic drag yet complete monodispersity to enable separation of larger DNA sizes (longer sequencing reads).

Gene Construction. The synthetic oligonucleotide encoding Gene 1 (Figure 1) was PCR-amplified and digested by EarI to generate the macromonomer. Multimers were generated by self-ligation. A multimer containing 18 repeats of the GAGTGSA amino acid sequence (or six macromonomers) was selected. Sequencing showed that the selected gene had two serine-to-arginine mutations and that the actual sequence is $(GAGTGSA)_4GAGTGRA(GAGTGSA)_7GAGTGRA(GAGTGSA)_5$. A small number of positively charged residues are actually beneficial to FSCE separations as the charges "pull" the drag-tag in the opposite direction of the negatively charged DNA in an electric field, effectively increasing the hydrodynamic drag.[21] This sequence was used as the first protein polymer drag-tag (designated PN-18 for the 18 repeats and N for the N-terminal affinity tag used in its production) for FSCE DNA sequencing. Because of the successful demonstration of FSCE sequencing of ~180 bases using this drag-tag, larger multimers of 36 and 72 repeats of this sequence were generated via controlled cloning, which allows for production of large genes from smaller multimers in a well-controlled fashion. These genes were inserted into either the MpET-19b vector for expression with an N-terminal His tag or the MpET41a vector for expression with an N-terminal T7 tag and C-terminal His tag.

N-Terminal His Tag Protein Expression and Purification. Controlled cloning was utilized to generate genes 756 and 1512 bases in length (encoding 36 and 72 repeats of the seven amino acid "monomer") from the 378-bp multimer gene (18 repeats) through two rounds of gene doubling. These genes were inserted into the MpET-19b expression vector and expressed in E. coli BLR(DE3) cells. Proteins were purified on Talon cobalt resin under denaturing conditions, and the fractions were analyzed by SDS-PAGE. The expressed proteins were designated PN-36 and PN-72 according to the number of seven amino acid repeats. Figure 2 is the SDS-PAGE gel for a purification of PN-72. Note that the protein migrates higher on the gel than its expected molecular weight, likely due to the non-natural sequence and its lack of charged amino acids (besides the sparse arginine mutations and the His tag). Average protein yields ranged from 15 to 25 mg/L culture with the larger protein having lower yields. The N-terminal His tag was removed by cyanogen bromide cleavage because of the presence of an existing enterokinase restriction site engineered into the tag design (DDDDK). The tag must be
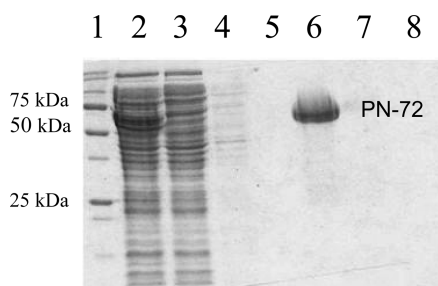
Figure 2. 12% SDS-PAGE gel of PN-72 purified from *E. coli* cell lysate by immobilized metal affinity chromatography (IMAC) on a column containing Talon cobalt-chelated resin. Lane 1: protein standards; lane 2: clarified cell lysate; lane 3: column flow through; lanes 4—5: washes; lanes 6—8: elutions.

Table 1. MALDI-TOF Analysis Results for PN Proteins

	expected mass (Da)	observed mass (Da)
PN-36	18 405	18 590
PN-72	36 736	37 085

removed to prevent the lysine acting as another reactive site for the sulfo-SMCC conjugation to ssDNA. Additionally, the inclusion of multiple negatively charged amino acids would significantly reduce the effective drag of the protein drag-tag. Amino acid analysis of the two proteins matched expected molar compositions (Supporting Information), and analytical RP-HPLC appeared to confirm that each protein was pure, consisting of a single peak on the chromatogram (data not shown). Likewise, MALDI-TOF analysis showed that the protein masses nearly matched expected values (Table 1), being only slightly higher than predicted. CD spectroscopy confirmed that the proteins exhibited random-coil conformations as designed. (See the Supporting Information.)

Characterization of N-Terminal His Tag Protein by FSCE. Protein polymers were further characterized by FSCE to determine the actual protein monodispersity and their suitability as drag-tags for free-solution DNA sequencing. PN-36 and PN-72 were conjugated to ssDNA primers and analyzed by free-solution capillary electrophoresis. No polymer matrix was utilized for the DNA separation beyond the small amount used as a dynamic wall coating agent. Figure 3 shows the FSCE electropherograms of PN-36 and PN-72. The peak on the far left of each electropherogram corresponds to the free (unconjugated) DNA, whereas the larger peak(s) on the far right corresponds to the drag-tag-DNA conjugate(s), which eluted later because of attachment of the drag-tag. The larger size and increased number of arginines of PN-72 leads to greater drag on the attached DNA, which elutes several minutes later in comparison with the DNA conjugated to PN-36. The smaller protein PN-18 displayed relatively minor impurities associated with a single peak,[5] but multiple distinct peaks of varying heights are observed in the electropherograms for PN-36 and PN-72, indicating that the attached protein polymer drag-tag is, in fact, rather heterogeneous. The polydispersity is more pronounced for the largest protein, PN-72, and appears to be related to the length of the protein polymer. The number of distinct bioconjugate peaks has increased from four to six. Additionally, the width of the distribution of peaks has increased from 1.7 min in Figure 3A to 6.6 min in Figure 3B.
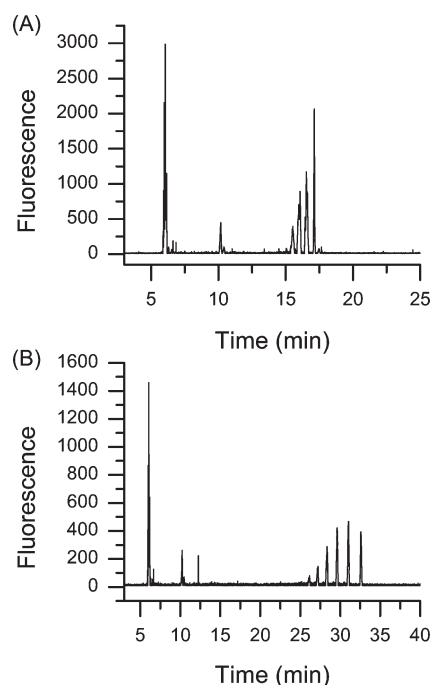

Figure 3. Free-solution capillary electrophoresis of drag-tag-DNA conjugates for (A) PN-36 (253 amino acids) and (B) PN-72 (505 amino acids) using a 30-base primer. ABI 3100, 36 cm array with 50 μM ID, 1× TTE, 7 M urea, 3%v/v POP5, 1 kV/1s injection, 312 V/cm, 55 °C.

Proteins in which the N-terminal His tag was removed by the enterokinase protease instead of chemical cleavage by cyanogen bromide exhibited similar profiles by FSCE (Supporting Information), indicating that the method of affinity tag removal is likely not the main contributing factor to the observed heterogeneity.

C-terminal His Tag Protein Expression and Purification. The highly repetitive nature of protein polymer amino acid sequences can be problematic for protein expression if the desire is to obtain completely monodisperse product. Premature protein truncation during synthesis has been previously observed for silk-based protein polymers.[24,25] Termination errors in protein synthesis may be due to depletion of available tRNA pools for certain codons, particularly any that are of low usage in *E. coli* synthesis.[24,26−28] Additionally, the Gene 1 sequence is ~43% glycine and 28% alanine. Therefore, it is likely that even relatively abundant species of tRNA can become depleted as well, despite utilizing a variety of Gly and Ala codons in the gene sequence. As a further precaution, protein expression is performed in the BLR(DE3) cell strain, which has an additional recombinase gene (*rec*A) knocked out compared with the more commonly utilized BL21 strain. Thus, potential repetitive gene recombination events are reduced or eliminated that could lengthen or shorten the gene within the plasmid, another possible source of heterogeneity.[24,29]

The 756 and 1512 bp genes (36 and 72 repeats of the seven amino acid "monomer") used above for expression were also inserted into MpET-41a and expressed in *E. coli* BLR(DE3) cells. A T7 tag was included at the N-terminus to enhance protein expression. These new proteins were designated PC-36 and PC-72. Although the sizes of the repetitive regions are comparable to that of the PN proteins, the addition of the T7 tag actually makes these proteins slightly larger. Specifically, PC-36 and PC-72
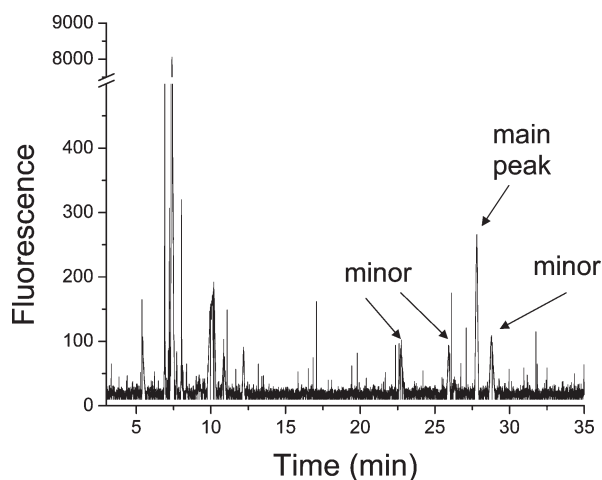
Figure 4. Free-solution capillary electrophoresis of drag-tag-DNA conjugates for PC-36 (277 amino acids) using a 23-base primer. ABI 3100, 36 cm array with 50 μM ID, 1× TTE, 7 M urea, 0.5% v/v POP6, 1 kV/20s injection, 312 V/cm, 55 °C.

have molecular weights of 21.1 and 39.4 kDa, respectively, when both the T7 tag and the His tag are attached. Protein yields ranged from 5 to 10 mg/L culture depending on the size of the protein being expressed. If truncation was occurring, then only full-length expressed proteins would have the C-terminal His tag and be isolated and purified by affinity chromatography. A reduction in yield would be expected with the exclusion of incomplete proteins from the purified product. Proteins expressed with the C-terminal His tag and no N-terminal T7 tag had even poorer yields in comparison and were not further studied (data not shown). Most likely due to the lower overall expression levels of the desired proteins using a C-terminal His tag, visible amounts of native protein contaminants were observed in the elution fractions by SDS-PAGE after affinity chromatography. Preparative RP-HPLC on a C18 column was used as a second purification step to remove these impurities. Although RP-HPLC cannot readily distinguish between protein polymers of widely varying sizes, the protein polymers do separate well from typically more hydrophobic natural *E. coli* proteins. (See the Supporting Information.) MALDI-TOF confirmed the molecular masses of the proteins.

Characterization of C-Terminal His Tag Protein by FSCE. Because neither the T7 tag nor the C-terminal His tag contain any lysine residues, neither tag was removed prior to conjugation to thiolated ssDNA using the heterobifunctional linker, sulfo-SMCC. However, analysis of the bioconjugates by free-solution capillary electrophoresis showed unexpectedly poor conjugation yields. Figure 4, the FSCE result for PC-36, shows that the free DNA peak is >10 times stronger in fluorescence intensity than the largest conjugate peak, indicating a very poor conjugation reaction yield in which most of the DNA did not attach to the protein polymer. This is in contrast with Figure 3A, the FSCE result for PN-36, which shows a free DNA peak only 1.5 times greater than the largest conjugate peak. Although Figure 4 does not demonstrate a noticeable improvement in protein purity when expressed with a C-terminal affinity tag, the peak pattern is distinctly different in comparison with previous results.

A plausible explanation for the observed low conjugation efficiency is that histidine is reacting with the sulfo-SMCC reagent during the conjugation step. Histidine can react with

N-hydroxysuccinimide (NHS) esters, effectively accelerating the rate of hydrolysis of the NHS groups in solution.[30,31] The unstable reaction product that is formed rapidly hydrolyzes. Typically, the NHS-ester reaction is performed first (i.e., drag-tag activation) to minimize hydrolysis because it is less resistant to hydrolysis than the maleimide group in sulfo-SMCC.[30] The histidines on the affinity tag may essentially be accelerating the hydrolysis of the reagent. Unlike a natural protein, there is only a single primary amine at the N-terminus of the protein polymer that may not be a strong enough nucleophile compared with the eight adjacent histidines at the C-terminus. Therefore, the sulfo-SMCC reagent preferentially reacts with the histidines, accelerating hydrolysis of the cross-linker and thus rendering it ineffective for conjugation as the cross-linker is now two separate molecules. Higher concentrations of sulfo-SMCC could be used to overcome this behavior. However, 100-fold excess of sulfo-SMCC reagent as opposed to the standard 10-fold molar excess showed no noticeable improvement in conjugation efficiency. Unfortunately, other commonly used small affinity tags such as FLAG (DYKDDDDK) or Strep tag (WSHPQFEK) cannot be used in place of the His tag due to the presence of lysines in the sequence. Sulfo-SMCC was determined to be the most suitable heterobifunctional protein−DNA linker because of its water-solubility and its chemoselective reactivity in the relevant pH range (near-neutral) and, finally, its robustness to the thermal cycling conditions required for a Sanger cycle sequencing reaction. Therefore, the purification method was changed as described below to enable continued use of this highly desirable linker molecule.

Removal of the Affinity Tag. It would be ideal to remove the C-terminal affinity tag completely to eliminate any possible side reactions that are causing either low conjugation yields or additional bioconjugate peaks to appear in the electropherogram. However, insertion of a methionine to act as a reactive site for cyanogen bromide cleavage was not as effective in this situation as it was for removal of the N-terminal affinity tag. After the cleavage reaction, the Met residue becomes the new C-terminus of the protein polymer. As part of the reaction, the methionine residue is converted into an equilibrium mixture of homoserine and homoserine lactone, which would result in at least two distinct peaks in an FSCE analysis that is performed at pH 8.5. (See the Supporting Information.)[19,32−35] Alternatively, site-specific proteases are commonly used to remove N-terminal affinity tags. However, enzymatic removal of a C-terminal affinity tag will result in part or all of the protease recognition sequence becoming the new C-terminus of the cleaved protein.

The protease Factor Xa was selected because only four additional amino acids (IEGR) from its recognition site would be added to the C-terminus of the cleaved protein (cleavage site is after Arg). Three of the amino acids have already been used in past or present protein polymer designs and are not expected to cause complications. Only one hydrophobic residue (isoleucine) is added to the protein. The negative charge of the glutamic acid is counteracted by the addition of a positively charged arginine. Adding two charged residues may also balance out the hydrophobicity of the isoleucine. Test cleavages were performed on PC-36 to determine appropriate reaction conditions for a larger scale reaction. The protein polymer, in general, does not stain well by Coomassie Blue because of its near neutral sequence, and at the low amounts used for these the studies, the proteins were unable to be visualized by SDS-PAGE. Therefore, the results were analyzed by the more sensitive Western blot method
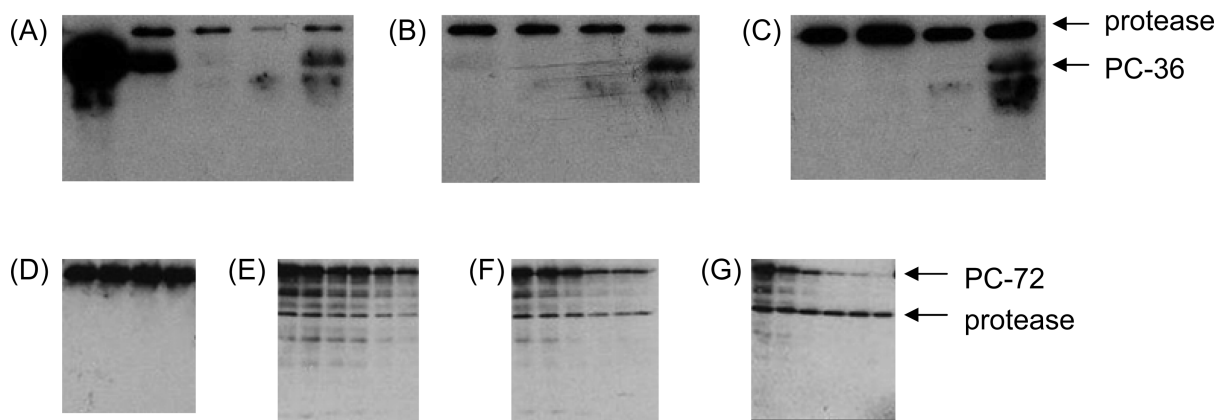
Figure 5. Western blot of endoproteinase GluC digestion of PC-36 protein over 16 h: (A) PC-36 no protease; 2, 4, 8, 16 h after 1:100 μg protease/μg protein added, (B) 2, 4, 8, 16 h after 1:50 μg protease/μg protein added, and (C) 2, 4, 8, 16 h after 1:20 μg protease/μg protein added. Western blot of endoproteinase GluC digestion of PC-72 protein over 12 h: (D) PC-72 after 0, 4, 8, 12 h with no protease added, (E) 2, 4, 6, 8, 10, 12 h after 1:100 μg protease/μg protein addition, (F) 2, 4, 6, 8, 10, 12 h after 1:50 μg protease/μg protein addition, and (G) 2, 4, 6, 8, 10, 12 h after 1:20 μg protease/μg protein addition.
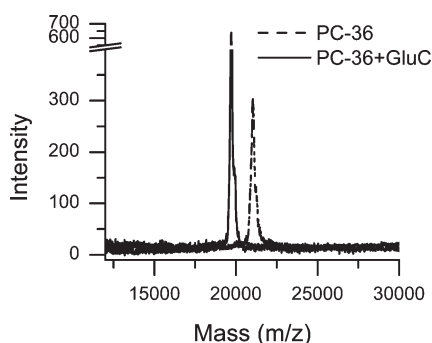

Figure 6. Overlay of MALDI-TOF results of PC-36 before and after digestion by endoproteinase GluC.

for Factor Xa to target protein ratios of 1:100, 1:50, and 1:20 (unit/μg) at 2, 4, 8, and 16 h time points at 20 °C. (See the Supporting Information.) The Western blot can only identify protein bands with a His tag still attached using the penta-His antibody. The addition of the protease unexpectedly showed evidence of digestion of the PC-36 target protein into multiple distinct bands. These bands most likely resulted from recognition of the four Gly-Arg mutation sites in the PC-36 sequence as cleavage sites by the enzyme. A reduction in temperature to 4 °C or 10-fold dilution of the protease concentration either completely halted enzymatic activity or failed to prevent nonspecific cleavage (data not shown).

Fortuitously, the addition of the IEGR recognition sequence also introduced a unique glutamic acid residue to the protein sequence. Endoproteinase GluC is a serine protease that can cleave specifically after Glu residues. This enzyme is typically used for peptide digestion and identification using mass spectrometry and not for affinity tag cleavage because natural proteins commonly contain Glu residues. The version sold by New England BioLabs includes a histidine tag at its C-terminus. Consequently, after protease digestion, the cleaved His tag, uncleaved protein, and the protease can all be removed in a single chromatographic step from the cleaved protein. Digestion was done at 25 °C in the provided reaction buffer using the protease-to-target-protein (μg/μg) ratios of 1:100, 1:50, and
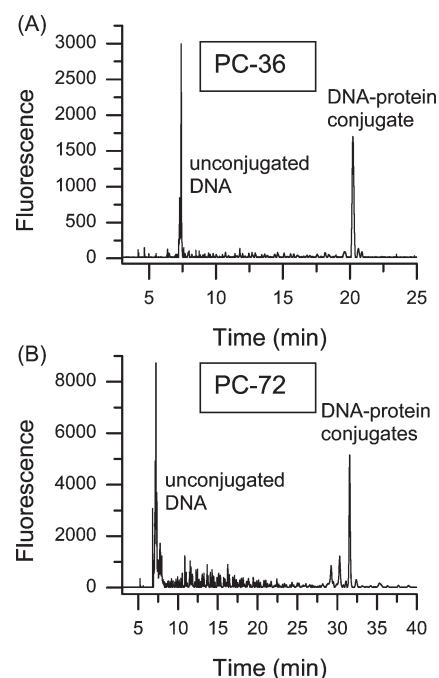

Figure 7. (A) Free-solution capillary electrophoresis of drag-tag-DNA conjugates for PC-36 with His tag removed (267 amino acids) using a 30-base primer. (B) Free-solution capillary electrophoresis of drag-tag-DNA conjugates for PC-72 with His tag removed (516 amino acids) using a 30-base primer. ABI 3100, 36 cm array with 50 μM ID, 1× TTE, 7 M urea, 0.5% v/v POP6, 1 kV/20s injection, 312 V/cm, 55 °C.

1:20. These reactions were monitored over the course of 16 h, and the results were analyzed by Western blot (Figure 5A–C). Cleavage was successful in <8 h for all protease concentrations tested. Endoproteinase GluC has a different mass than the PC-36 protein, and thus they are easily distinguishable on the blot. Additionally, the protease band remains essentially constant for each concentration. Curiously, faint bands are detected in all three 16 h reactions. These unexpected bands may be associated with side reactions of the long incubation time such as autolysis of the protease. A larger scale reaction was performed using 50 μg
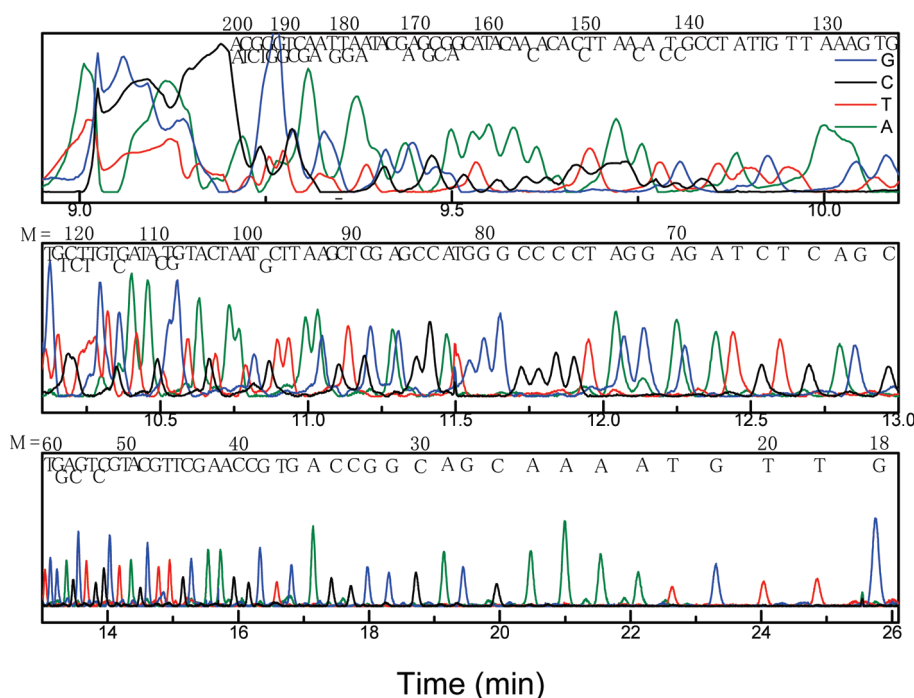
Figure 8. FSCE DNA sequencing using PC-36 drag-tag with no sieving polymer solution. Drag-tag was conjugated to M13 sequencing primer prior to Sanger reaction. Sequence is "read" backward starting at the bottom right. M is the base pair length of the Sanger fragment at that time point. ABI 3100, 36 cm array with 50 μM ID, 1X TTE, 7 M urea, 0.5% v/v POP6, 312 V/cm injection for 30 s, 312 V/cm, 55 °C.

of endoproteinase GluC and 5 mg of PC-36 for 6 h at 25 °C. After dialysis and lyophilization, the reaction mixture was resuspended in denaturing buffer and purified on Talon resin. However, in this situation, the flow-through and wash fractions were collected because they contained the desired, completely cleaved protein polymer. MALDI-TOF confirmed that the affinity tag was successfully removed by the enzyme, as evidenced by the mass shift shown in Figure 6, comparing measurements made before and after the reaction. We recovered ~5 mg of material in the flow-through fraction, indicating complete removal of the affinity tag.

The same endoproteinase GluC treatment was applied to the larger PC-72 protein. Western blot analysis of the test cleavages (Figure 5D−G) shows the presence of multiple bands upon the addition of the protease, but the PC-72 protein itself was stable in the reaction buffer over the entire 12 h if no protease was added. Interestingly, these bands were not as well-defined as those resulting from Factor Xa digestion, indicating that nonspecific cleavage was likely occurring at multiple locations and not specifically at Gly-Arg regions. We reacted 6 mg of PC-72 for 12 h at 25 °C using a 1:20 μg/μg ratio of protease to PC-72. The reaction was purified by column chromatography. All material was recovered in the flow through and wash fractions, indicating complete removal of the affinity tag.

Analysis by FSCE of Endoproteinase GluC-Digested Proteins. The cleaved PC-36 protein was conjugated via sulfo-SMCC to ssDNA primer and analyzed by free-solution capillary electrophoresis. Figure 7A is an electropherogram showing that the bioconjugate is almost completely monodisperse and that the conjugation efficiency has significantly improved with the removal of the histidine-containing affinity tag. A couple minor peaks of unknown origin can be observed in the electropherogram. These may be due to protease cleavage at other sites along the affinity tag such as the G or R residues adjacent to the

glutamic acid. Overall, the PC-36 protein is significantly more monodisperse than the version previously expressed using an N-terminal affinity tag when both are analyzed by FSCE (Figure 3A). This protein has double the mass and hydrodynamic drag of the PN-18 protein used previously for successful FSCE DNA sequencing and is expected to produce even longer sequencing reads with its improved ability (i.e., greater hydrodynamic drag) to separate larger DNA sizes in free solution.

The cleaved PC-72 was also conjugated to DNA and analyzed by FSCE (Figure 7B). At least two smaller peaks are clearly visible that elute prior to the main peak. The baseline is also noisier compared with the PC-36 analysis, likely the byproduct of nonspecific enzymatic cleavage. As expected, the larger size of PC-72 allows it to better separate two different sizes of DNA by FSCE in comparison with PC-36. (See the Supporting Information.) Although not as monodisperse as the GluC-cleaved PC-36 protein, the monodispersity is noticeably improved over Figure 3B. MALDI-TOF matches the expected mass of the PC-72 protein before and after removal of the His tag. However, the protein polymer appears as a single, broad peak by MALDI-TOF that could not be further resolved. This is typical for its size (38 kDa), regardless of whether the protein is expressed with an N- or C-terminal affinity tag. It is interesting to note that MALDI-TOF of PC-36 and PC-72 are at the exact expected sizes in contrast with the previous PN proteins. This may indicate that post-translational modification in addition to protein truncation contributed to the observed polydispersity. It is apparent that removal of the affinity tag by endoproteinase GluC has drawbacks when applied to proteins much larger than PC-36, and further studies are needed to identify the cause of and reduce the polydispersity observed in the PC-72 protein.

The 1:20 through 1:100 (w/w) ratio of enzyme to substrate used for removal of the His tag is within the recommended range

for this protease. It is not expected that the random coil structure of these proteins is preventing protease accessibility. A reduction in reaction time to 6 h from 12 h had no noticeable effect on the final FSCE analysis for PC-72 nor did reducing the enzyme concentration five-fold to match the reaction conditions used for PC-36 (data not shown). In other words, using the same reaction conditions as the PC-36 digestion along with the same mass of protein (albeit half the molar amount) did not affect results beyond lowering the yield of fully cleaved protein. The addition of more protease may allow for these side reactions to proceed further toward completion, thus reducing the size of the secondary peaks but likely also reducing the final amount of the main peak. This strategy is currently being investigated.

FSCE DNA Sequencing Using New Drag-Tags. PC-36 and PC-72 were both tested as drag-tags in FSCE DNA sequencing reactions. Unfortunately, sequencing was not successful with the PC-72 drag-tag attached to the primer, possibly because of interference from the larger protein with primer binding or polymerase activity, which consequently prevented the generation of Sanger fragments. Most likely a post-Sanger conjugation strategy will need to be devised to avoid potential interference of the enzymatic reaction by the attached large drag-tag. Preliminary FSCE DNA sequencing with the monodisperse PC-36 drag-tag was successful, although the initial read length was only slightly longer than the 127 aa protein previously used. The raw sequencing data is shown in Figure 8. Knowing the sequence ahead of time, the peaks can be read out to ~200 bases. FSCE sequencing data is "read" backward starting at the bottom right and moving toward the upper left. The last peaks to elute are the smallest fragments, which were slowed the most by the attachment of the drag-tag. Although this protein is twice the size and charge of the drag-tag used in the previously published result[5] and was made using a new cloning and expression strategy, the peaks are still quite sharp, indicating that the protein is not sticking noticeably to the capillary walls. Additional changes to injection or running conditions may improve the quality of the data and increase the read length further.

CONCLUSIONS

We have demonstrated a cloning and purification strategy that has enabled the generation of a completely monodisperse, 267-amino acid protein polymer (PC-36) and a nearly monodisperse 516-amino acid protein polymer (PC-72) as analyzed by a highly sensitive, fluorescence-based detection method called FSCE. This method has been shown to be more sensitive than MALDI-TOF, SDS-PAGE, or RP-HPLC in detecting small variations in size or charge when the protein polymer is conjugated to ssDNA and analyzed in free solution in an electric field. These proteins are significantly larger in size than the 127-amino acid protein previously used successfully to demonstrate DNA sequencing by FSCE using protein polymer drag-tags. Generation of these new drag-tags using the improved cloning and purification strategy has overcome a key obstacle in the further development of FSCE for sequencing. We expect that use of these larger, yet still monodisperse, drag-tags for sequencing will enable even longer reads in free-solution electrophoresis than the current 180 bases, potentially obtaining read lengths close to 400 bases. Preliminary sequencing results with the PC-36 protein show that at least 200 bases of read length is possible with the new drag-tags. Additional FSCE sequencing results and analyses performed with these protein polymer drag-tags are detailed in a companion paper.[36] When applied to microfluidic devices, FSCE DNA sequencing will enable fast, low-cost, and accurate sequencing of ~400 reads for biomedical applications with a sufficiently large and monodisperse drag-tag and optimized sequencing protocol. The methods discussed herein may also be beneficial to protein-based polymers being used for biomaterials or other research applications where precise characterization and true monodispersity is important.

ASSOCIATED CONTENT

ⓢ Supporting Information. Experimental details of oligonucleotide sequences, amino acid compositional analysis, CD spectroscopy, MALDI-TOF spectra, and additional Western blot and FSCE results. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*Phone: (650) 721-1151. Fax: (650) 723-9801. E-mail: aebarron@stanford.edu.

ACKNOWLEDGMENT

REFERENCES

(1) Mayer, P.; Slater, G. W.; Drouin, G. *Anal. Chem.* **1994**, 66, 1777–1780.

(2) Noolandi, J. *Electrophoresis* **1992**, 13, 394–395.

(3) Noolandi, J. *Electrophoresis* **1993**, 14, 680–681.

(4) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, 74, 5463–5467.

(5) Meagher, R. J.; Won, J. I.; Coyne, J. A.; Lin, J.; Barron, A. E. *Anal. Chem.* **2008**, 80, 2842–2848.

(6) Meagher, R. J.; Won, J. I.; McCormick, L. C.; Nedelcu, S.; Bertrand, M. M.; Bertram, J. L.; Drouin, G.; Barron, A. E.; Slater, G. W. *Electrophoresis* **2005**, 26, 331–350.

(7) Haynes, R. D.; Meagher, R. J.; Won, J. I.; Bogdan, F. M.; Barron, A. E. *Bioconjugate Chem.* **2005**, 16, 929–938.

(8) van Hest, J. C. M.; Tirrell, D. A. *Chem. Commun.* **2001**, 1897–1904.

(9) Davis, N. E.; Karfeld-Sulzer, L. S.; Ding, S.; Barron, A. E. *Biomacromolecules* **2009**, 10, 1125–1134.

(10) Kiick, K. L. *Polym. Rev.* **2007**, 47, 1–7.

(11) Vreeland, W. N.; Desruisseaux, C.; Karger, A. E.; Drouin, G.; Slater, G. W.; Barron, A. E. *Anal. Chem.* **2001**, 73, 1795–1803.

(12) Vreeland, W. N.; Slater, G. W.; Barron, A. E. *Bioconjugate Chem.* **2002**, 13, 663–670.

(13) Won, J. I.; Meagher, R. J.; Barron, A. E. *Biomacromolecules* **2004**, 5, 1624–1624.

(14) Won, J. I.; Meagher, R. J.; Barron, A. E. *Electrophoresis* **2005**, 26, 2138–2148.

(15) Hert, D. G.; Fredlake, C. P.; Barron, A. E. *Electrophoresis* **2008**, 29, 4618–4626.

(16) Won, J. I.; Barron, A. E. *Macromolecules* **2002**, 35, 8281–8287.

(17) Garnier, J.; Gibrat, J. F.; Robson, B. *Methods Enzymol.* **1996**, 266, 540–553.

(18) Rydzanicz, R.; Zhao, X. S.; Johnson, P. E. *Nucleic Acids Res.* **2005**, *33*, W521–525.

(19) Gross, E. *Methods Enzymol.* **1967**, *11*, 238–255.

(20) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: San Diego, CA, 1996; p 785.

(21) Coyne, J. C.; Lin, J. S.; Barron, A. E. In *Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques*, 3rd ed.; Landers, J. P., Ed.; CRC Press: New York, 2008; p 1567.

(22) Meagher, R. J. Ph.D. Dissertation, Northwestern University, Evanston, IL, 2005.

(23) Meagher, R. J.; McCormick, L. C.; Haynes, R. D.; Won, J. I.; Lin, J. S.; Slater, G. W.; Barron, A. E. *Electrophoresis* **2006**, *27*, 1702–1712.

(24) Fahnestock, S. R.; Irwin, S. L. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23–32.

(25) Huang, J.; Valluzzi, R.; Bini, E.; Vernaglia, B.; Kaplan, D. L. *J. Biol. Chem.* **2003**, *278*, 46117–46123.

(26) Robinson, M.; Lilley, R.; Little, S.; Emtage, J. S.; Yarranton, G.; Stephens, P.; Millican, A.; Eaton, M.; Humphreys, G. *Nucleic Acids Res.* **1984**, *12*, 6663–6671.

(27) Rosenberg, A. H.; Goldman, E.; Dunn, J. J.; Studier, F. W.; Zubay, G. *J. Bacteriol.* **1993**, *175*, 716–722.

(28) Kane, J. F. *Curr. Opin. Biotechnol.* **1995**, *6*, 494–500.

(29) Prince, J. T.; McGrath, K. P.; Digirolamo, C. M.; Kaplan, D. L. *Biochemistry* **1995**, *34*, 10879–10885.

(30) Mattson, G.; Conklin, E.; Desai, S.; Nielander, G.; Savage, M. D.; Morgensen, S. *Mol. Biol. Rep.* **1993**, *17*, 167–183.

(31) Cuatrecasas, P.; Parikh, I. *Biochemistry* **1972**, *11*, 2291–2299.

(32) Kuliopulos, A.; Walsh, C. T. *J. Am. Chem. Soc.* **1994**, *116*, 4599–4607.

(33) Armstrong, M. D. *J. Am. Chem. Soc.* **1949**, *71*, 3399–3402.

(34) Murphy, C. M.; Fenselau, C. *Anal. Chem.* **1995**, *67*, 1644–1645.

(35) Lee, T. D.; Shively, J. E. *Methods Enzymol.* **1990**, *193*, 361–374.

(36) Albrecht, J. C.; Lin, J. S.; Barron, A. E. *Anal. Chem.* **2011**, *83*, 509–515.