

Jong-In Won*
Robert J. Meagher
Annelise E. Barron

Department of Chemical and
Biological Engineering,
Northwestern University,
Evanston, IL, USA

Protein polymer drag-tags for DNA separations by end-labeled free-solution electrophoresis

We demonstrate the feasibility of end-labeled free-solution electrophoresis (ELFSE) separation of DNA using genetically engineered protein polymers as drag-tags. Protein polymers are promising candidates for ELFSE drag-tags because their sequences and lengths are controllable not only to generate monodisperse polymers with high frictional drag, but also to meet other drag-tag requirements for high-resolution separations by microchannel electrophoresis. A series of repetitive polypeptides was designed, expressed in *Escherichia coli*, and purified. By performing an end-on conjugation of the protein polymers to a fluorescently labeled DNA oligomer (22 bases) and analyzing the electrophoretic mobilities of the conjugate molecules by free-solution capillary electrophoresis (CE), effects of the size and charge of the protein polymer drag-tags were investigated. In addition, the electrophoretic behavior of bioconjugates comprising relatively long DNA fragments (108 and 208 bases) and attached to uncharged drag-tags was observed, by conjugating fluorescently labeled polymerase chain reaction (PCR) products to charge-neutral protein polymers, and analyzing *via* CE. We calculated the amount of friction generated by the various drag-tags, and estimated the potential read-lengths that could be obtained if these drag-tags were used for DNA sequencing in our current system. The results of these studies indicate that larger and uncharged drag-tags will have the best DNA-resolving capability for ELFSE separations, and that theoretically, up to 233 DNA bases could be sequenced using one of the protein polymer drag-tags we produced, which is electrostatically neutral with a chain length of 337 amino acids. We also show that denatured (unfolded) polypeptide chains impose much greater frictional drag per unit molecular weight than folded proteins, such as streptavidin, which has been used as a drag-tag before.

Keywords: Bioconjugation / Capillary electrophoresis / End-labeled free-solution electrophoresis / Monodisperse / Polydisperse / Protein polymer drag-tag DOI 10.1002/elps.200410042

1 Introduction

DNA separation by electrophoresis generally requires the use of a polymer matrix (*e.g.*, a gel or an entangled polymer solution) because the electrophoretic mobilities of DNA molecules in free solution do not normally depend on chain length to a useful degree [1]. However, it was first suggested in the early 1990s that the constant charge-to-friction ratio of DNA that prevents its free-solution separation by electrophoresis could be overcome if DNA fragments were attached to a perturbing entity [2, 3]. This approach is known as end-labeled free-solution electrophoresis (ELFSE) [4], and this perturbing entity can be

called a “drag-tag” [5]. Since the time that the ELFSE concept was proposed, and the theoretical aspects of this concept were examined [4, 6–8], researchers have shown that this bioconjugate approach is potentially useful for the separation of short oligonucleotides [5, 9–11] as well as long dsDNA fragments [12]. ELFSE is a particularly promising method for DNA analysis requiring a high-resolution, size-based separation of relatively small DNA fragments, such as for genotyping or sequencing, by either capillary or microfluidic chip electrophoresis. A significant advantage of ELFSE is that it eliminates the need for loading viscous polymer networks into electrophoresis microchannels, since the separation of DNA conjugates can be performed in an aqueous buffer with a viscosity of ~ 1 cP. However, to accomplish DNA separations with high performance in free solution, ELFSE requires totally monodisperse drag-tags, which create a suitable amount of frictional drag when pulled behind DNA, and which have other properties appropriate for

Correspondence: Professor Annelise E. Barron, Northwestern University, Department of Chemical and Biological Engineering, 2145 Sheridan Road, Rm. E136, Evanston, IL 60208, USA
E-mail: a-barron@northwestern.edu
Fax: +847-491-3728

Abbreviations: aa, amino acid; DEAE, diethylaminoethyl; ELFSE, end-labeled free-solution electrophoresis; IPTG, isopropylthiogalactoside; sulfo-SMCC, sulfosuccinimidyl 4-(*N*-maleimidomethyl)-1-cyclohexane carboxylate

* Present address: Department of Chemical Engineering, Hongik University, Seoul, South Korea

microchannel electrophoresis, such as water-solubility, charge-neutrality, and a minimal tendency for nonspecific interaction with the walls of the microchannel [13, 14].

Three different classes of molecules have been considered as potential ELFSE drag-tags, including (i) chemically synthesized polymers [15]; (ii) oligo-*N*-substituted glycines (oligopeptoids) produced on an automated peptide synthesizer [5, 11]; and (iii) natural proteins [12, 16]. Chemically synthesized polymers, such as polyethylene glycol (PEG), are not suitable as drag-tags for high-resolution DNA separations, because they are inevitably polydisperse in terms of molecular weight distribution [15]; even a polydispersity index (M_w/M_n) of 1.01 is too large for sequencing or genotyping applications. Oligopeptoids, produced using an automated peptide synthesizer and purified to monodispersity by HPLC, can be useful for the separation of small oligonucleotides, but are not applicable to DNA sequencing with long reads because they are too small. In order to separate large DNA fragments (e.g., > 150 bases), a drag-tag with large frictional drag is essential, but there is a limit to the practical usefulness of organic synthetic methods for the production of monodisperse, long polypeptides. However, solid-phase synthesis is certainly useful for the synthesis of short polypeptides or peptoids (e.g., shorter than ~ 60-mer), and ELFSE using short hydrophilic drag-tags (20-, 40-, and 60-mer) was successfully implemented for the separation of DNA oligonucleotides (20- and 21-base ssDNA) [5] as well as for genotyping the oligonucleotide products of a single-base extension (SBE) reaction between 19 and 24 bases in length [11].

Natural proteins might seem to be excellent candidates for sequencing drag-tags, because they can be relatively large, monodisperse, and water-soluble. In fact, the first, and only demonstration of DNA sequencing by ELFSE in microchannels was performed by Ren and co-workers [16] using a recombinant streptavidin protein as the drag-tag. Streptavidin is a compactly folded homo-tetramer of 159-residue polypeptide chains, and can be attached easily to biotinylated DNA, because of its strong binding to biotin. With the use of a suitable wall-coating agent (e.g., POP6) and gel-purified streptavidin (attached to biotinylated sequencing fragments after the Sanger reaction), they were able to sequence up to ~ 110 bases in 18 min. Although these remarkable results provided the first demonstration that ELFSE could be used for DNA sequencing, they also highlighted some of the limitations of natural proteins as sequencing drag-tags. Obviously, the DNA sequencing read-length obtained with streptavidin-based ELFSE was not competitive with that of conventional gel or polymer matrix-based electrophoresis (the typical read-length of matrix-based electrophoresis

with an ABI PRISM 3730™ is ~ 700 bases at 99% accuracy in 1 h [17]). According to the analyses of Ren *et al.*, analyte-wall interactions were responsible for most of the decrease in peak resolution seen with ELFSE under high electric fields; it was suggested that these interactions could be minimized by the use of a specially designed, less adsorptive drag-tag. Although streptavidin is water-soluble and its charge is nearly zero at pH 8–9, local electrostatic interactions can occur during free-solution electrophoresis, because numerous positively and negatively charged amino acids are distributed on the surface of the protein. Hence, a natural water-soluble protein such as streptavidin can be said to be sub-optimal for use as a drag-tag molecule, because it will likely present a chemically “patchy” (charged and/or hydrophobic) surface to the capillary wall, which could facilitate its adsorption to the wall, leading to band-broadening. If, instead, a non-natural polypeptide composed of only hydrophilic and uncharged amino acids is used as a drag-tag, it may allow for better ELFSE performance. Additionally, and perhaps most importantly, streptavidin gives a relatively small α value (where α is the total hydrodynamic drag created by the drag-tag, in units of the drag created by one base of ssDNA), due to its globular, folded conformation and charge distribution (the measured α value of streptavidin is ~ 30 [16]). The obvious way of separating larger DNA fragments is to use a larger friction-generating label for ELFSE, because larger α values will increase the resolution of ELFSE separation. Based on the physical equations they derived as well as experimental data, Ren *et al.* [16] predicted that a read-length of about 625 bases could be obtained with $\alpha = 300$, $E = 1000$ V/cm, and L (channel length) = 40 cm; more recently, Slater *et al.* (personal communication) have predicted that 1300 bases could be sequenced in 700 s (< 12 min), under optimized conditions, with a drag-tag having $\alpha = 400$, $E = 2000$ V/cm, and $L = 20$ cm. Many challenges remain to be overcome to fulfill this exciting prediction, the first and most important of which is the creation of a monodisperse drag-tag with large α .

In this paper, we demonstrate the feasibility of ELFSE separation of DNA using non-natural, protein-based polymers (or “protein polymers”) as drag-tags. Protein polymers are highly repetitive polypeptides provided by genetic engineering technology [18]. They offer significant advantages for this application over natural proteins, in that their sequences and lengths are controllable not only to generate monodispersity and high frictional drag, but also to meet other drag-tag requirements for microchannel electrophoresis. We designed, produced, and purified a series of different repetitive proteins in *Escherichia coli*. By attaching the protein polymers to a 22-base oligonucleotide and analyzing the electrophoretic mobili-

ties of the conjugate molecules by free-solution capillary electrophoresis, the effects on α of varying the size and charge of the protein polymer drag-tags were explored. To study the effects of charge in drag-tags for ELFSE, the overall charge of the protein polymer sequence was modified by including either a negatively charged glutamic acid residue [(GAGQGEA)_nG] or a neutral serine residue [(GAGQGSA)_nG] in the 6th position out of 7 in a repeating sequence, and investigating the electrophoretic mobilities of these differently charged drag-tags. To determine the effect of drag-tag size, the electrophoretic mobilities of DNA-protein polymer conjugates were investigated, using charge-neutral protein polymers of three different sizes ranging from 85 to 337 amino acids, and three different negatively charged protein polymers ranging from 127 to 505 amino acids. In addition, the effect of including a short polyhistidine tag (His-tag) in the protein polymers on bioconjugate peak mobility and peak shape during electrophoresis was investigated, by comparing electropherograms obtained with fluorescently labeled protein polymers with and without the His-tag. Since both positively charged and negatively charged amino acids are found in the His-tag, this analysis provides some insight into how the electrophoretic behavior of a DNA-protein bioconjugate is affected if electrostatic interactions between the protein and the capillary wall (or between the protein and itself or another protein) can occur.

One specific aim of this work was to observe and quantify the α values of the obtained drag-tags as a function of polypeptide chain length and charge, and to estimate the potential read-lengths that could be obtained if these drag-tags were used for DNA sequencing in our current system. To our knowledge, these are the first examples of ELFSE analysis using genetically engineered protein polymer drag-tags. Our results show that the protein polymers are promising candidates for ELFSE drag-tags. The ELFSE technique should provide an attractive alternative to matrix-based DNA sequencing, if even more homogeneous and higher-friction protein polymer drag-tags are developed. (Note that we do not present DNA sequencing separations in this paper, as other challenges remain in applying these protein polymers to this demanding application.)

2 Materials and methods

2.1 Gene design, synthesis, and multimerization

Two single-stranded, synthetic oligonucleotides (102 bases each; see Fig. 1), which encode three tandem repeats of -(Gly-Ala-Gly-Gln-Gly-Ser-Ala)- or -(Gly-Ala-

Gly-Gln-Gly-Glu-Ala)-, were amplified *via* PCR with two corresponding oligonucleotide primers [5'-ATATAGAA-TTCCTCTTCAGGT-3'; 5'-TATATGAATTCCTCTTCTACC-3'], which were obtained from the Northwestern University Biotechnology Laboratory (Chicago, IL, USA). This reaction was performed using an MJ Research DNA Thermal Cycler with 30 cycles of 95°C for 1 min (denaturing), 55°C for 1 min (annealing), and 72°C for 2 min (elongation). A high concentration of dsDNAs was obtained from the PCR, and subsequently digested with *Eam*1104 I endonuclease (Stratagene, La Jolla, CA, USA) for 12 h at 37°C to produce DNA monomers with cohesive termini. The 63-bp DNA monomers were then fractionated and recovered *via* 3% agarose gel electrophoresis onto a diethylaminoethyl (DEAE)-cellulose membrane. The isolated DNA monomers were purified by standard phenol/chloroform extraction and ethanol precipitation. Higher-order multimers were constructed by the previously reported strategy developed in our laboratory, which we have termed the "controlled cloning method" [19], and genes encoding 4, 8, and 16 repeats of Gene 1 (252, 504, 1008 bp, respectively) and 6, 12, and 24 repeats of Gene 2 (378, 756, 1512 bp, respectively) were obtained.

2.2 Protein expression and purification

A pUC18 plasmid (Invitrogen, Carlsbad, CA, USA) with an ampicillin resistance gene was used as a cloning vector, and a pET-19b plasmid (Novagen, Madison, WI, USA) with ampicillin resistance was used for protein expression. Protein expression was performed using a 2-L Erlenmeyer flask. Single colonies of *E. coli* expression strain BLR(DE3)(pET-19b) (Novagen) were used to inoculate a seed culture of Luria broth (LB) medium (50 mL) containing both carbenicillin (50 µg/mL) and tetracycline (30 µg/mL) antibiotics. The seed culture was grown to saturation and used to inoculate 500 mL of the same medium in a 2-L culture flask. The main culture was incubated for several hours at 37°C with agitation at 250 rpm until the OD₆₀₀ reached 0.6–0.8. At that time, isopropyl-β-galactoside (IPTG) was added to a final concentration of 1 mM to initiate recombinant protein production. Cells were cultured for an additional 3 h, and then harvested by centrifugation at 10 000 × *g* and 4°C for 15 min. The cell paste was resuspended in 30 mL of the lysis buffer (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride; pH 7.8). The cells were lysed by several freeze (–80°C)/thaw (37°C) cycles and subsequently sonicated at 30 s intervals for 3 min. The lysate was centrifuged at 10 000 × *g* and 4°C for 30 min to pellet the cell debris, and the supernatant was loaded onto Ni-chelating resin (ProBond; Invitrogen). The target proteins were selectively eluted at low pH (pH 4.0) after the endogenous host

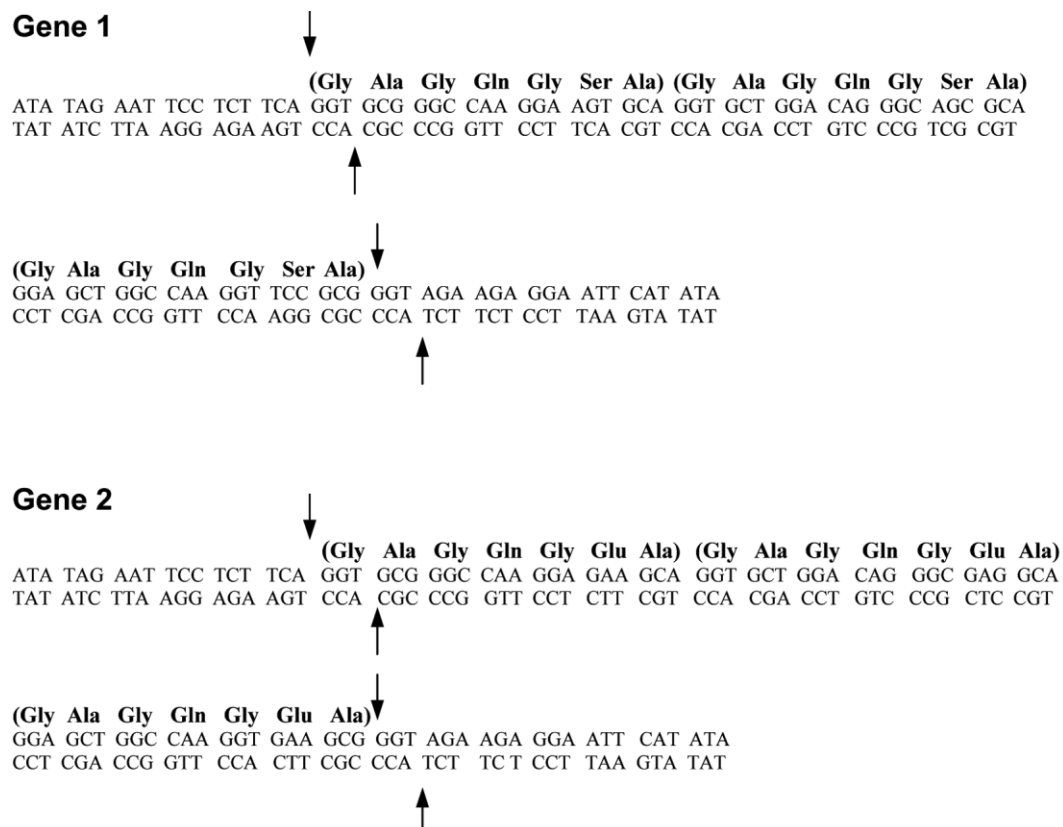


Figure 1. DNA sequences (Gene 1 and Gene 2) of the designed genes, and the amino acid repeat that the central gene region encodes.

proteins had been washed off the resin extensively with higher pH buffers (pH 7.8, pH 6.0, and pH 5.3 buffers). The eluents were analyzed by 12% discontinuous SDS-PAGE. Purified proteins were dialyzed against deionized water for three days, and lyophilized to a dry powder.

2.3 Cyanogen bromide cleavage reaction and post-purification

The fusion polypeptides (1 mg each) were dissolved in 70% formic acid/water (Aldrich, Milwaukee, WI, USA) (final concentration of protein \approx 1 mg/mL), and each solution was purged with N₂ gas for 3 min to remove dissolved oxygen. Cyanogen bromide (Sigma, St. Louis, MO, USA) was added to the mixture at a final concentration of about 5 mg/mL. The reaction mixture was purged again with N₂ gas for 5 min. The reaction tubes were capped, covered with aluminum foil, and stirred at room temperature. Each sample was reacted for 3 h. After reaction, the volatiles, cyanogen bromide, and the N-terminal oligopeptide tag were removed by centrifugal ultrafiltration using a Microcon filter with an appropriate molecular weight cutoff (Millipore, Billerica, MA, USA) and

by several subsequent washes with distilled, deionized water. The cleavage products were lyophilized to a dry powder. The cleavage products were resuspended in 500 μ L of native protein purification buffer (50 mM NaH₂PO₄, 500 mM NaCl, pH 8.0). To separate the target protein (*i.e.*, a protein polymer without a His-tag) from the residual, uncleaved fusion protein (with His-tag) completely, each protein mixture was mixed with 0.5 mL of Ni-chelating resin (Probond; Invitrogen) and centrifuged at 1000 \times g for 5 min. After collecting the column flow-through solution, salts in the solution were removed by centrifugal ultrafiltration using a Microcon filter with an appropriate molecular weight cutoff.

2.4 Protein analysis and characterization

The purified proteins were dissolved in water and analyzed by gradient reversed-phase HPLC on C18 packing (Vydac, 5 μ m, 300 Å, 2.1 \times 250 mm). A linear gradient of 0–95% B in A was run over 50 min at a flow rate of 0.1 mL/min (solvent A = water with 0.1% TFA, solvent B = acetonitrile with 0.1% TFA) at 60°C; peaks were detected by UV absorbance at 220 nm. A Voyager-DE™ PRO mass spec-

trometer (Analytical Services Laboratory, Northwestern University) was used for MALDI-TOF analysis, and sinapinic acid was used as the matrix. Circular dichroism spectra were recorded on a Jasco J-715 spectropolarimeter (Keck Biophysics Facility, Northwestern University). Protein samples were dissolved in sterile water at a concentration of 1 mg/mL. Spectra were obtained from 260 to 190 nm as a scanning range, using the average of 40 consecutive data accumulations.

2.5 Preparation of the DNA sample and conjugation of the protein polymers to DNA

A 22-base M13 oligonucleotide primer [5'- X_1 GG X_2 -TTTCCCAGTCACGACGTTG-3'] (Oligos, Etc., Wilsonville, OR, USA) was prepared for conjugation to the protein polymers. The DNA oligomers included an Oregon Green fluorescent label on the 4th residue (X_2) from the 5'-terminus to allow for laser-induced fluorescence (LIF) detection, and a thiol (-SH) functionality on the 5'-terminus (X_1) to allow for covalent conjugation with the maleimide functionality introduced at the amino terminus of the polypeptides. Thiol group reduction and conjugation of the protein polymers to a DNA oligomer were performed by procedures described in a previous report [20].

2.6 FITC labeling of the protein polymers

Protein polymers which were not conjugated to DNA were instead fluorescently labeled by reaction with fluorescein-5-isothiocyanate (FITC) (Acros Organics, Morris Plains, NJ, USA) in an aqueous solution containing 0.1 M sodium carbonate, pH 9.0, for 1 h at room temperature with continuous mixing. A protein concentration of roughly 10 mg/mL and the same molar ratio of the FITC was used. The unreacted FITC was then removed by centrifugal ultrafiltration using Microcon filtration devices (Millipore) or by Centri-Sep columns (Princeton Separations, Princeton, NJ, USA).

2.7 Preparation of PCR products and denaturing of the bioconjugates

A 30-base M13 oligonucleotide [5'- X_1 CC X_2 TTTAGGG-TTTCCCAGTCACGACGTTG-3'] (Oligos, Etc.) was prepared as a forward primer for PCR amplification. This forward primer includes an Oregon Green fluorescent label on the 4th residue (X_2) from the 5'-terminus, and a thiol (-SH) functionality on the 5'-terminus (X_1). Two different 22-base oligonucleotides [5'-ACGAATTCGAGCTCGGTACCC-3'; 5'-TAGGCACCCCAGGCTTTACAC-3'], which were purchased from IDT (Coralville, IA, USA) and served as M13

reverse primers, were prepared for generating PCR products of 108 and 208 bp in length. M13mp18 ssDNA, which was obtained from a Thermo Sequenase Dye Terminator Cycle Sequencing Kit (Amersham Biosciences, Piscataway, NJ, USA), was used as the template. PCR reaction and DNA purification were performed using the same protocol described above. The sizes of the two PCR products (dsDNA) were confirmed *via* 2.5% agarose gel electrophoresis. Thiol group reduction and the conjugation of the protein polymers to the PCR products were performed by the same procedures described above (Section 2.5). Single-stranded DNA(ssDNA)-polypeptide conjugates were prepared by heating the PCR products (dsDNA)-polypeptide conjugates for 3 min at 95°C in formamide solution, and consecutive snap-cooling on ice.

2.8 Electrophoretic analysis of the conjugate molecules

The polypeptide-oligonucleotide conjugates or FITC-labeled polypeptides were analyzed *via* capillary electrophoresis (CE). CE separations were conducted at 55°C in 25 μ m ID fused-silica capillaries with various lengths (Polymicro Technologies, Phoenix, AZ, USA) filled with a 1 \times TTE, 7 M urea buffer (50 mM Tris, 50 mM TAPS, 2 mM EDTA, 7 M urea, pH 8.4). 7 M urea and a temperature of 55°C was used to keep the protein polymers and DNA in an unstructured state. The internal surface of the capillary was coated with an adsorbed layer of POP5 polymer (Applied Biosystems, Foster City, CA, USA), using a low-viscosity, 3% v/v aqueous dilution of the commercially available POP5 solution, to reduce electroosmotic flow to negligible levels. All analyses were performed with a Bio-Focus 3000 Capillary Electrophoresis system equipped with a laser-induced fluorescence (LIF) detector (BioRad, Hercules, CA, USA). Immediately prior to sample injection, the injection end of the capillary was briefly dipped into pure water to remove any residual buffer salts on the outer surface of the capillary. The samples were introduced into the capillary by a pressure injection, with pressure-time constants of 20–60 psi·s. Electrophoresis was conducted at 400 V/cm until all peaks had eluted. Detection of the analytes was accomplished by excitation of the fluorescent label using the 488 nm line of an argon-ion laser, with emission detected at 520 nm.

3 Results and discussion

3.1 Gene construction and multimerization

Two 102-base synthetic oligonucleotides (ssDNA), which encode three tandem repeats of GAGQGS or GAGQ-GEA (Fig. 1), were designed and PCR-amplified with two

primers to generate a large amount of dsDNA. The coding sequences of the gene (shown between the arrows) were flanked by two *Eam*1104 I recognition sites, the cleavage sites of which are indicated by arrows in Fig. 1. After PCR amplification, the 102-bp genes were digested with *Eam*1104 I restriction enzyme to generate 63-bp DNA "monomers". These DNA fragments were isolated from the reaction mixture by agarose gel electrophoresis onto a DEAE-cellulose membrane. The isolated monomers were then successfully multimerized through a controlled cloning process we have described before [19], and the final genes encoding 4, 8, and 16 repeats of Gene 1 and 6, 12, and 24 repeats of Gene 2 were constructed. The multimerized genes were sequenced to confirm that they had the correct, expected DNA sequences.

3.2 Protein expression and purification

The proteins of interest were produced in *E. coli* strain BLR(DE3) (Novagen) using a modified pET-19b expression vector (Novagen). Isolation of the target proteins was accomplished by immobilized metal affinity chromatography (IMAC) under denaturing conditions with a pH gradient, and the purity and molecular weight of proteins were characterized and confirmed by SDS-PAGE and MALDI. The purified proteins were dialyzed for three days to remove salts, and then lyophilized. The final fusion polypeptides were obtained in yields of ~ 10–20 mg/L.

3.3 Cyanogen bromide cleavage reaction and post-purification

The fusion polypeptides were reacted with CNBr for 3 h to remove the decahistidine affinity tag. After reaction, the purity and identity of each protein was observed by HPLC and MALDI-TOF analysis. Uncleaved fusion proteins (with His-tag) were still found in the cleavage mixture in amounts of 5–10 weight% (data not shown). To separate the target protein (without a His-tag) from the residual, uncleaved fusion protein (with His-tag) completely, each protein mixture was mixed with Ni-resin and centrifuged. While the uncleaved fusion protein bound to the resin strongly due to its histidine residues, the target protein did not bind to the resin because it has no histidine residues. Thus, virtually pure proteins were obtained by collecting the column flow-through solution. The purity (> 99% in each case) and identity of each target protein was confirmed by HPLC and MALDI-TOF mass spectrum analysis (data not shown). The amino acid sequences and molecular weights of the obtained protein polymers are presented in Table 1. Note that in each molecule's name, the number of amino acids (aa) is given; for example, P1–85 derives from Gene 1 and comprises 85 amino acids.

Table 1. Amino acid sequences and molecular weights of the obtained protein polymers

Name ^{a)}	Amino acid sequence	M.W. (kDa)
P1-85	(GAGQGSA) ₁₂ G	6.42
P1-169	(GAGQGSA) ₂₄ G	12.76
P1-337	(GAGQGSA) ₄₈ G	25.45
P2-127	(GAGQGGEA) ₁₈ G	10.35
P2-253	(GAGQGGEA) ₃₆ G	20.62
P2-505	(GAGQGGEA) ₇₂ G	41.17
P2-127 with His-tag	MG(H) ₁₀ SSGHI(D) ₄ KHM(GAGQGGEA) ₁₈ G	13.25
P2-253 with His-tag	MG(H) ₁₀ SSGHI(D) ₄ KHM(GAGQGGEA) ₃₆ G	23.52
P2-505 with His-tag	MG(H) ₁₀ SSGHI(D) ₄ KHM(GAGQGGEA) ₇₂ G	44.06

a) P1–85 denotes protein polymer sequence motif 1, with a chain length of 85 amino acids, and so on.

3.4 Electrophoretic analysis of DNA-polypeptide conjugates

Each purified, neutral polypeptide (P1-85, -169, and -337) was conjugated end-on, *via* the amino terminus, to the 22-base, fluorescently labeled oligonucleotide using sulfosuccinimidyl 4-(*N*-maleimidomethyl)-1-cyclohexane carboxylate (SMCC) as a coupling reagent [20]. The resulting conjugates were then analyzed by free-solution CE. Electropherograms resulting from the CE analysis of three different DNA-polypeptide conjugates are shown in Fig. 2. Note that this family of polypeptides (P1) has no charged aa residues, and hence these are neutral drag-tags, with the minor exception of the one charge on the carboxylate terminus. The residual, unconjugated DNA peaks elute first around 7.5 min, followed by the elution of DNA-polypeptide conjugate peaks at 13, 18, and 32 min, as the polypeptide chain length increases from 85, to 169, to 377 aa, respectively. As shown in Fig. 2, the elution time of DNA-protein conjugate peaks increases with polypeptide size, for the obvious reason that hydrodynamic drag must scale (by some factor) with drag-tag chain length. An interesting feature of this data set is that the polydispersity of the conjugate peaks is also seen to increase with polypeptide size. As reported previously [20], this polydispersity results from the deamidation of some of the glutamine residues in the polypeptides (to glutamic acids). Also, since a relatively small (22-mer) DNA was conjugated to polypeptides of three different sizes, the polydispersity of the larger polypeptides may be better resolved because of the relatively slow migration of the larger conjugate molecules, and hence longer effective separation length.

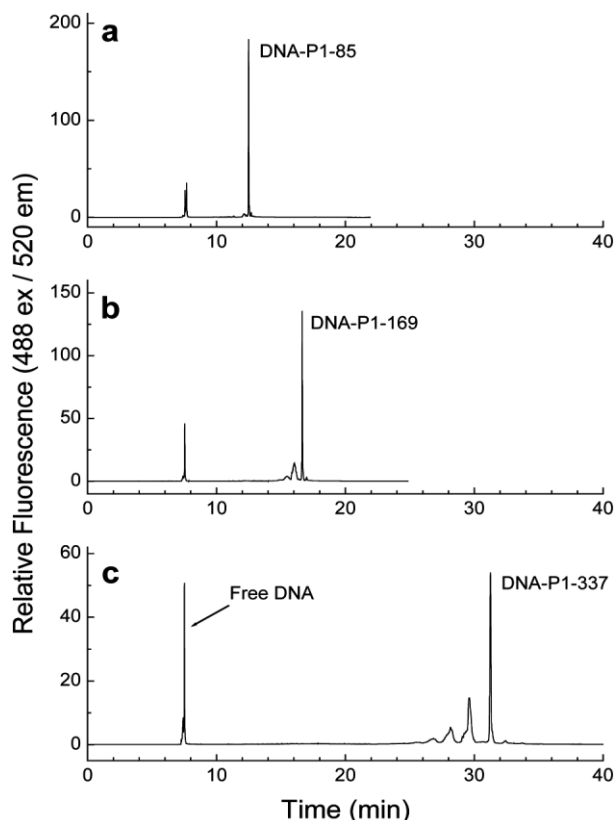


Figure 2. Capillary electropherograms obtained from the analysis of protein-DNA conjugates, with protein samples of P1-85, -169, and -337. The fused silica capillary was 25 μm ID \times 44 cm long (39.6 cm to the detection window), with the inner surface was coated with POP 5 polymer. The running buffer was 50 mM Tris, 50 mM TAPS, 2 mM EDTA with 7 M urea, pH 8.4, with 3% v/v POP5 solution added for wall-coating purposes. Injection was hydrodynamic, with a pressure-time constant of 30 psi \cdot s. The electric field strength was 17.6 kV (400 V/cm), and the temperature was maintained 55°C.

The peak efficiencies of the conjugate peaks were examined in terms of the number of theoretical plates. The efficiencies of the main conjugate peaks were calculated from the electropherograms shown in Fig. 2, and compared with those of DNA-streptavidin conjugates, which were studied by Ren *et al.* [16]. The efficiencies for the DNA-P1-85, -169, and -337 conjugate peaks were 7.7×10^5 , 1.0×10^6 , and 6.3×10^5 theoretical plates, respectively, in a capillary with an effective length of 39.6 cm. Ren *et al.* [16] report the height of a theoretical plate as a function of capillary length or electrophoretic velocity for two sizes of DNA-streptavidin conjugates. For a capillary of similar effective length (40 cm), Ren *et al.* report a peak efficiency of about 5.7×10^5 plates for a 103-base DNA-streptavidin conjugate. In a slightly shorter capillary (34 cm effective length), at electropho-

retic velocities similar to those in the current study, Ren *et al.* observed efficiencies of about 2.5×10^5 plates for a 61-base DNA-streptavidin conjugate. In addition, they also showed that peak width was larger (and thus peak efficiency was lower) for smaller sizes of DNA, probably because these elute more slowly, and thus have more time for diffusion to broaden the peaks. Although direct comparison between the different experimental setups is difficult, it can be said that our results with a small, 22-base DNA and a hydrophilic, uncharged protein polymer drag-tag show substantially better peak efficiencies than were observed for the DNA-streptavidin conjugates. This leads us to conclude that the P1 protein polymer succeeds in meeting the design objective of reduced interaction with capillary walls.

To investigate the electrophoretic behavior of bioconjugates with negatively charged polypeptide drag-tags, three different protein polymers (P2-127, -253, and -505) were also conjugated to the 22-base, fluorescently labeled DNA and analyzed by free-solution CE. As shown in Fig. 3, the peak patterns obtained are similar to those in

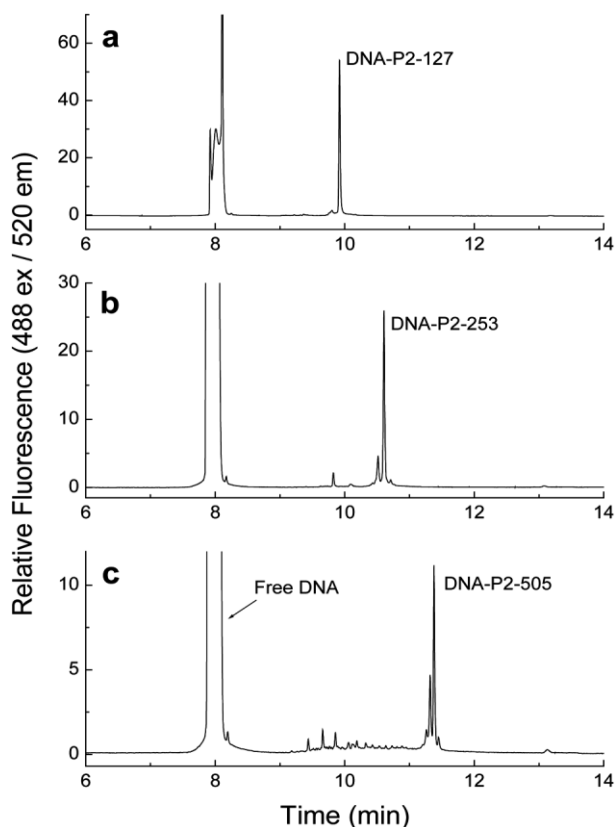


Figure 3. Capillary electropherograms resulting from the analysis of protein-DNA conjugates, with protein samples of P2-127, -253, and -505. Electrophoresis conditions as in Fig. 2.

Fig. 2, in that the polydispersity of the DNA-protein polymer conjugates apparently increases as the polypeptide chain length increases, and that the elution times of the peaks increase with drag-tag chain length. However, the elution times for the conjugate peaks in Fig. 3 are significantly shorter than those in Fig. 2 (ranging from only ~10–12 min despite the large size of these proteins). Remarkably, the DNA-P2–505 conjugate, with a 505-aa drag-tag, shows higher mobility than the DNA-P1–85 conjugate, which has a much shorter, 85-aa drag-tag. This can be explained by the fact that the negatively charged glutamic acid residues, which exist in a ratio of 1 out of every 7 residues in the repeating P2 aa sequence, substantially increase the electrophoretic mobility of the DNA-polypeptide conjugates during electrophoresis, and that this charge-induced increase in net mobility exceeds the decrease in mobility that results from the difference in drag-tag chain length (505 vs. 85 aa). Based on these experimental results, it is clear that negatively charged polypeptides, comprising anionic amino acids in the same ratio as (or higher than) P2 polypeptides, are most likely not suitable as high-performance ELFSE drag-tags, especially for applications (such as DNA sequencing) which require tags with high-effective frictional drag. (Cationic drag-tags would be assumed to be unsuitable as well, since they would be electrostatically attracted to both the capillary walls and the DNA itself.)

3.5 Electrophoretic analysis of fluorescently labeled polypeptides

The negatively charged protein polymers we created (P2-127, -253, and -505) have their own electromotive driving force toward the anode during electrophoresis, and therefore did not in fact need to be conjugated to DNA for CE analysis. To study the electrophoretic mobilities of the protein polymers by themselves, these negatively charged polypeptides were labeled with a FITC fluorescent dye and analyzed *via* free-solution CE. Electropherograms obtained from analyses of the FITC-labeled polypeptides are shown in Fig. 4A. The peak pattern observed in Fig. 4A is similar to that seen in Fig. 3, except that the peaks in Fig. 4A elute a little later due to the absence of DNA. The extent of polydispersity in the three samples is consistent with what is seen in the respective DNA-protein conjugates (Fig. 3), indicating that the polydispersity does not result from the conjugation reaction. The three sizes of P2 migrate with very similar electrophoretic mobilities, indicating a nearly constant scaling of charge and friction for these protein polymers. Note that the addition of the FITC label modifies the net charge of the polypeptide by a total of negative three units (–1 from the reaction of the terminal amine, and –2 from the dye,

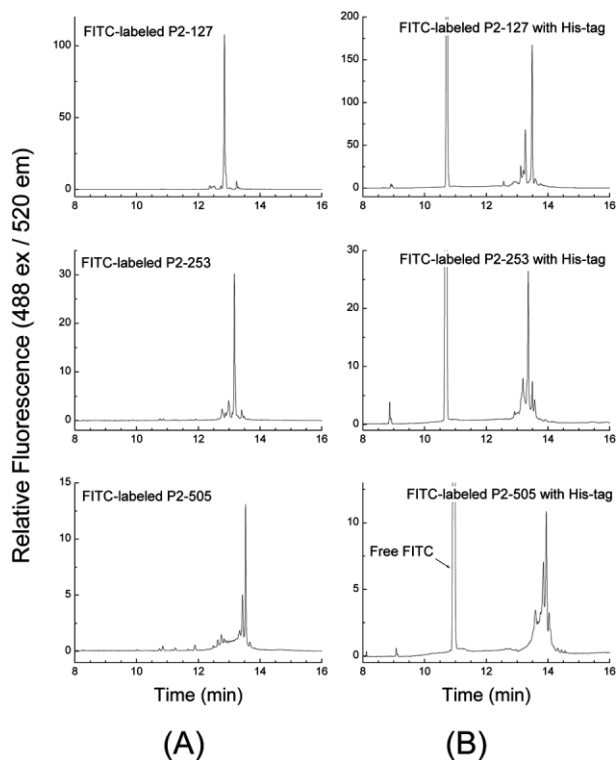


Figure 4. Capillary electropherograms of FITC-labeled P2 polypeptides, (A) without the His-tag and (B) with the His-tag. Electrophoresis conditions as in Fig. 2.

which is a dianion at basic pH). The additional negative charge from the label affects the electrophoretic mobility slightly, and may be responsible for the small variation in mobility between the three sizes of P2.

The influence of the decahistidine fusion tag (His-tag) on the electrophoretic behavior of the P2 protein polymers was investigated by CE analysis of FITC-labeled polypeptides that still included the His-tag. Figure 4B presents ELFSE electropherograms obtained for these FITC-labeled molecules. The His-tag we used is composed of 24 amino acids, and contains several negatively charged amino acids, along with several amino acids that have cationic side groups (see Table 1 for sequence). When the electropherograms in Fig. 4A are compared with those in Fig. 4B, a substantially greater number of unresolved, overlapping peaks is clearly observed in Fig. 4B. This likely results from a combination of effects: (i), either one or two FITC molecules can react with a His-tagged protein polymer, with either the primary amine group of the *N*-terminus or with the one lysine residue in the His-tag; and (ii), an electrostatic interaction may occur between the His-tag and the capillary wall (the latter type of interaction was considered to make a major contribution to the increase in peak broadness and decrease in resolution

observed for a streptavidin drag-tag under high electric fields [16]). In addition, we observe that the mobility of P2–127 with a His-tag is similar to, or even a little lower than, that of P2–253 with the His-tag (Fig. 4B). This feature is presumably caused by the intrinsic drag effect of the His-tag. Since these three different fusion protein polymers of different sizes all carry an identical His-tag on their *N*-termini, and since the *pI* of the His-tag is higher than the *pI* of the P2 protein polymers, the His-tag exerts greater drag on the small fusion protein polymers, because of its greater relative proportion in the fusion peptides as the molecular size of protein polymers decreases.

3.6 Electrophoretic analysis of PCR product-polypeptide conjugates

The electrophoretic behavior of bioconjugates comprising relatively long DNA fragments, attached to the various uncharged drag-tags, was observed by conjugating PCR products of two different lengths (108 and 208 bp) to the neutral protein polymers (P1–85, –169, and –337), and analyzing *via* CE. To create these DNAs, PCR amplification of M13 was performed with a modified forward primer, which includes a fluorescent label as well as a thiol functionality, and two unmodified reverse primers, which

were designed to generate PCR products of 108 or 208 bp in length. After amplification, the PCR products of two different lengths were conjugated (in six separate reactions) to P1–85, –169, and –337. PCR product (dsDNA)-polypeptide conjugates were then denatured by heating for 3 min at 95°C in formamide solution to prepare ssDNA-polypeptide conjugates, and the resulting conjugates were analyzed by CE in a buffer containing 7 M urea (Fig. 5). Since the 208-base DNA acts as a more powerful “engine” than 108-base DNA, the conjugates comprising 208-base DNA (Fig. 5b) have a higher electrophoretic mobility than those of 108-base DNA (Fig. 5a), for the same protein polymer drag-tags. In addition, while the polydispersity of P1–169 and P1–337 is easily seen with the 108-base DNA, it is much less apparent for 208-base DNA. Most likely, the polydispersity of P1–169 (or P1–337) with the 208-base DNA engine is not fully resolved because of the relatively high mobilities of the conjugate molecules. Note also that the polydispersity of P1–337 with a 22-base DNA engine is quite apparent in Fig. 2c, but this is not the case for P1–337 with 208-base DNA in Fig. 5b. These results imply that, potentially, protein polymers that are not perfectly monodisperse (for reasons of glutamine deamidation, *etc.*), may still be useful as drag-tags for high-resolution ELFSE separations of larger DNA fragments.

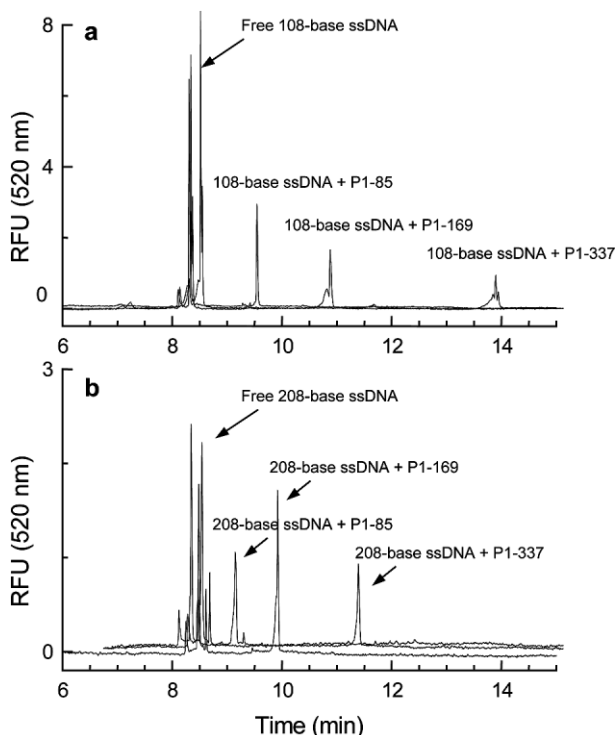


Figure 5. Capillary electropherograms obtained for the analysis of P1 protein-PCR product conjugates, with (A) a 108-base DNA and (B) a 208-base DNA. Electrophoresis conditions as in Fig. 2.

3.7 Calculation of hydrodynamic friction: the α value

To quantify the effective drag imposed on the DNA fragments by the various drag-tags, and to measure the potential resolving capabilities of the different drag-tag molecules, the effective hydrodynamic drag of each drag-tag (*i.e.*, its α value) was calculated using the free-draining model of Mayer *et al.* [4]. In this model, a DNA molecule with *N*-monomers is conjugated to a drag-tag, and the drag-tag is said to have a net free-flow mobility equal to β times that of DNA in free solution (μ_0), and a total hydrodynamic drag equivalent to α bases of DNA. In the absence of segregation between the DNA and the drag-tag (such a segregation could happen, for example, at extremely high fields), the new mobility of the DNA/drag-tag conjugate can be approximated as:

$$\mu(N) = \frac{V}{E} = \mu_0 \frac{(N - \beta)}{(N + \alpha)} \quad (1)$$

where $\mu(N)$ is the electrophoretic mobility of a DNA molecule with *N*-monomers conjugated to a drag-tag, μ_0 is the mobility of unlabeled DNA, *V* is the velocity of the analyte, and *E* is the electric field strength [4]. Since both the mobility and the α value may be strongly affected by the solution ionic strength, the free-draining model may not be

entirely accurate [7, 8]; however, the use of this model equation for calculating the α value enables us to at least estimate α , and also to compare directly the effective α values of our protein polymers with that reported for streptavidin, which was obtained using the same equation [16].

Since we know from structural information that the neutral protein polymers (P1-85, -169, and -337) are uncharged at the pH of analysis, for these we can set $\beta = 0$. Because $\mu(N)$ and μ_0 can be calculated directly from the peak elution times provided by electropherograms, the values of α for the neutral protein polymers can be determined from Eq. (1). The obtained α values for the P1-85, -169, -337 protein polymers are presented in Table 2, and are $\alpha = 15$, 29, and 70, respectively. There is an almost linear dependence of α on P1 chain length in this case, as previously observed with shorter polypeptoid drag-tags [5]. Additionally, apparent or "effective" α values for the P2-127, -253, -505 protein polymers are also presented in Table 2. Since these P2 protein polymers have negative charge, the actual β values are not 0. However, by intentionally setting $\beta = 0$ for these protein polymers, we can calculate the apparent amount of friction (or, an effective α value) imposed on the DNA fragment by the drag-tags during electrophoresis. Note that the calculated α value for streptavidin, composed of 636 amino acids, is only ~ 30 , and that the Slater and Drouin group performed DNA sequencing up to ~ 110 bases with this drag-tag (whereas the predicted read-length calculated with Eq. (2) is 141) in 18 min [13]. We obtain a very similar α value (29) for the P1-169 polypeptide, which is composed of only 169 aa. This similarity in α , despite the much smaller molecular weight of our protein polymer, most likely results from differences in protein conformation. Whereas streptavidin is compactly folded and hence roughly spherical, the P1 polypeptide adopts a random coil in solution, as evidenced by circular dichroism spectroscopy (data not shown). Both proteins have an α of roughly the same

magnitude, although the number of amino acids in streptavidin is almost 4 times that in P1-169 polypeptide. This, along with the higher peak efficiencies we observed, provides the strongest evidence that genetically engineered, unfolded protein polymers are better than natural, folded proteins as drag-tags.

The potential sequencing read-length of an ELFSE system, which depends on the potential DNA-resolving capabilities of the given drag-tag molecules, was estimated using the equation provided by Ren *et al.* [16], given as follows:

$$\frac{M_R^{1/2}(M_R + \alpha)^3}{\alpha^2} = \frac{L\mu_0 E}{16\ln(2)D_1} \quad (2)$$

where M_R is the potential read-length, L is an effective capillary length, and D_1 is the diffusion coefficient of ssDNA. With $\alpha = 70$, $L = 39.6$ cm, $E = 400$ V/cm, $\mu_0 = 1.95 \times 10^{-4}$ cm²/Vs (this value is obtained in a typical electrophoresis buffer and is independent of the applied field strength or DNA size), and $D_1 = 3.2 \times 10^{-6}$ cm²/s [16], which are the parameters obtained in our current system, we predict that we could obtain $M_R = 233$ bases using the P1-337 protein polymer. Other calculated potential read-lengths for obtained protein polymers are listed in Table 2. These results imply that, theoretically, up to 233 DNA bases could be sequenced using the P1-337 protein polymer drag-tag. However, this could only be done if the polypeptide can be obtained with greater homogeneity, since the polydispersity of P1-337 would make DNA sequence difficult to interpret. (Our next generation of drag-tags excludes chemically unstable amino acids, such as glutamine.) Moreover, if a homogeneous protein polymer substantially longer than P1-337 is successfully produced (which should be possible, using our recently developed controlled cloning method [19]), we would anticipate that these polypeptides could serve as excellent drag-tags for long-read DNA sequencing. Using Eq. (2) and assuming a linear dependence of α on drag-tag size [5], we predict that we would need an uncharged drag-tag comprised of 1461 aa ($\alpha = 319$) in order to obtain a read length of 700 bases, if the capillary length were 25 cm and $E = 2000$ V/cm. These conditions of analysis could be obtained in either a modified CE system or, perhaps better, on a microfluidic device.

Table 2. Measured α value and calculated potential read-length for the obtained protein polymers

Name	α value	Estimated potential read-length (M_R)
P1-85	15	108
P1-169	29	152
P1-337	70	233
P2-127	5.0 ^{a)}	60
P2-253	7.1 ^{a)}	73
P2-505	9.2 ^{a)}	84

a) Apparent or effective α values

4 Concluding remarks

We have presented the first ELFSE separations of DNA in which non-natural protein polymers were used as drag-tags. Size and charge effects of the protein polymer drag-

tags on ELFSE electrophoretic behavior were studied by conjugating the protein polymers to a 22-base oligonucleotide, and analyzing them *via* free-solution CE. The effect of a 24-residue His-tag on bioconjugate peak mobility and peak shape during electrophoresis was also explored by the electrophoretic analysis of FITC-labeled, negatively charged protein polymers. In addition, the electrophoretic behavior of bioconjugates comprising PCR products attached to the various uncharged drag-tags was investigated. The effective hydrodynamic drag of each drag-tag was calculated using the free-draining model of Mayer *et al.*, and the potential sequencing read-lengths, which represent the potential DNA-resolving capabilities of the given drag-tag molecules, were estimated. From the data analyses, we demonstrated that larger and uncharged drag-tags have the best DNA-resolving capability for ELFSE separations, and that theoretically, up to 233 DNA bases could be sequenced using one of the protein polymer drag-tags we created for these studies. We also show that, based both on the dependence of molecular hydrodynamic friction on molecular weight, and on experimentally obtained peak efficiencies, unstructured, genetically engineered protein polymers are better drag-tags than a natural, folded protein such as streptavidin. Work towards the accomplishment of long-read, four-color sequencing with protein polymer drag-tags is ongoing.

We thank Gary W. Slater for helpful discussions, and Russell Haynes for assistance with MALDI analysis. We acknowledge support from the U.S. Department of Energy, Office of Biological and Environmental Research (Grant DE-FG02-99ER6278) and the National Institutes of Health (Grant # NHGRI R01 HG002918-01). We acknowledge the use of the Keck Biophysics Facility at Northwestern University for circular dichroism measurements.

Received June 21, 2004

5 References

- [1] Olivera, B. M., Baine, P., Davidson, N., *Biopolymers* 1964, 2, 245–257.
- [2] Noolandi, J., *Electrophoresis* 1992, 13, 394–395.
- [3] Noolandi, J., *Electrophoresis* 1993, 14, 680–681.
- [4] Mayer, P., Slater, G. W., Drouin, G., *Anal. Chem.* 1994, 66, 1777–1780.
- [5] Vreeland, W. N., Barron, A. E., *Abstr. Pap. Am. Chem. Soc.* 2000, 219, 555–556.
- [6] Long, D., Dobrynin, A. V., Rubinstein, M., Ajdari, A., *J. Chem. Phys.* 1998, 108, 1234–1244.
- [7] McCormick, L. C., Slater, G. W., Karger, A. E., Vreeland, W. N., Barron, A. E., Desruisseaux, C., Drouin, G., *J. Chromatogr. A* 2001, 924, 43–52.
- [8] Desruisseaux, C., Long, D., Drouin, G., Slater, G. W., *Macromolecules* 2001, 34, 44–52.
- [9] Volkel, A. R., Noolandi, J., *Macromolecules* 1995, 28, 8182–8189.
- [10] Long, D., Ajdari, A., *Electrophoresis* 1996, 17, 1161–1166.
- [11] Vreeland, W. N., Meagher, R. J., Barron, A. E., *Anal. Chem.* 2001, 74, 4328–4333.
- [12] Heller, C., Slater, G. W., Mayer, P., Dovichi, N., Pinto, D., Viovy, J.-L., Guy, D., *J. Chromatogr. A* 1998, 806, 113–121.
- [13] Slater, G. W., Kist, T. B. L., Ren, H., Drouin, G., *Electrophoresis* 1998, 19, 1525–1541.
- [14] Slater, G. W., Desruisseaux, C., Hubert, S. J., Mercier, J.-F., Labrie, J., Boileau, J., Tessier, F., Pepin, M. P., *Electrophoresis* 2000, 21, 3873–3887.
- [15] Vreeland, W. N., Desruisseaux, C., Karger, A. E., Drouin, G., Slater, G. W., Barron, A. E., *Anal. Chem.* 2001, 73, 1795–1803.
- [16] Ren, H., Karger, A., Oaks, F., Menchen, S., Slater, G. W., Drouin, G., *Electrophoresis* 1999, 20, 2501–2509.
- [17] http://www1.amershambiosciences.com/aptrix/upp01077.nsf/Content/auto-dna_highthroughput^abi_prism_kit
- [18] Haider, M., Megeed, Z., Ghandehari, H., *J. Control. Release* 2004, 95, 1–26.
- [19] Won, J.-I., Barron, A. E., *Macromolecules* 2002, 35, 8281–8287.
- [20] Won, J.-I., Meagher, R. J., Barron, A. E., *Biomacromolecules* 2004, 5, 618–627.