BIOMACROMOLECULES
Article
pubs.acs.org/Biomac

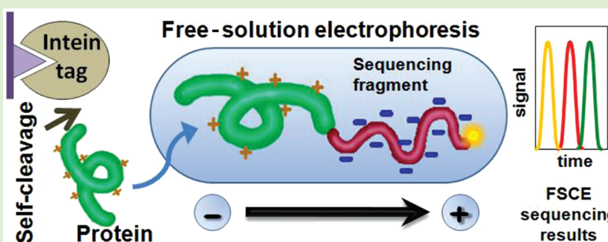
Monodisperse, "Highly" Positively Charged Protein Polymer Drag-Tags Generated in an Intein-Mediated Purification System Used in Free-Solution Electrophoretic Separations of DNA

Xiaoxiao Wang,[1] Jennifer Coyne Albrecht,[2] Jennifer S. Lin,[1] and Annelise E. Barron*,[1]

[1]Departments of Bioengineering and [2]Chemical Engineering, Stanford University, Stanford, California 94305, United States

S *Supporting Information*

ABSTRACT: Free-solution conjugate electrophoresis (FSCE) is a method of DNA sequencing that eliminates the need for viscous polymer solutions by tethering a carefully designed, mobility modifying "drag-tag" to each DNA molecule to achieve size-based separations of DNA. The most successful drag-tags to date are genetically engineered, highly repetitive polypeptides ("protein polymers") that are designed to be large, water-soluble, and completely monodisperse. Positively charged arginines were deliberately introduced at regular intervals into the amino acid sequence to increase the hydrodynamic drag without increasing drag-tag length. Additionally, a one-step purification method that combines affinity chromatography and on-column tag cleavage was devised to achieve the required drag-tag monodispersity. Sequencing with a read length of approximately 180 bases was successfully achieved with a known sequence in free-solution electrophoresis using one of these positively charged drag-tags. This preliminary result is expected to lead to further progress in FSCE sequencing with ~400 bases read length possible when more "highly" positively charged protein polymers of larger size are generated with the intein system.



INTRODUCTION

Size-based separations of DNA are used in many fields, including molecular biology, forensic analysis, and DNA sequencing, to name a few. The ability to achieve high-resolution separations of DNA for sequencing generally requires a highly viscous sieving polymer network to separate Sanger fragments[1] by length by electrophoresis. To date, the only reliable electrophoretic method able to separate DNA by size without a sieving polymer solution is Free-Solution Conjugate Electrophoresis (FSCE, which is also sometimes called, End-Labeled Free-Solution Electrophoresis or ELFSE).

FSCE achieves size-based separation of DNA in free-solution electrophoresis by attaching a mobility modifier ("drag-tag") to each DNA molecule.[2,3] DNA has a constant charge-to-friction ratio under free-solution electrophoresis conditions and thus does not separate on basis of size. In FSCE, the balance of charge-to-friction is broken by conjugating each DNA molecule to a monodisperse drag-tag with a different charge-to-friction ratio than DNA. Therefore, the electrophoretic mobility of each DNA/drag-tag bioconjugate becomes size-dependent without the need for a polymeric sieving matrix. DNA sequencing by FSCE is achieved by separating Sanger fragments with single-base resolution. FSCE is ideal for implementation on automated, parallel microfluidic devices since it avoids the use of the difficult to load viscous polymer solutions. Engineering appropriate macromolecular materials suitable as drag-tags is key to the success of FSCE sequencing.

An ideal drag-tag is completely monodisperse, water-soluble, has minimal adsorption to microchannel walls, and can be uniquely and stably attached to DNA, as described previously.[3,4] One of the most important properties for a drag tag is complete monodispersity, where every tag is identical in sequence, charge, molecular weight and mean hydrodynamic radius. If a polydisperse drag-tag is used, each DNA length will be represented in the electropherogram by multiple peaks instead of a single peak, potentially making accurate DNA sizing very difficult. Another important property for a drag tag is its hydrodynamic drag. A higher overall hydrodynamic drag enables resolution of longer sequencing fragments and, consequently, a longer sequencing read length is obtainable. The dual requirements of monodispersity and large hydrodynamic size remove all commonly available nano-particles and synthetic polymers, including polyethylene glycol (PEG)[5] as well as small, chemically synthesized polypeptoids (poly-*N*-substituted glycines)[6,7] from consideration as useful drag-tag candidates for long-read FSCE sequencing.

The first demonstration of FSCE sequencing achieved 110 bases of sequencing with a purified (but still slightly polydisperse) recombinant streptavidin protein drag-tag.[8] However, the chemically diverse (charged and hydrophobic) surface of streptavidin interacted with the capillary walls, producing broad peaks. Additionally, the globular structure of natural proteins renders them suboptimal drag-tag candidates, because it is the hydrodynamic radius of the drag-tag that

Received: September 22, 2011
Revised: November 18, 2011
Published: December 15, 2011

© 2011 American Chemical Society

dx.doi.org/10.1021/bm2013313 | *Biomacromolecules* 2012, 13, 117−123

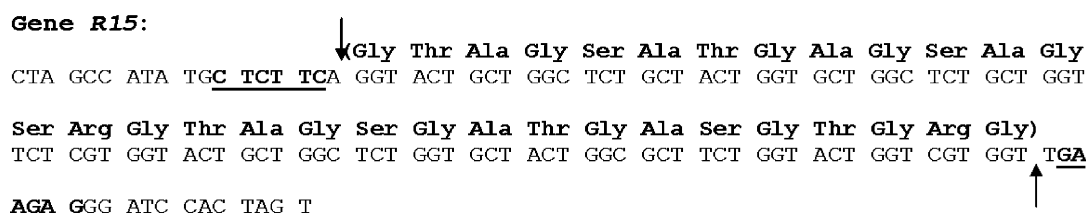


Figure 1. DNA sequence of the macromonomer used to generate the repetitive genes. *Ear* I restriction sites are in bold and underlined and arrows indicate the cleavage site of the endonuclease.

matters, not its molecular weight per se. In contrast, genetically engineered "protein polymers" can be designed to meet the many stringent requirements of an "ideal" drag-tag. Protein polymers are repeats of a peptide monomer sequence that can either mimic a natural protein motif[9−12] or be a highly non-natural and specifically designed sequence.[13,14] Potentially problematic charged or reactive amino acids can be reduced or eliminated from the sequence design. Our group has developed a series of genetically engineered, highly repetitive, random-coil protein polymers that meet the requirements of FSCE sequencing.[15,16] Protein polymer drag-tags were originally designed to be neutrally charged. Positive charges might cause undesirable interactions with either the negatively charged DNA or microchannel walls.. Negative charges would actually reduce the effective drag and thus separating power of the drag-tag by "pulling" the protein in the same direction as the DNA in an electric field.

FSCE sequencing of ~180 bases using a 127-aa protein was the first reported sequencing with a genetically engineered protein polymer with a +1 net positive charge.[16] However, achieving longer DNA sequencing lengths (>400 bases) requires an essentially monodisperse drag-tag with significantly higher hydrodynamic drag (i.e., radius of gyration), and producing larger monodisperse protein polymers has been a challenging task.[15] The histidine tag used for purification had to be removed after affinity chromatography due to adverse interactions with the heterobifunctional linker (sulfo-SMCC) used for coupling DNA to drag-tags.[15,17,18] Enzymatic removal of the C-terminal His tag was used to obtain larger protein polymers with improved monodispersity leading to DNA sequencing of 265 bases.[19] However, the enzymatic cleavage method is not ideal because it is expensive, and time-consuming, and digestion at nonspecific sites produces some polydispersity in the larger proteins. To address this issue, new strategies for both better purification and increasing the hydrodynamic drag need to be developed and are discussed herein.

The intein-mediated purification system is a simple, efficient option for obtaining pure, tag-free recombinant proteins in a one-step purification through on-column cleavage and removal of the affinity tag under mild conditions.[20,21] It was previously found that a small number of positive charges (3−8) in longer protein polymers may increase the hydrodynamic drag without detrimental interaction with the capillary walls.[15,16] The additional cationic residues introduced by *E. coli* mutations during cloning work provided a slight backward "pull" on the drag-tag during electrophoresis, thus increasing its "drag power." in FSCE. Building on these results, our lab sought to explore the possibility of intentionally including the positively charged amino acid arginine in sequences of protein polymer drag-tags. Lysine could not be used here because our conjugation method requires a unique free amine site (N-terminus) to react with the cross-linker.

A new series of "highly" positively charged protein polymers was thus designed to increase hydrodynamic drag for a given protein polymer length and then purified using the intein method. We say "highly" because the protein polymers designed and generated in this paper are much more charged than previous protein polymer drag-tags. Two "highly" positively charged (with 6 or 12 Arg in the entire sequence) protein polymers with 110 and 182 amino acids, respectively, were constructed by molecular cloning, expressed in *E. coli*, and purified using an intein-mediated purification method that was adapted to our cloning system.[22] Both protein polymers showed improved monodispersity in FSCE separations compared with previous purifications with enzymatic His-tag removal. Also, the inclusion of positive charges was proven to increase their overall hydrodynamic drag. One 110-aa drag-tag (6 Arg) was successfully used for FSCE sequencing with ~180 bases read length of a known sequence. Longer read length is expected using "highly" charged protein drag-tags with greater length and positive net charge in the near future. A discussion of the physics of FSCE using slightly cationic drag-tags will be presented in an upcoming paper.

■ MATERIALS AND METHODS

All molecular biology techniques were conducted according to standard protocols or from instructions provided by manufacturers unless otherwise noted. All enzymes were obtained from New England Biolabs (NEB, Ipswich, MA), except specifically stated. The intein-mediated purification system (including the unmodified pTXB1 vector and chitin beads) was also purchased from NEB. General reagents for cloning and protein expression were obtained from Fisher Scientific (Pittsburgh, PA) unless noted otherwise.

Gene Generation for Repetitive Protein Polymers. A 113-bp single-stranded synthetic oligonucleotide was designed to encode a 30-amino acid sequence consisting of 4 neutral, hydrophilic amino acids (Ala, Gly, Thr, Ser) with 1 Arg per 15 amino acids [i.e., 30 amino acids including 2 Arg residues per single-stranded DNA (ssDNA)]. The charge density chosen here was expected to add overall drag without significantly causing peak broadness based on previous studies. The gene and amino acid sequences are 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 90-bp fragment was isolated and purified using agarose gel electrophoresis and the QIAEX II Gel Extraction Kit (Qiagen, Valencia, CA). Multimers of the gene were then 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.[22] The plasmids were transformed via heat shock into NovaBlue competent cells (Novagen, Madison, WI). The resulting colonies were screened by DNA sequencing to verify the identity and size of the insert DNA. Two genes with desired sequences were selected from the sequencing

results, one consisting of 6 repeats of the initial gene, while the other contained spontaneous mutations which caused the average Arg spacing to increase from 1 in 15-aa to 1 in 18-aa (3 repeats of the mutant gene).

Generation of Expression Vector with C-Terminal Intein-Tag. The C-terminal intein-tag vector (the intein-tag will be fused to the C-terminus of the target protein), pTXB1, was chosen to eliminate polydispersity that can result from premature protein truncation during synthesis. A further advantage is that the C-terminal intein-tag exhibits better on-column cleavage yields than the N-terminal version. The pTXB1 vector was modified to be suitable for our controlled cloning system. A 54-bp adapter oligonucleotide containing a newly designed cloning region (5′-CAT ATG GGT TGA AGA GCC GTA CAT GAG CTC TGC ACG GGC TCT TCA GGT GCG TGC-3′) was generated by annealing two complementary single-stranded oligonucleotides (Stanford University Protein and Nucleic Acid Facility). The annealing reaction was conducted with 12.5 μM of each ssDNA. After denaturing at 95 °C for 5 min, the two ssDNA were annealed with a temperature ramping protocol that decreased from 85 to 75 °C in 30 min and then rapidly decreased to 4 °C. The multiple cloning site (MCS) of the original pTXB1 vector was removed by double digestion using two restriction enzymes, *Nde* I and *Sap* I. The 54-bp adapter DNA was inserted to create the modified pTXB1 vector, MpTXB1. The new cloning region contained two *Sap* I recognition sites (GCTCTTC) for the controlled cloning system. An alanine codon was introduced just before the first codon of the intein sequence to yield higher on-column cleavage activity (according to the accompanying instruction manual for the IMPACT system). The recipient vector was prepared by digesting the circular plasmid with *Sap* I at 37 °C. Slab gel purification was performed to isolate the desired vector band. Finally, the vector was reacted with calf intestinal phosphatase (CIP) for 1 h at 37 °C to minimize recircularization of the plasmid in subsequent ligation steps.

Protein Expression and Purification via Intein-Mediated System. The desired genes were excised from the pUC18 cloning vector via *Sap* I digestion and were ligated into the MpTXB1 recipient vector. The identity of the resulting plasmid DNA was confirmed by sequencing. Plasmids were transformed into *E. coli* BLR(DE3) expression cells (Novagen) via heat shock. Protein expression was performed using Terrific Broth media (EMD Biosciences, San Diego, CA) under tetracycline (12.5 μg/mL) and carbenicillin (50 μg/mL) antibiotic selection. A 25 mL overnight culture grown in LB media was inoculated into 1 L of Terrific Broth and grown at 37 °C. After the cells reached OD_{600} = 0.4, the temperature was decreased to 16 °C and isopropyl-β-D-thiogalactoside (IPTG, U.S. Biologicals, Swampscott, MA) was added at a final concentration of 0.4 mM to induce protein synthesis. After 20 h of additional growth time at 16 °C, cells were harvested by centrifugation at 6000 g at 4 °C for 12 min.

Affinity purification and on-column cleavage was performed according to the protocols in the instruction manual for the IMPACT system with a couple of modifications in order to obtain better results for our proteins. First, the concentration of NaCl in the column buffer was changed from 500 to 350 mM to decrease the ionic strength of the buffer and reduce the presence of impurities. Second, the clarified cell extract was gently mixed with prepared chitin beads at 4 °C for 2 h before loading onto the column for enhanced binding. Overnight on-column cleavage was conducted at 25 °C with 50 mM DTT added to the column buffer. Cell lysate, flow through, washes, samples before DTT treatment, elutions and samples after elution were all analyzed on a discontinuous 12% SDS-PAGE gel stained by Coomassie Blue R-250. Elutions containing the target protein were combined and dialyzed three days against deionized water at 4 °C. Finally the protein was lyophilized into a dry powder.

To further purify the proteins and produce monodisperse drag-tags for FSCE applications, reversed-phase high performance liquid chromatography (HPLC) was performed after affinity purification to eliminate coeluted intein-tag as well as other impurities. Approximately 10−15 mg of protein was dissolved in 4 mL of water and then loaded onto a Phenomenex Jupiter C18 column (10 μm, 300 Å, 21.2 × 250 mm). A linear gradient of 5−95% solvent B (0.1% TFA in acetonitrile (ACN) (v/v)) in solvent A (0.1% trifluoroacetic acid (TFA) in water (v/v)) over 35 min at a flow rate of 15 mL/min was used. The target protein eluted at approximately 35% ACN. Fractions were lyophilized, resuspended in water, and pH adjusted to near neutral and then lyophilized again.

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 (Protein and Nucleic Acid Facility, Stanford University) was used for MALDI-TOF analysis of the protein using sinapinic acid as the matrix.

Protein Analysis and Characterization by FSCE. Proteins polymers were characterized by FSCE to further determine their purity and, consequently, their suitability as drag-tags for free-solution DNA sequencing. Because these drag-tags are charged, it is no longer possible to estimate the effective drag coefficient of the tag using published theoretical predictions and formulas. A new proposed method to estimate this drag coefficient will be discussed in an upcoming paper. Oligonucleotides with a 5′-thiol and fluorescein label were purchased from Integrated DNA Technologies (SH-CCT* TTT AGG GTT TTC CCA GTC ACG ACG TTG, 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 70 mM sodium phosphate buffer, pH 7.2.[16,19,23] Protein polymers were activated at the N-terminus with a maleimide by the addition of the heterobifunctional cross-linker sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (sulfo-SMCC, Pierce). A 10:1 molar excess of sulfo-SMCC was added to 1.2 mg 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 removed 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, and then resuspended in water at 10 mg/mL concentration.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.

For sequencing, the protein drag-tag was instead conjugated to a thiol-containing M13 sequencing primer (SH-GTT TTC CCA GTC ACG AC from IDT) containing no fluorophore. Eight microliters of BigDye terminator v1.1 cycle sequencing mix (Applied Biosystems, Foster City, CA), 0.16 μg of M13mp18 ssDNA template, and water were mixed and combined 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, 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 (Applied Biosystems, Foster City, CA) with a 16-capillary array of fused silica capillaries (50 μM inner diameter) and 4-color laser-induced fluorescence (LIF) detection with a 488 nm laser was used to analyze the protein−DNA conjugates. Capillary electrophoretic separations of the conjugates were performed in denaturing buffer consisting of 89 mM Tris, 89 mM TAPS, 2 mM EDTA, and 7 M urea. A 0.5% (v/v) POP-6 polymer solution was used as a dynamic wall coating agent to suppress electro-osmotic flow and minimize adsorption to capillary walls. Capillaries with an effective length of 36 cm were used for FSCE separations. Typical electrophoresis conditions include electrokinetic injection with a potential of 1−2 kV applied for 5−30 s and a running voltage of 15 kV, all at 55 °C.

■ RESULTS AND DISCUSSION

A streamlined cloning and purification protocol was devised to achieve completely monodisperse repetitive polypeptides ("protein polymers") for Free-Solution Conjugate Electrophoresis (FSCE). A new series of protein polymers with a novel sequence and regularly spaced arginine amino acids was designed and produced to test the inclusion of cationic residues. Sequencing of approximately 180 bases of a known sequence was successfully achieved in free-solution electrophoresis with no polymer sieving matrix using one of these "highly" positively charged protein polymer drag-tags.

Gene Construction. The synthetic "monomer" gene *R15* (Figure 1) was designed to encode a water-soluble protein polymer that adopts a random-coil structure (in order to maximize its hydrodynamic radius). Two evenly spaced arginine residues in the sequence of gene *R15* were used to introduce cationic charges. These positive charges increase the effective hydrodynamic drag by exerting a "pulling" force on the drag-tag in the opposite direction of the negatively charged DNA in an electric field. Two multimer genes cloned from *R15* were selected: a 540-bp gene *R15–6* with six repeats of *R15*, and a 324-bp gene with three repeats of a mutant version of *R15*. The mutation inserted a sequence encoding GTAGSA after the second Ala, resulting in the average arginine spacing increasing to one per 18 amino acid residues. Therefore, this new gene was renamed *R18–3*. The amino acid sequences encoded by *R15–6* and *R18–3*, respectively, are $(\text{GTAGSATGAGSAGSR-GTAGSGATGASGTGR})_6$ and $(\text{GTAGSAGTAGSATGAGSAGSR-GTAGSGATGASGTGR})_3$. These sequences contain a higher number of positive charges (*R15–6:* 12 Arg in 182 amino acids, and *R18–3*: 6 Arg in 110 amino acids) than our previous protein polymer drag-tags, and therefore, these proteins were expected to have a larger amount of effective hydrodynamic drag per unit length. These genes were either inserted into the MpET-41a vector for expression with a C-terminal His tag or into the modified pTXB1 for expression with a C-terminal intein-CBD tag.

C-Terminal His Tag Protein Polymers: Expression and Purification. The *R15–6* gene was cloned into MpET-41a, expressed with a C-terminal His tag and N-terminal T7 tag (for improved yield), and purified using the same methods described previously.[15] The expressed protein was designated CR15–6 with the "C", indicating the C-terminal His tag was used in its production. The average protein yield is about 10 mg/L culture. The SDS-PAGE gel for the purified protein CR15–6 showed trace amounts of native protein contaminants in the elution fractions (Figure S2 in Supporting Information). Therefore, preparative RP-HPLC on a C18 column was used as a second purification step to remove these impurities (Figure S3 in Supporting Information). MALDI-TOF confirmed the molecular mass of the protein (Figure 2).

Endoproteinase GluC was chosen for His tag removal as in previous work.[15] Digestion was performed at 25 °C for 6 h in the provided reaction buffer using 50 μg of enzyme for 5 mg of the target protein (CR15–6). After protease digestion, the cleaved His tag, uncleaved protein, and the protease (which also has a His tag) were removed from the cleaved protein in a single chromatographic step. SDS-PAGE confirmed the successful removal of the His tag as seen by a size shift (Figure S4 in Supporting Information). Almost all of the starting material was recovered from flow-through and wash fractions, indicating complete removal of the affinity tag.

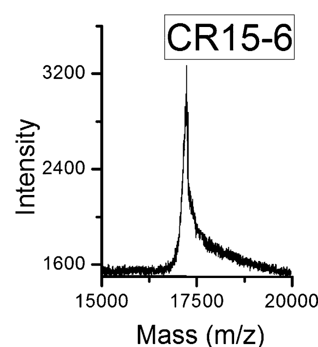


Figure 2. MALDI-TOF results for CR15–6 before His tag removal. The observed mass is 17233.21 Da, whereas the estimated mass for CR15–6 with His tag is 17216.68 Da.

Characterization of Endoproteinase GluC Digested "Highly" Charged Protein Polymers by FSCE. The cleaved CR15–6 protein was conjugated to a 30-nt ssDNA oligomer and analyzed by FSCE to determine its monodispersity and suitability as a drag-tag for free-solution DNA sequencing. The electropherogram is shown in Figure 3. The "free" (unconjugated) DNA elutes first and corresponds to the peak on the far left of the electropherogram. The DNA drag-tag conjugate elutes later due to the additional hydrodynamic drag (larger peaks on the far right of Figure 3). Although CR15–6 is about three-quarters the size of the previous protein polymer drag-tag PC-36[15] (195- vs 267-aa), it showed comparable drag (i.e., elution times) to PC-36 when analyzed by FSCE under identical conditions. The smaller size of CR15–6 is compensated for by the increased number of arginine residues (12 in CR15–6 compared to 4 in PC-36), indicating that increased hydrodynamic drag can be achieved without greatly increasing the protein size by intentionally including a few positive charges in the drag-tag sequence.

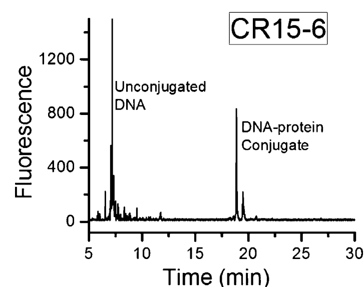


Figure 3. Free-solution capillary electrophoresis of drag-tag-DNA conjugates for CR15–6 without His tag (195 amino acids, 12 Arg) using a 30-base ssDNA oligomer. ABI 3100, 36 cm array with 50 μM ID, 1X TTE (buffer of Tris, TAPS, and EDTA), 7 M urea, 0.5%v/v POP-6 (for dynamic coating), 1 kV/20s injection, 310 V/cm, 55 °C.

The extra peaks observed in Figure 3 that are clustered around the DNA–protein conjugate peak indicate that CR15–6 is heterogeneous. A two-peak pattern is seen in MALDI-TOF for CR15–6 after endoproteinase GluC digestion (Supporting Information, Figure S5) which suggests that at least some of the polydispersity is likely caused by the His tag removal step. The endoproteinase GluC digestion was performed at pH 8.5, where a deamidation reaction is possible at the two Gln residues in the T7 tag that changes Gln into Glu, the residue that is recognized and cleaved by endoproteinase GluC. Thus, the cleavage product can contain two different components: the

protein polymer with the full length T7 tag and the protein polymer without the first eight residues of the T7 tag. This assumption is supported by the mass differences shown in the MALDI-TOF result (see Supporting Information, Table S2).

Protein Polymers Generated by Intein-Mediated Purification System. Although endoproteinase GluC showed good cleavage specificity for certain sequences and lengths of protein polymers,[15] polydispersity remained an issue when treating other protein polymers (e.g., the two-peak pattern for CR15−6). However, keeping the C-terminal affinity tag attached to the protein polymers leads to very low conjugation efficiency as the His tag can react with the sulfo-SMCC, accelerating the hydrolysis of the heterobifunctional linker during the conjugation step.[15,17] Alternative affinity tags all contain at least one lysine, rendering them unusable with our current conjugation scheme. In consideration of these limitations, an alternative method with better and more consistent results for obtaining monodisperse proteins is required. An intein-mediated purification system was chosen for its simple and highly efficient method of obtaining purified recombinant proteins. Additionally, it does not require an expensive enzyme like the GluC method, and it decreases the number of overall steps in the purification protocol.

A commercially available intein vector from New England Biolabs can be used to fuse an intein and chitin binding domain (CBD) tag to the target protein. During affinity purification with chitin beads, a reducing reagent such as dithiothreitol (DTT) induces specific self-cleavage at the C-terminus of the first cysteine in the protein-tag junction, releasing the target protein from the chitin-bound intein tag. The absence of cysteine in our drag-tag sequences ensures high specificity and efficiency of DTT-induced self-cleavage. With the two steps of affinity purification and tag cleavage combined into one, the intein-mediated purification system minimizes the potential for material loss or protein degradation that may occur with the previously described method.

For our applications, the C-terminal intein-tag vector pTXB1 was chosen to prevent polydispersity caused by premature protein truncation. The vector was adapted as described earlier to the controlled cloning system. Gene *R15−6* was inserted into MpTXB1 and expressed in *E. coli* BLR(DE3) cells with an induction temperature of 16 °C for 20 h. The new protein (182 amino acids) expressed and purified in the intein system was designated IR15−6. No T7-tag was included in IR15−6 as it did not show a significant improvement in expression yield (data not shown). A lower concentration of NaCl in the column buffer was used as well as an additional 2 h mixing step between the cell extract and prepared chitin beads at 4 °C for better binding. Overnight on-column cleavage was conducted at 25 °C with 50 mM DTT to achieve optimal cleavage results. Successful on-column cleavage and tag-free target protein was confirmed by SDS-PAGE, but visible amounts of coeluted intein-CBD tag could still be observed with target protein polymers in the elution fractions (see Supporting Information, Figure S6). Preparative RP-HPLC on a C18 column was used as a second purification step to remove hydrophobic impurities from the hydrophilic protein polymers (see Supporting Information, Figure S7). The same intein expression and purification strategy was applied to another gene, *R18−3*, and the resultant 110-aa protein polymer was designated IR18−3. MALDI-TOF confirmed the molecular masses of these two proteins after secondary purification by RP-HPLC (Figure 4). About 2−5 mg of protein polymer was obtained from 1 L of expression culture after RP-HPLC purification.

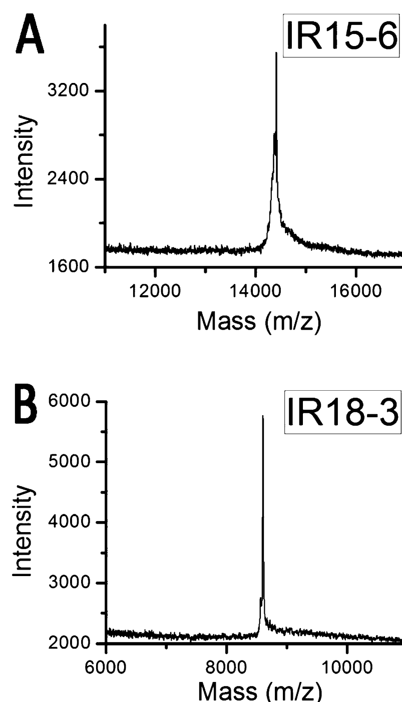


Figure 4. MALDI-TOF results for intein-purified protein polymer drag-tags. (A) MALDI-TOF result for IR15−6. The observed mass is 14416.13 Da, whereas the expected mass for IR15−6 is 14416.63 Da. (B) MALDI-TOF result for IR18−3. The observed mass is 8618.92 Da whereas the expected mass for CR15−6 with His tag is 8614.72 Da.

Characterization of Protein Polymers Generated from Intein-Mediated Purification System by FSCE. The tag-free protein polymer drag-tags IR15−6 and IR18−3 were conjugated to a 30-nt ssDNA oligomer and then analyzed by FSCE. Figure 5 shows a mostly clean single peak with a strong signal for both DNA drag-tag bioconjugates, demonstrating protein monodispersity and efficient conjugation reactions. Both of these attributes are key to successful FSCE DNA sequencing Due to the smaller size and lower number of arginine residues (only 6 Arg in IR18−3), IR18−3 provides less drag than IR15−6. Although IR15−6 is slightly smaller than GluC-cleaved CR15−6 due to the absence of a T7-tag at the N-terminus, it shows larger effective drag than CR15−6 (Figure 3) in FSCE separations.. The greater the difference in elution times between the unconjugated DNA and conjugated DNA, the greater the drag of the attached protein polymer. The increased drag of IR15−6 compared to CR15−6 is likely due to the absence of the negatively charged C-terminal Glu residue required for the enzymatic removal method.

FSCE Sequencing with an Intein-Purified Protein Polymer. Sanger sequencing reactions were performed using either IR18−3 or IR15−6 conjugated to the sequencing primer prior to the extension reaction. The sequencing fragments were separated by free-solution electrophoresis without any entangled polymer network present. The raw electropherogram (without corrections that normalize peak heights or mobility shifts induced by different dyes) for sequencing with the IR18−3 drag-tag is shown in Figure 6. Unlike traditional CGE separations, in FSCE, the smaller fragments migrate slower

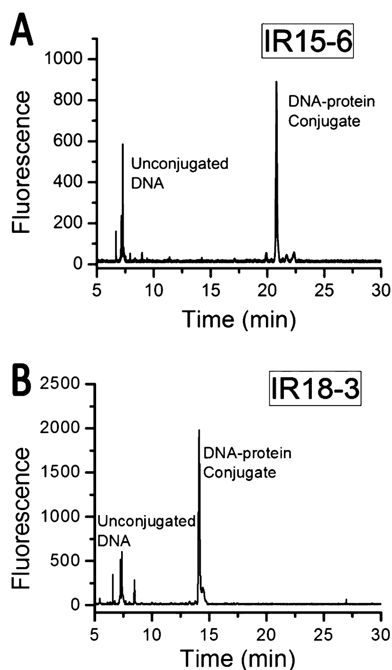


Figure 5. (A) Free-solution capillary electrophoresis of DNA drag-tag conjugate for IR15−6 (182 amino acids, 12 Arg) using a 30-base ssDNA oligomer. (B) Free-solution capillary electrophoresis of DNA drag-tag conjugate for IR18−3 (110 amino acids, 6 Arg) using a 30-base ssDNA oligomer. ABI 3100, 36 cm array with 50 μM ID, 1X TTE, 7 M urea, 0.5%v/v POP-6 (for dynamic coating), 1 kV/20s injection, 310 V/cm, 55 °C.

than the larger ones; thus, the sequencing data is "read" starting from the lower right corner of the figure and going "backwards" to the upper left corner. When using the known sequence of the M13mp18 template for alignment, the sequencing data obtained by IR18−3 can be read easily to approximately 125 bases, with the exception of a dye blob in the "A" channel due to copurification of the ddATP with the cationic drag-tag that obscures the sequence from 95 to 100 bases. Between about 120 and 180 bases, identification of the sequence cannot be done by eye due to slight differences in the mobility of the four different dye terminators and repeated peaks that are unresolved; the observed peaks, however, can be aligned with the known sequence of the template to at least 180 bases. For this drag-tag to be used to sequence an unknown sequence, more advanced data processing software should be developed. The read length is comparable to the sequencing data generated by the 127-aa drag-tag, a protein polymer drag-tag larger in size but with only two Arg (+1 net positive charge after conjugation at N-terminus).[16] This result further confirmed that deliberately introducing some positive charges into the drag-tag sequence is another way to obtain higher drag for longer sequencing read lengths in FSCE. A detailed peak analysis combined with an analysis of FSCE theory using charged drag-tags will be discussed in an upcoming paper. Although IR18−3 showed a clean single peak when characterized by FSCE, there was a minor peak present in addition to the main peak in the single-base extension[16] test (see Supporting Information, Figure S8), which may have caused increased peak widths in the sequencing separation.

No sequencing data could be generated by the IR15−6-conjugated sequencing primer. Similar to the 390-aa and 516-aa protein polymers from the previous family of protein polymer

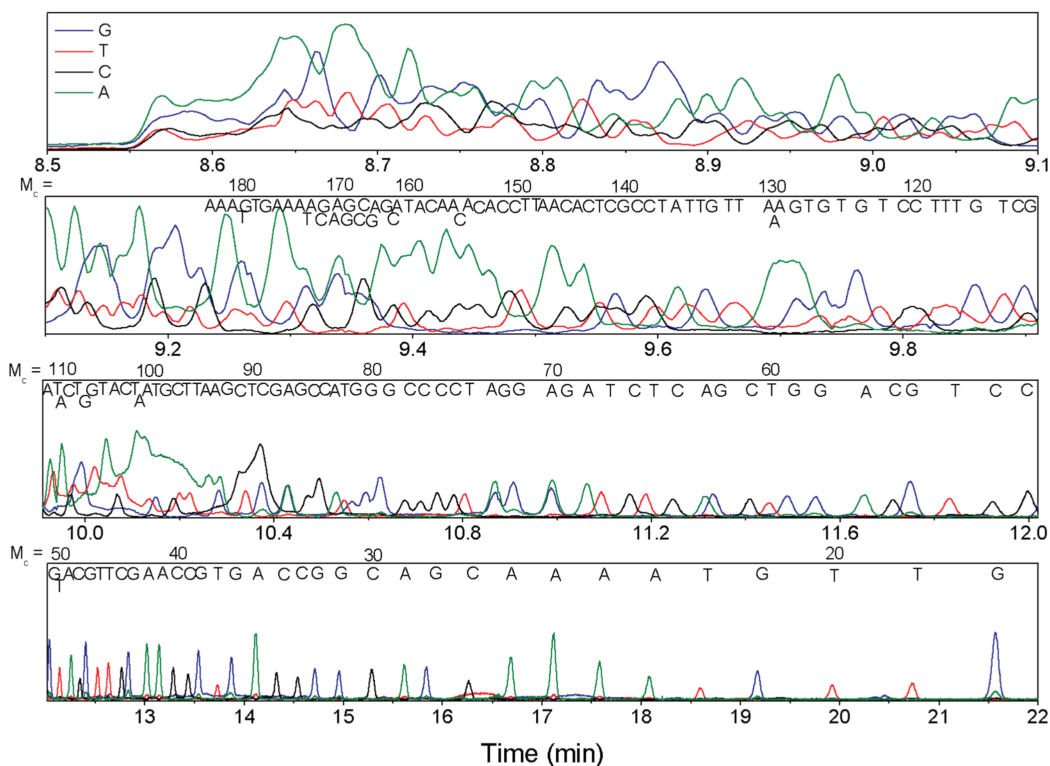


Figure 6. Sequencing separation of Sanger fragments coupled to the IR18−3 (110 aa, 6 Arg) protein drag-tag. The shortest fragments elute last, therefore the sequence is "read" backward from the right side of the bottom panel., ABI 3100, 36 cm array with 50 μM ID, 1X TTE, 7 M urea, 0.5% v/v POP-6 (for dynamic coating), 1 kV/20s injection, 287 V/cm, 55 °C.

drag-tags, enzymatic extension reactions were somehow inhibited by the presence of the IR15−6 (182-aa) drag-tag on the 5′ end of the sequencing primer. The 182-aa long IR15−6, with 12 positively charged arginine residues, may have enough charge to interact strongly with either the sequencing primer or the ssDNA M13mp18 template. An alternative conjugation method for appending protein drag-tags to Sanger fragments would avoid the inhibition described above, and is currently under investigation. By using this new post-Sanger conjugation strategy, a much longer read length (>400 bases) is expected by utilizing drag-tags with greater size and more positive charges.

■ CONCLUSIONS

We demonstrated a one-step purification method that combines affinity chromatography and on-column tag cleavage to generate tag-free protein polymers with improved level of monodispersity for Free-Solution Conjugate Electrophoresis (FSCE) separations. Two "highly" charged protein polymers, IR15−6 (182 aa, 12 Args) and IR18−3 (110 aa, 6 Args), were produced and found to be essentially monodisperse when analyzed by FSCE. These proteins are smaller in size yet had comparable hydrodynamic drag to previous nearly neutral protein polymer drag-tags, which indicates that besides increasing the sizes of proteins, intentional inclusion of positive charges in drag-tag designs could be another strategy for obtaining larger hydrodynamic drag and thus improved separations of longer DNA sequencing fragments. The first sequencing read by FSCE with a "highly" positively charged protein polymer drag-tag was obtained with IR18−3 and described herein; about 180 bases of sequencing read length was achieved with a known sequence. The IR15−6 protein polymer inhibited the enzymatic amplification reaction, indicating that the current method of using preconjugated, drag-tagged sequencing primers will not work when the net charge of the drag-tag is over a certain limit. A new, post-Sanger conjugation reaction method is currently under investigation along with the production of more "highly" positively charged protein polymer drag-tags with different lengths and net charges. Longer read length FSCE sequencing is expected with the utilization of larger and more charged protein polymer drag-tags in the near future. Additionally, the fundamental theory of electrophoresis of drag-tagged DNA will be further developed and tested for the first time with a non-neutral drag-tag using data from these "highly" charged drag-tags in an upcoming paper.

■ ASSOCIATED CONTENT

Supporting Information

Experimental details of oligonucleotide sequences, SDS-PAGE results, RP-HPLC purifications, single base extension test, and additional MALDI-TOF spectra. 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.

■ ACKNOWLEDGMENTS

We thankfully acknowledge Stanford University's Protein and Nucleic Acid Facility for the use of their MALDI-TOF instrument. Additionally, we acknowledge Prof. Gary Slater, Dr. Mytyta Chubynsky, and Dr. Samantha Desmarais for their assistance with the manuscript. This work was funded by the NIH R01 HG002918-01.

■ REFERENCES

(1) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463−5467.
(2) Mayer, P.; Slater, G. W.; Drouin, G. *Anal. Chem.* **1994**, *66*, 1777−1780.
(3) 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.
(4) Won, J. I.; Meagher, R. J.; Barron, A. E. *Electrophoresis* **2005**, *26*, 2138−48.
(5) Vreeland, W. N.; Desruisseaux, C.; Karger, A. E.; Drouin, G.; Slater, G. W.; Barron, A. E. *Anal. Chem.* **2001**, 73, 1795−1803.
(6) Haynes, R. D.; Meagher, R. J.; Barron, A. E. *Biopolymers* **2011**, *96*, 702−707.
(7) Haynes, R. D.; Meagher, R. J.; Won, J. I.; Bogdan, F. M.; Barron, A. E. *Bioconjugate Chem.* **2005**, *16*, 929−938.
(8) Ren, H.; Karger, A. E.; Oaks, F.; Menchen, S.; Slater, G. W.; Drouin, G. *Electrophoresis* **1999**, *20*, 2501−2509.
(9) Huang, J.; Foo, C. W. P.; Kaplan, D. L. *Polym. Rev.* **2007**, *47*, 29−62.
(10) Prince, J. T.; McGrath, K. P.; Digirolamo, C. M.; Kaplan, D. L. *Biochemistry* **1995**, *34*, 10879−10885.
(11) Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2009**, *10*, 229−236.
(12) Simnick, A. J.; Lim, D. W.; Chow, D.; Chilkoti, A. *Polym. Rev.* **2007**, *47*, 121−154.
(13) Farmer, R. S.; Argust, L. M.; Sharp, J. D.; Kiick, K. L. *Macromolecules* **2006**, *39*, 162−170.
(14) Farmer, R. S.; Top, A.; Argust, L. M.; Liu, S.; Kiick, K. L. *Pharm. Res.* **2008**, *25*, 700−708.
(15) Lin, J. S.; Albrecht, J. C.; Meagher, R. J.; Wang, X.; Barron, A. E. *Biomacromolecules* **2011**, *12*, 2275−2284.
(16) Meagher, R. J.; Won, J. I.; Coyne, J. A.; Lin, J.; Barron, A. E. *Anal. Chem.* **2008**, *80*, 2842−2848.
(17) Cuatreca., P; Parikh, I. *Biochemistry* **1972**, *11*, 2291−2299.
(18) Mattson, G.; Conklin, E.; Desai, S.; Nielander, G.; Savage, M. D.; Morgensen, S. *Mol. Biol. Rep.* **1993**, *17*, 167−183.
(19) Albrecht, J. C.; Lin, J. S.; Barron, A. E. *Anal. Chem.* **2011**, *83*, 509−515.
(20) Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. *Nat. Biotechnol.* **1999**, *17*, 889−92.
(21) Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M. Q. *Gene* **1997**, *192*, 271−81.
(22) Won, J. I.; Barron, A. E. *Macromolecules* **2002**, *35*, 8281−8287.
(23) 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.