Article

Sequencing of DNA by Free-Solution Capillary Electrophoresis Using a Genetically Engineered Protein Polymer Drag-Tag

Robert J. Meagher, Jong-In Won, Jennifer A. Coyne, Jennifer Lin, and Annelise E. Barron

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Sequencing of DNA by Free-Solution Capillary Electrophoresis Using a Genetically Engineered Protein Polymer Drag-Tag

Robert J. Meagher,†,‡ Jong-In Won,†,§ Jennifer A. Coyne,†,‖ Jennifer Lin,† and Annelise E. Barron*,†,⊥,⊗

Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, and Department of Chemistry, Northwestern University, Evanston, Illinois

We demonstrate the first use of a non-natural, genetically engineered protein polymer drag-tag to sequence DNA fragments by end-labeled free-solution electrophoresis (ELFSE). Fluorescently labeled DNA fragments resulting from the Sanger cycle sequencing reaction were separated by free-solution capillary electorophoresis, with much higher resolution and cleaner results than previously reported for this technique. With ELFSE, size-based separation of DNA in the absence of a sieving matrix is enabled by the end-on attachment of a polymeric "drag-tag" that modifies the charge-to-friction ratio of DNA in a size-dependent fashion. Progress in ELFSE separations has previously been limited by the lack of suitable large, monodisperse drag-tags. To address this problem, we designed, constructed, cloned, expressed, and purified a non-natural, genetically engineered 127mer protein polymer for use as an ELFSE drag-tag. The Sanger cycle sequencing reaction is performed with the drag-tag covalently attached to the sequencing primer, a major advance over previous strategies for ELFSE sequencing. The electrophoretic separation is diffusion-limited, without significant adsorption of the drag-tag to capillary walls. Although the read length (at about 180 bases) is still short, our results provide evidence that larger protein polymer drag-tags, currently under development, could extend the read length of ELFSE to more competitive levels. ELFSE offers the possibility of very rapid DNA sequencing separations without any of the difficulties associated with viscous polymeric sieving networks and hence will be amenable to implementation in microchannel and chip-based electrophoresis systems.

Exciting advances have recently been made in genome-scale DNA sequencing,[1,2] but for the majority of research laboratories, low- to medium-throughput sequencing operations are still performed using Sanger dye-terminator sequencing chemistry with electrophoretic DNA separation and analysis. The electrophoretic separation is typically performed in a viscous, entangled polymer sieving matrix using capillary array electrophoresis and increasingly by microchannel electrophoresis on microfludic devices.[3] Separation of DNA sequencing products with single-base resolution is a demanding process and has been the subject of intensive research and optimization;[4] further miniaturization of the process onto microfluidic chips still presents numerous technical challenges.

Sieving polymers can be difficult to work with due to their high viscosity in solution and also impose limitations on read length and electric field strength (and hence speed of separation) due to the mechanism of DNA migration through sieving matrixes.[5,6] End-labeled free-solution electrophoresis (ELFSE) has been proposed as an alternative mechanism of separation, which would be subject to different limitations than matrix-based sequencing.[7] In ELFSE, each DNA molecule is modified at one end with a neutral perturbing entity or "drag-tag", which serves to add additional hydrodynamic friction during electrophoresis, altering the constant scaling of charge and friction that is typical of DNA. Free-solution electrophoretic separation of DNA sequencing

* Corresponding author. Professor Annelise E. Barron, Department of Bioengineering, Stanford University, James H. Clark Center, 318 Campus Drive, Stanford, CA 94305-5444. Phone: (650) 721-1151. Fax: (650) 723-9801, E-mail: aebarron@stanford.edu.
 † Department of Chemical and Biological Engineering, Northwestern University.
 ‡ Present address: Sandia National Laboratories, Livermore, CA.
 § Present address: Department of Chemical Engineering, Hongik University, Seoul, South Korea.
 ‖ Present address: Department of Chemical Engineering, Stanford University, Stanford, CA.
 ⊥ Department of Chemistry, Northwestern University, Evanston, IL.
 ⊗ Present address: Department of Bioengineering, Stanford University, Stanford, CA.

(1) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z. T.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P. G.; Begley, R. F.; Rothberg, J. M. *Nature* **2005,** *437,* 376−380.
(2) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X. X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005,** *309,* 1728−1732.
(3) Blazej, R. G.; Kumaresan, P.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103,* 7240−7245.
(4) Kan, C. W.; Fredlake, C. P.; Doherty, E. A. S.; Barron, A. E. *Electrophoresis* **2004,** *25,* 3564−3588.
(5) Slater, G. W.; Kenward, M.; McCormick, L. C.; Gauthier, M. G. *Curr. Opin. Biotechnol.* **2003,** *14,* 58−64.
(6) Viovy, J. L. *Rev. Mod. Phys.* **2000,** *72,* 813−872.
(7) 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.
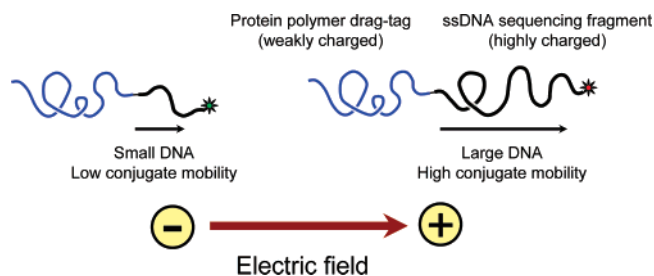
Figure 1. Schematic illustration of end-labeled free-solution electrophoresis of DNA sequencing fragments. Each ssDNA fragment is conjugated to an identical, neutral (or nearly neutral) protein polymer drag-tag, which is present during the Sanger cycle sequencing reaction and then adds additional friction during free-solution microchannel electrophoresis. The additional drag impacts the mobility of DNA fragments inversely according to size, such that conjugates comprising large DNA fragments migrate close to the free-solution mobility of DNA, while conjugates with smaller DNA are slowed down substantially, according to their size.

fragments conjugated to random-coil polymeric drag-tags is illustrated schematically in Figure 1. Because biased reptation and molecular orientation are not relevant in free-solution electrophoresis, ELFSE is potentially capable of performing longer reads at higher electric field strengths than matrix-based sequencing, without the practical difficulties or expense associated with sieving polymers. These features make ELFSE particularly attractive for implementation on microfluidic chips, where the inherently small length scales could allow very rapid separations at electric field strengths that would be impractical in traditional capillaries.

The concept of separating DNA sequencing fragments by electrophoresis in free solution by attachment of a perturbing entity was first proposed in 1992[8] and has been the subject of extensive theoretical analyses in the years since.[9–15] In 1999, Ren et al. provided the first experimental evidence that this technique could be used for DNA sequencing separations.[16] In this initial report of sequencing by ELFSE, recombinant streptavidin was used as a drag-tag, allowing the separation of biotinylated ssDNA fragments up to about 110 bases in size in free solution.

Ren et al. found that streptavidin provided hydrodynamic drag (α) equivalent to about 30 bases of DNA, which limited the sequencing read length to about 110 bases. The authors estimated that a read-length of 625 bases would require an uncharged and monodisperse drag-tag with a friction equivalent to about 300 uncharged ssDNA bases. Streptavidin is a roughly spherical, globular protein, which is an inefficient shape for generating hydrodynamic drag. The friction coefficient of a sphere increases linearly with the radius of the sphere, while the effective radius of a globular protein increases as roughly the $^{1}/_{3}$ power of its

molecular weight, meaning that a very large globular protein ($>30 \times 10^6$ Da) would be required to achieve the requisite value of α. Other morphologies, including unstructured or random-coil polymer chains, would provide better scaling of friction with molecular weight than compact, spherical proteins.

No suitable drag-tags are commercially available to produce a sequencing read even comparable to that obtained with streptavidin. Most other large natural proteins suffer from the same drawbacks as streptavidin: besides the compact globular structure, many proteins are heterogeneous in charge or molecular weight, which can dramatically affect mobility in electrophoresis. Proteins are also notoriously prone to adsorption to surfaces. Finally, most proteins lack a convenient technique for unique and stable conjugation to DNA. Synthetic polymers and "large" organic molecules tend to be either too small,[17] too polydisperse,[18] or both[19] for separation of large DNA molecules. Progress has been made in the synthesis of larger, branched polypeptoid drag-tags,[20] but even the largest of these synthesized to date only provided $\alpha \approx 16$, which is significantly less drag than streptavidin.

Therefore, we have applied genetic engineering techniques to create long, repetitive polypeptides or "protein polymers" to be used as large drag-tag molecules.[21] Artificial genes are constructed that encode for polypeptides with simple repetitive sequences that are expected to be useful for ELFSE.[22] Our initial attempts to create protein polymer drag-tags for ELFSE[21,23] highlighted the advantages of unstructured protein polymers compared to streptavidin, both in terms of the friction they generated and also in terms of improved peak efficiency, suggesting minimal adsorption of the protein polymers to capillary walls. The initial sequences included glutamine, which proved to be unsuitable for our purposes because of a tendency to deamidate during protein purification[23] and during thermal cycling.

We report here a new protein polymer sequence based on simple modifications to our previous sequences.[21,23] The new protein polymer consists of repeats of the amino acid sequence (Gly−Ala−Gly−Thr−Gly−Ser−Ala), where glutamine has been replaced by a stable, hydrophilic threonine. We constructed an artificial gene encoding for a protein polymer with 18 repeats of the sequence (a total of 127 amino acids). This protein polymer, although small, was tested for sequencing, with excellent results (presented here) that surpassed the original sequencing study performed with streptavidin.[16] Although still not truly "long reads," the results we present here demonstrate that protein polymer drag-tags do work for sequencing and have extensive potential for further development.

In keeping with the previous work by Ren et al.,[16] we have explored the effects of several experimental parameters on the

(8) Noolandi, J. *Electrophoresis* **1992**, *13*, 394−395.

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

(10) Hubert, S. J.; Slater, G. W. *Electrophoresis* **1995**, *16*, 2137−2142.

(11) Long, D.; Ajdari, A. *Electrophoresis* **1996**, *17*, 1161−1166.

(12) Long, D.; Dobrynin, A. V.; Rubinstein, M.; Ajdari, A. *J. Chem. Phys.* **1998**, *108*, 1234−1244.

(13) Desruisseaux, C.; Drouin, G.; Slater, G. W. *Macromolecules* **2001**, *34*, 5280−5286.

(14) Desruisseaux, C.; Long, D.; Drouin, G.; Slater, G. W. *Macromolecules* **2001**, *34*, 44−52.

(15) 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.

(16) Ren, H.; Karger, A. E.; Oaks, F.; Menchen, S.; Slater, G. W.; Drouin, G. *Electrophoresis* **1999**, *20*, 2501−2509.

(17) Vreeland, W. N.; Barron, A. E. *Abstr. Pap. Am. Chem. Soc.* **2000**, *41*, 1018−1019.

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

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

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

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

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

(23) Won, J. I.; Meagher, R. J.; Barron, A. E. *Biomacromolecules* **2004**, *5*, 618−627.

quality of separation, with the goals of optimizing sequencing read-length and determining the factors limiting separation. Experimental data from the current study are also compared to a theoretical model for chain stretching and deformation in a recently published companion paper.[24]

MATERIALS AND METHODS

Production of Protein Polymer Drag-Tag. A synthetic oligonucleotide encoding three repeats of the amino acid sequence (Gly−Ala−Gly−Thr−Gly−Ser−Ala) was purchased from Oligos Etc. (Wilsonville, OR) and multimerized by the controlled cloning process, using techniques described previously.[21−23] A multimer encoding 18 repeats of the amino acid sequence was isolated and expressed as a fusion protein with an N-terminal polyhistidine tag in *Escherichia coli*. The N-terminal polyhistidine tag was chemically cleaved using cyanogen bromide in 70% formic acid, and the target protein was obtained in high purity, with the molecular weight confirmed by MALDI-TOF mass spectrometry.

Conjugation of Drag-Tag to DNA Sequencing Primers. The protein polymer was activated at the N-terminus with Sulfo-SMCC by adding a 10:1 molar excess of Sulfo-SMCC to 1.2 mg of the protein in 80 μL of 100 mM sodium phosphate buffer, pH 7.2. The mixture was vortexed for 1 h, and excess Sulfo-SMCC was removed using a Centri-Sep gel filtration column (Princeton Separations, Adelphia, NJ). The purified, activated drag-tag was frozen and lyophilized, then resuspended in pure water at a concentration of 10 mg/mL.

A 17-base, thiolated M13 (-40) forward sequencing primer (5′-**X₁**GTTTTCCCAGTCACGAC, where **X₁** is a 5′-C6 thiol linker) was purchased from Integrated DNA Technologies (Coralville, IA). The 5′-thiol group was reduced by incubating 2 nmol of the DNA with a 20:1 molar excess of TCEP at 40 °C in a total volume of 20 μL of 70 mM sodium phosphate buffer, pH 7.2, for 2 h. The reduced oligonucleotide was desalted with a Centri-Sep column and immediately mixed with a 100-fold molar excess of activated drag-tag in 25 mM sodium phosphate buffer, pH 7.2, with a final DNA concentration of 4.2 pmol/μL. The conjugation reaction was allowed to proceed at room temperature for 4 h before further use. The efficiency of conjugation was determined by performing a single-base extension reaction with the conjugated primer, separating the products by free-solution electrophoresis, and measuring the areas of the peaks for conjugated and unconjugated DNA.[25]

Sequencing Reactions and Cleanup. DNA sequencing reactions were carried out using the SNaPshot Multiplex single-base extension (SBE) kit (Applied Biosystems, Foster City, CA), with deoxyribonucleotide triphosphates (dNTPs) added to facilitate Sanger sequencing instead of single-base extension. An amount of 5 μL of the SNaPshot premix was mixed with 8 nmol of dNTPs (1.8 nmol of dCTP, 1.8 nmol of dTTP, 2.2 nmol of dGTP, and 2.2 nmol of dATP), 4.2 pmol of the drag-tag-labeled primer, and 0.16 μg of the M13mp18 control DNA template (Amersham Biosciences, Piscataway, NJ) in a total volume of 10 μL. The reaction was cycled in an MJ Research Products Thermal Cycler, with 26 cycles of denaturation at 96 °C for 5 s, annealing at 50 °C

for 5 s, and chain extension at 60 °C for 30 s. Upon completion of thermal cycling, the reaction products were purified using a Centri-Sep column (Princeton Separations, Adelphia, NJ) and diluted to a final volume of 20 μL.

Capillary Electrophoresis. DNA sequencing separations with 4-color LIF detection were performed using an Applied Biosystems Prism 3100 Genetic Analyzer with arrays of 16 fused silica capillaries with inner diameters of 50 cm and effective lengths of 36, 50, or 80 cm. Separations were carried out at 55 °C in a denaturing buffer consisting of 1X TTE (89 mM Tris, 89 mM TAPS, 2 mM EDTA) with 7 M urea, with a 1:200 dilution of POP-5 solution added as a dynamic wall coating agent. The capillary was flushed with fresh buffer between each run, and the buffer reservoirs were replenished every 1−3 runs.

Sequencing samples were denatured prior to analysis by heating (in water) to 95 °C for 30 s, followed by snap-cooling on ice. Samples were introduced into the capillaries by electrokinetic injection at 1 kV for 20 s. Run voltages ranging from 3 to 15 kV were tested.

RESULTS

In the standard theory of ELFSE,[7,14,15] the electrophoretic mobility of a composite DNA−drag-tag copolymer is determined by a weighted average of the electrophoretic mobilities of charged DNA and uncharged drag-tag monomers. For a chain with M_c charged DNA monomers and M_u uncharged monomers, the mobility μ is

$$\mu = \mu_0 \frac{M_c}{M_c + \alpha_1 M_u} \tag{1}$$

where μ_0 is the free-solution electrophoretic mobility of DNA (independent of size) and α_1 is a weighting factor that rescales the number of uncharged monomers based on differences in size and persistence length as compared to the DNA monomers. Typical values of α_1 for flexible uncharged polymer chains range from $^1/_5$ to $^1/_6$. The product $\alpha \equiv \alpha_1 M_u$ represents the overall drag provided by a drag-tag and can easily be calculated from experimental data.

Protein Polymer Drag-Tag Production and Conjugation. We used a controlled cloning technique[22] to create an artificial gene encoding for a repetitive protein polymer with the repeating sequence (GAGTGSA)₁₈G, which became (GAGTGSA)₄−GAGT-GRA−(GAGTGSA)₇−GAGTGRA−(GAGTGSA)₅−G, after an unexpected mutation arose in the *E. coli* cell line used for expression of the protein. This mutation, which was discovered by DNA sequencing of the insert, gives the drag-tag a net charge of +1 following conjugation to DNA at the N-terminus.

The drag-tag was conjugated to a 5′-thiolated sequencing primer using the heterobifunctional crosslinker Sulfo-SMCC, in a high-yielding two-step reaction. The efficiency of conjugation was measured by performing a single-base extension reaction with the drag-tag labeled primer and an M13mp18 template, using the ABI SNaPshot single-base extension (SBE) kit. The sequencing primer (including any primer without an attached drag-tag) is extended by one base, incorporating a fluorescent dye terminator (ddGTP in this case). The resulting SBE product was purified with a Centri-Sep spin column and analyzed by free-solution

(24) Nedelcu, S.; Meagher, R. J.; Barron, A. E.; Slater, G. W. *J. Chem. Phys.* **2007**, *126*, 175104.

(25) Meagher, R. J.;. Coyne, J. A.; Hestekin, C. N.; Chiesl, T. N.; Haynes, R. D.; Won, J.-I.; Barron, A. E. *Anal. Chem.* **2007**, *79*, 1848−1854.
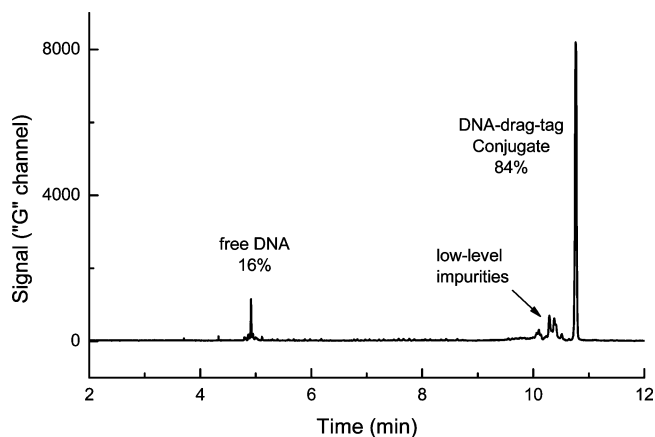
Figure 2. Electrophoretic analysis of single-base extension products from the protein polymer drag-tag conjugated to a 5′-thiolated 17-base M13mp18 sequencing primer. The analysis shows the mono-dispersity of the drag-tag and the efficiency of conjugation of the thiolated primer. Separation was carried out at 55 °C in a 36 cm capillary, with 1X TTE buffer, along with a 0.5% v/v of ABI POP-5 solution to provide a wall coating. Samples were injected at 1 kV for 20 s and separated at a field strength of 312 V/cm.

electrophoresis. The resulting electropherogram is presented in Figure 2. Conjugation efficiency was assessed by measuring the relative areas (normalized by migration time) for the unconjugated or "free" DNA peak (around 5 min) and the DNA−drag-tag conjugate peak (around 11 min). This analysis also reveals 3−4 low level "impurity" peaks, which represent small amounts of (unknown) impurities or modified drag-tags present along with the primary drag-tag. The impurities may result from truncated expression of the gene transcript in *E. coli*; truncated protein polymers would be copurified with the desired drag-tag due to the N-terminal histidine tag.

A notable result of this study is that a cycle sequencing reaction can be performed with the protein polymer drag-tag already attached to the sequencing primer; the presence of the drag-tag does not seem to interfere with the action of the sequencing polymerase nor disrupt the stability of the primer-template−enzyme complex. In the previous work with streptavidin,[16] thermal cycling was performed in the absence of streptavidin to prevent denaturation and aggregation of the protein. The resulting sequencing products were conjugated to streptavidin afterward, a tricky process dependent upon using exactly the right ratio of streptavidin to DNA. With the current protein polymer drag-tag, the efficiency of conjugation as well as the monodispersity of the drag-tag can be assessed prior to sequencing, which is an important improvement in efficiency as well as quality control.

Sequencing Reactions with Drag-Tag-Labeled Primer. The ABI SNaPshot kit, intended for single-base extension genotyping reactions, was used as a DNA sequencing kit simply by the addition of dNTPs. A total concentration of 800 μM dNTP in the sequencing reaction was found to generate a ladder of sequencing products up to about 250 bases long, an appropriate size range for this relatively small drag-tag. A typical sequencing electropherogram, obtained in a capillary with an effective length of 36 cm and a field strength of 312 V/cm, is shown in Figure 3. Apart from correction for spectral overlap of the dyes (performed automatically by the instrument) and simple baseline addition or

subtraction, this electropherogram shows "raw", unmodified data, without normalization of peak heights or correction for mobility shifts that are commonly performed on matrix-based sequencing traces. The sequence is read "backward", from the lower right to the upper left. The smaller fragments in the lower panel are resolved far more than necessary for unambiguous identification of each base. Determination of the sequence is quite straightforward in the bottom two panels, up to $M_c \approx 115$, although slight mobility shifts introduced by the different dye terminators cause certain peaks to overlap or shift out of the expected position. Beyond about 120 bases, the mobility shifts and unresolved peaks for repeated bases make identification of the sequence more difficult, although the observed peaks can be aligned with the known M13mp18 sequence to at least 180 bases with minimal effort. For a truly unknown template, more advanced data processing should be able to correct for the mobility shifts of the different terminators (as is commonly done for gel-based sequencing) and analyze peak widths to determine if a single broad peak might represent two or more repeated bases. Simple visual comparison of Figure 3 of this study with the previous streptavidin sequencing result (Figure 2 of Ren et al.[16]) shows that we obtained sharper, cleaner peaks and significantly better resolution with the protein polymer drag-tag.

The sequencing data in Figure 3 lead to a value of $\alpha = 25.0 \pm 0.1$. Fitting the data for the fragments labeled with each dye terminator separately yields slightly different α values for each terminator, which result from differences in hydrodynamic properties or net charge for each of the dyes. These slight differences in α could be used to correct for the mobility shifts for the different terminators observed in Figure 3.

Effects of Experimental Parameters on Sequencing Separations. Numerous electrophoretic separation variables may be manipulated to obtain better resolution and a longer sequencing read length apart from the drag-tag and sequencing chemistry. Variables tested in this study include electrophoretic velocity (a function of electric field), capillary length, and ionic strength of the buffer. Performance was quantified by tracking both theoretical plate height (*H,* a normalized measure of peak width) and resolution (*R,* a ratio of peak width to peak spacing).

Several factors can lead to an observation of band-broadening in ELFSE, including the initial injection zone width, analyte−wall interactions, polydispersity of the drag-tag, and thermal diffusion, the latter of which is unavoidable. Following the approach of Ren et al.,[16] the relative magnitudes of these different effects upon the theoretical plate height *H* can be quantified using an equation reminiscent of the van Deemter equation of chromatography:

$$H = \frac{A}{L} + \frac{2D}{u} + Wu + BL \qquad (2)$$

where *L* and *u* refer to the effective length of the capillary and the electrophoretic velocity, *D* is the diffusion coefficient, and *A, W,* and *B* are constants related to the magnitudes of the injection plug width, analyte−wall interactions, and polydispersity of the drag-tag, respectively. The height of a theoretical plate is the ratio of spatial variance σ_x^2 to capillary length *L*; for Gaussian peaks the plate height *H* can readily be determined from the peak migration time (which establishes the velocity *u*), the temporal
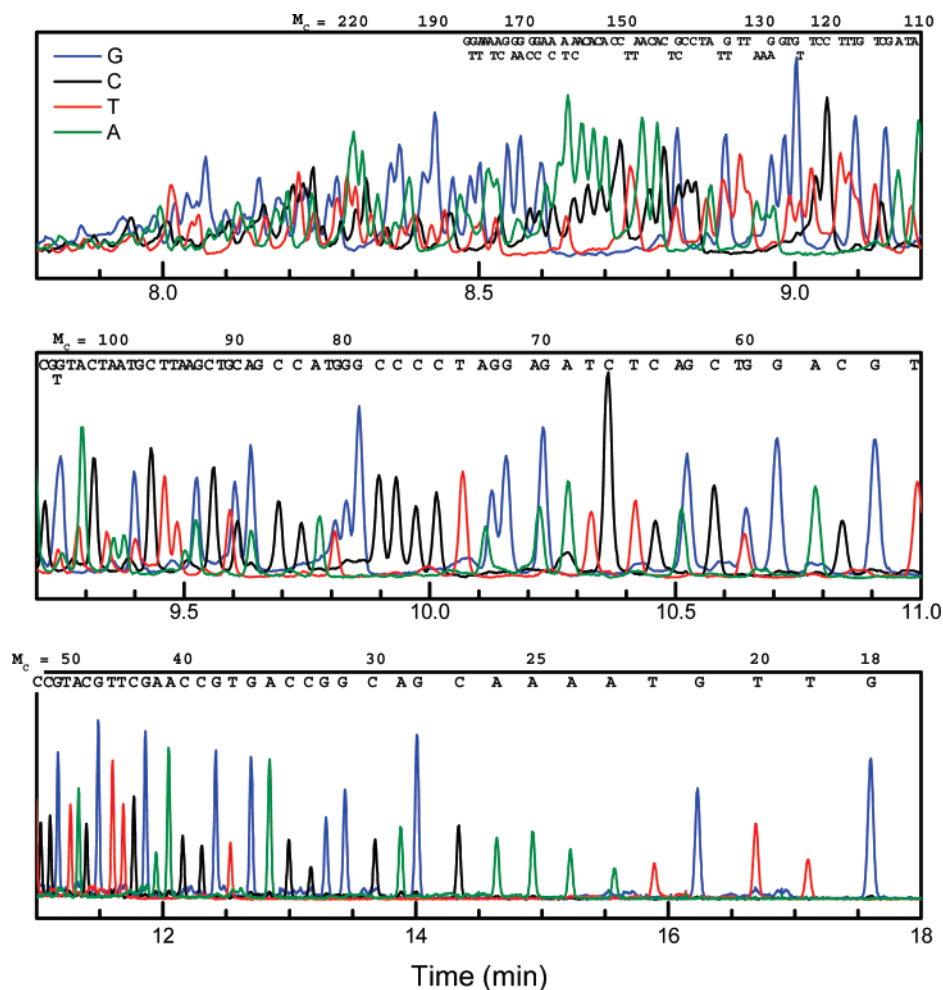
Figure 3. Four-color electropherograms for the products of a sequencing reaction carried out using the 127mer protein polymer drag-tag-labeled M13 primer and the SNaPshot SBE kit with dNTPs at a total concentration of 800 μM in the sequencing reaction. Conditions for capillary electrophoresis are identical to those described for Figure 2, with the addition of 7 M urea to the running buffer.

peak width at half-max (w), and the capillary length as

$$H = \frac{\sigma_x^2}{L} = \frac{u^2}{L} \frac{w^2}{8 \ln 2} \tag{3}$$

The relative contributions of the different effects on band broadening can thus be determined using eq 2 by independently varying either L or u and calculating H for a given size of DNA. Ideally, for a narrow injection zone and a monodisperse, nonadsorbing drag-tag, the primary contribution to band-broadening would simply be diffusion.

In addition to the plate height H, the resolution between adjacent peaks must also be considered for determining read lengths. The resolution factor R is essentially the ratio of peak width at half-maximum to the spacing between two peaks. The resolution for two peaks for DNA of size M_1 and M_2 can be defined as

$$R = \frac{(w_1 + w_2)}{2} \frac{(M_2 - M_1)}{(t_1 - t_2)} \tag{4}$$

where w_1 and w_2 refer to the peak widths and t_1 and t_2 refer to

the peak migration times. In practice, peaks for which $R \leq 1$ are well-resolved, whereas peaks for which $R > 1$ are run together, which makes determining the sequence difficult.

Effect of the Electric Field Strength. With the use of a capillary array with an effective length of 36 cm (total length of 47 cm), the electrophoretic velocity was varied by adjusting the applied voltage between 3 and 15 kV, the maximum allowed by the instrument. The electrophoretic velocity u and theoretical plate height H were tracked for two sizes of DNA, $M_c = 61$ and 103 (terminated by C and A, respectively). The electrophoretic velocities u for these two fragments increased linearly with the electric field strength (data not shown). This linear behavior indicates that changes in the conformation of the DNA–drag-tag conjugate (if any) with the electric field strength have only a small effect on the mobility of the conjugate in this regime. A more detailed treatment of the behavior of the DNA–drag-tag conjugate as a function of field strength, including comparison of the experimental data to a theoretical model for chain stretching and deformation in the low electric field regime, is presented in a companion paper.[24]

Theoretical plate heights H for the same two sizes of DNA–drag-tag conjugates are plotted with respect to the inverse of velocity ($1/u$) in Figure 4A. The plate heights H for both sizes of
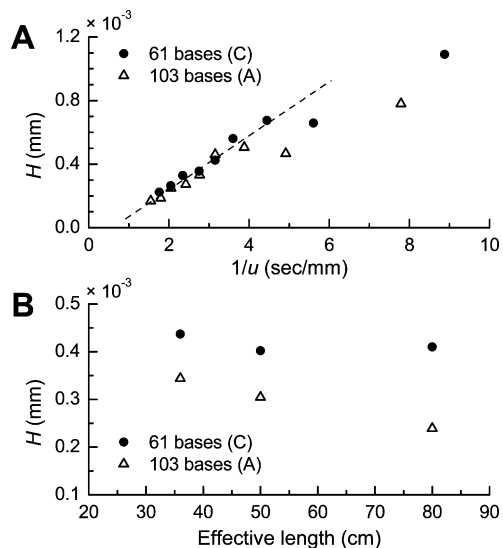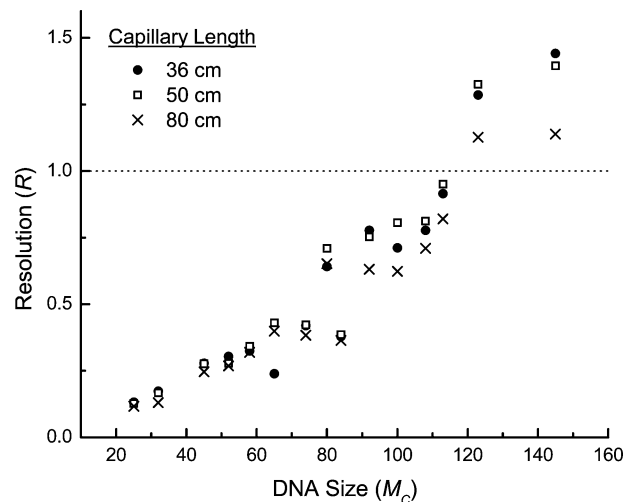
Figure 4. Effects of electrophoretic velocity and capillary length on the quality of electrophoretic separation. (A) Theoretical plate height H as a function of the inverse of electrophoretic velocity u for the 61- and 103-base sequencing fragments. Electric field strength was varied between 60 and 310 V/cm; other experimental conditions are as described in Figure 3. (B) Plate height H as a function of capillary length. The electric field was varied to give similar velocities in each length of capillary. Values of H for the 36 and 50 cm capillaries were interpolated from nearby values to match the velocities observed in the 80 cm capillary.

DNA are quite linear with respect to $1/u$ for all but the lowest velocities tested, with an intercept approaching zero at high velocity (i.e., as $1/u$ approaches zero). According to eq 2, this indicates that diffusion is the primary contributor to band-broadening at these conditions. The measured slope of 1.7×10^{-4} mm²/s suggests a diffusion coefficient D of about 8.5×10^{-7} cm²/s for the DNA−drag-tag conjugates, which agrees well with previous measurements for DNA in free-solution electrophoresis.[26]

Effect of Capillary Length. Sequencing separations were performed with three different lengths of capillary arrays, with 36, 50, and 80 cm effective lengths. The ABI 3100 offers only four lengths of arrays: the three tested here along with a short 22 cm array, at a considerable expense for each array. Thus, an exhaustive study of capillary length could not readily be performed. Plate heights were calculated for the 61- and 103-base, C- and A-terminated DNA fragments; these are plotted with respect to capillary length (at constant electrophoretic velocity u) in Figure 4B. Given the scarcity of data points, it is difficult to draw any conclusions about the injection plug width or the label polydispersity.

According to the standard theory of ELFSE,[7] the resolution of diffusion-limited ELFSE separations is essentially independent of capillary length L, with a dependence instead on the total applied potential (the product EL). The resolution R is plotted as a function of DNA size in Figure 5 for each length of capillary at a total applied potential of 14.7 kV. The read-length can be approximated as the point at which $R = 1$, which appears to occur between M_c = 110 and 120 bases, although more advanced base-calling

(26) Nkodo, A. E.; Garnier, J. M.; Tinland, B.; Ren, H. J.; Desruisseaux, C.; McCormick, L. C.; Drouin, G.; Slater, G. W. *Electrophoresis* **2001**, *22*, 2424−2432.

Figure 5. Resolution R, as defined by eq 3, calculated for selected pairs of closely eluting DNA fragments terminated with the same base. Separations were performed in capillaries with effective lengths of 36, 50, and 80 cm (total lengths of 47, 61, and 91 cm) at an applied potential of 14.7 kV, with all other conditions as described in Figure 3. The dotted line at $R = 1$ indicates the threshold value below which peaks can be considered well-resolved.
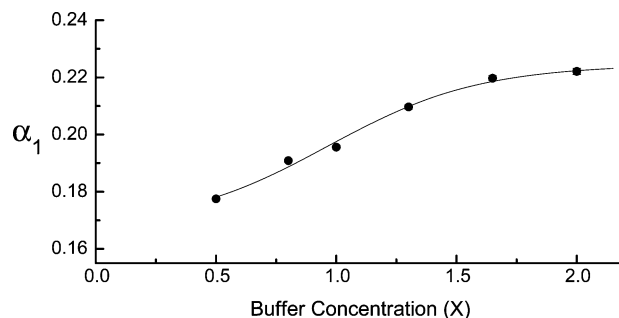
Figure 6. Effect of buffer concentration on the drag scaling parameter α_1. The solid curve is drawn to guide the eye. Buffer concentration 1X TTE ≡ 89 mM Tris, 89 mM TAPS, 2 mM EDTA; all buffers include 7 M urea and 0.5% v/v POP-5 solution. Separation conditions are as described in Figure 3, except the separation voltage was 11.3 kV (240 V/cm).

software may be able to accurately identify cases where a single broad peak actually represents two or more identical bases in a row. The values of R are slightly better for the 80 cm capillary, indicating some experimental deviations from the idealized theory.[7] This better resolution comes at the cost of significantly longer run time, especially given the limitation of the 15 kV total potential with the ABI 3100.

Effect of Buffer Concentration. The friction parameter α is a relative quantity which depends on the properties of the charged and uncharged monomers. Specifically, the scaling parameter α_1 depends on the ratio of sizes of the uncharged and charged monomers and the ratio of Kuhn lengths (related to chain stiffness) of the uncharged and charged monomers.[7] Increasing ionic strength is expected to decrease the Kuhn length of DNA without significantly affecting the drag-tag, thereby increasing α_1.[14]

To measure this effect, sequencing separations were carried out with TTE buffer concentrations varying by a factor of 4, from 90 to 360 mM total buffer concentration. The resulting electropherograms were used to calculate the values of α_1 for each buffer concentration, and the results are shown in Figure 6. As found

previously for streptavidin,[14] the value of α_1 depends strongly on ionic strength, increasing by 20% over the concentration range studied. However, it is not clear that this leads to significantly increased sequencing performance for high ionic strengths, as we observed no consistent trend between the buffer ionic strength and the resolution. Heat generation caused by high current with high ionic strength buffers may degrade sequencing performance, despite the larger values of α_1. The capillary inner diameter is also subject to optimization, although capillary arrays with diameters narrower than 50 μm are not commercially available for common sequencing instruments.

DISCUSSION

The results presented here show the first major progress in ELFSE DNA sequencing since the first report of sequencing with streptavidin in 1999 by Ren et al.[16] The protein polymer drag-tag used here has an effective drag similar to streptavidin, with α of about 25, but gives significantly cleaner results, with sharper peaks. This study presents the first strong evidence that protein polymer drag-tags, which we have been developing for several years, can actually be used for DNA sequencing. It is noteworthy that the presence of a 127-amino acid long protein polymer attached to the short (17-base) sequencing primer does not seem to interfere with the ability of the primer to hybridize to the template nor with the chain elongation activity of the sequencing polymerase. We have recently generated a longer version, 253 amino acids long, which when conjugated to a primer gives a clean peak with $\alpha \sim 55$ (data not shown). We anticipate that this longer drag-tag will soon allow a significantly longer sequencing read. The 127mer protein polymer drag-tag has also been employed successfully as part of a large set of distinct drag-tags used for highly multiplexed mutation detection by single-base extension followed by free-solution electrophoresis.[25]

Our detailed studies suggest that the separation is limited primarily by diffusion at high field strength. Small differences between the experimental results and theoretical predictions suggest that other mechanisms that were not quantified may have contributed to band broadening, possibly including temperature gradients arising from Joule heating. Further exploration of ELFSE sequencing should make use of narrower capillaries, although the design of commercial sequencing instruments place some limitation on the ability to fully optimize the separation. The short read length obtained with the current protein polymer drag-tag makes it difficult to assess sequence-specific effects such as GC content of different templates. Future studies of ELFSE with larger drag-tags should also employ a range of templates besides M13mp18, including templates with high GC content, to determine if the secondary structure of sequencing products impairs separation performance in free solution as it does in sieving matrixes.

ELFSE has frequently been mentioned as an enabling approach for sequencing on microfluidic devices,[7] which offer numerous advantages and greater flexibility than capillary sequencing instruments. The protein polymer drag-tag offers the first real possibility for testing ELFSE sequencing on microdevices, with no need for a polymer matrix for DNA separation, and implementation of this approach is underway.