Review

End-labeled free-solution electrophoresis of DNA

Robert J. Meagher[1]
Jong-In Won[1*]
Laurette C. McCormick[2]
Sorin Nedelcu[2]
Martin M. Bertrand[2]
Jordan L. Bertram[1]
Guy Drouin[3]
Annelise E. Barron[1]
Gary W. Slater[2]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA
[2]Department of Physics
[3]Department of Biology, University of Ottawa, Ottawa, Ontario, Canada

DNA is a free-draining polymer. This subtle but "unfortunate" property of highly charged polyelectrolytes makes it impossible to separate nucleic acids by free-flow electrophoresis. This is why one must typically use a sieving matrix, such as a gel or an entangled polymer solution, in order to obtain some electrophoretic size separation. An alternative approach consists of breaking the charge to friction balance of free-draining DNA molecules. This can be achieved by labeling the DNA with a large, uncharged molecule (essentially a hydrodynamic parachute, which we also call a drag-tag) prior to electrophoresis; the resulting methodology is called end-labeled free-solution electrophoresis (ELFSE). In this article, we review the development of ELFSE over the last decade. In particular, we examine the theoretical concepts used to predict the ultimate performance of ELFSE for single-stranded (ssDNA) sequencing, the experimental results showing that ELFSE can indeed overcome the free-draining issue raised above, and the technological advances that are needed to speed the development of competitive ELFSE-based sequencing and separation technologies. Finally, we also review the reverse process, called free-solution conjugate electrophoresis (FSCE), wherein uncharged polymers of different sizes can be analyzed using a short DNA molecule as an electrophoretic engine.

Keywords: Bioconjugates / Capillary electrophoresis / DNA-polymer conjugates / DNA sequencing / Drag-tags / End-labeled free-solution electrophoresis / Free-draining polyelectrolytes / Free-solution electrophoresis / Hydrodynamic friction / Review DOI 10.1002/elps.200410219

Contents

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

Abbreviations: ELFSE, end-labeled free-solution electrophoresis; **FSCE**, free-solution conjugate electrophoresis; **CGE,** capillary gel electrophoresis

1 Introduction

The complete sequencing of the Human Genome is certainly one of the greatest achievements of modern science [1, 2]. Capillary gel electrophoresis (CGE) made the completion of this enormous task possible. However, CGE is far from being a perfect analytical tool. For instance, it is well-known that the main physical mechanism responsible for the separation of the long DNA molecules, biased reptation, is limited to providing read-lengths of about 1000 bases because the DNA molecules tend to align with the applied field in the sieving

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

matrix during their migration [3]. In addition, the method is rather slow and expensive, and the loading of entangled polymer solutions in small-diameter capillaries is a process that creates many problems. Moving from capillaries to microfluidic devices will not change the latter situation since microchannels are essentially identical to capillaries; in fact, using narrower channels will simply exacerbate the polymer loading issue. Finally, CGE also requires long migration paths (typically on the order of tens of cm) in order for small differential molecular velocities to overcome the band-broadening processes.

A sieving matrix is required for electrophoresis to successfully separate DNA molecules of different sizes because DNA chains act like free-draining polymers during free-solution electrophoresis [3]. Each DNA monomer carries the same electric charge; therefore, the total force F applied to a molecule with M monomers increases linearly with M. If this external force were mechanical (*e.g.*, a sedimentation force), the friction coefficient ζ of the molecule would be that predicted by the Zimm theory of polymer dynamics [4]. Indeed, the different parts of the molecule would interact with each other *via* the fluid (*i.e.*, *via* long-range hydrodynamic interactions), and the random coil molecule would behave like an impermeable sphere of size R_H, the so-called hydrodynamic radius of the random coil. Since $R_H \sim M^{3/5}$ for long, semiflexible polymers [3], this would give $\zeta \sim R_H \sim M^{3/5}$ and a velocity that scales like $v = F/\zeta \sim M^{2/5}$; this hypothetical process would thus lead to a size-dependent velocity and the successful separation of DNA molecules of arbitrary size. Unfortunately, electrical forces generate a different situation, especially when there is salt in the buffer solution, which is always the case in practice. The electric field forces the DNA molecule and its counterions to move in opposite directions. The counterions move through the random coil of the DNA molecule such that the latter is no longer an impermeable sphere; instead, we say that the coil is free-draining. As we will discuss later, this process effectively screens the hydrodynamic interactions between the different parts of the DNA molecule so that the resulting electrophoretic friction coefficient actually scales like $\zeta \sim M$ instead of $\zeta \sim R_H$. This gives rise to a universal electrophoretic velocity $v = F/\zeta \sim M/M \sim M^0$, *i.e.*, the velocity is independent of the size of the DNA molecule and no separation is achieved [3]. Moreover, the local balance between force and friction means that the random coil conformation is not deformed by the migration. This symmetry between force ($F \sim M$) and friction ($\zeta \sim M$) is the unfortunate and unavoidable consequence of the subtle effects taking place during free-flow electrophoresis of large flexible polyelectrolytes in buffer solutions containing salt.

Gel electrophoresis resolves this problem by forcing the DNA molecules to collide with fixed obstacles during their electrophoretic migration [3]. Since longer DNA molecules collide more frequently with the gel fibers, they are slowed down to a larger extent and small molecules elute first, thus providing size-separation. This collision-driven process obviously results in longer elution times, but it also leads to conformational deformation of the DNA since the electrical forces are rather large, and this deformation tends to reduce the size-dependence of the net electrophoretic velocity [3, 5]. In practice, ssDNA molecules longer than about 1000 bases cannot be sequenced because of this molecular deformation effect.

However, gel electrophoresis is not the only way to overcome the free-draining properties of DNA during electrophoresis. In principle, one can also break the balance between force and friction by modifying the DNA at the molecular level. All that one needs to do is to modify the DNA in such a way that its ratio of electric charge to friction coefficient is rendered size-dependent. The only practical way to do this is to increase the friction coefficient of DNA, since it is not possible to substantially modify its average charge density. This idea has probably been around for quite some time; it is clear that many scientists looked at ways to do this back in the 1980s when the Human Genome Project was first conceived. However, it was not possible to test such ideas until the invention of capillary electrophoresis (CE), because free-solution electrophoresis leads to strong heat generation effects when it is not carried out in a narrow-diameter channel. Perhaps the first paper that specifically mentions this idea is that of Noolandi in 1992 [6]; however, the paper is purely speculative and does not provide experimental data or theoretical predictions. Moreover, this paper suggests attaching an object with a negative charge to ssDNA molecules prior to free-solution electrophoresis; most likely, this would be an inefficient way to restore a size-dependent DNA velocity, since the object would increase both the net force and the net friction on the ssDNA molecules.

The first paper to examine this concept quantitatively was published in 1994 by Mayer, Slater, and Drouin [7]. The theory developed by these authors made remarkable predictions and the name end-labeled free-solution electrophoresis (ELFSE) was coined for the process wherein an uncharged label is attached to the end of the DNA molecule in order to increase its friction coefficient without affecting its total charge. The paper mentions the possibility of sequencing more than 1000 or even 2000 bases in less than 1 h without the need for a sieving matrix. As we shall see later, the theory used by these authors needs major revisions, but their general conclusions are still valid.

The first paper showing positive experimental results for dsDNA was that of Heller *et al*. in 1998 [8], and ssDNA sequencing was achieved in 1999 by Ren *et al*. [9]. However, ELFSE has not yet become a competitive sequencing technology because the read-length has remained too low (slightly above 100 bases). The reason for this very limited read-length appears to be essentially chemical: it has been challenging to design and create monodisperse labels large enough to provide the extra friction necessary to enable successful sequencing of at least 500 bases. The ELFSE process is now well-understood, thanks to improvements in theory and detailed experimental studies. This article will review the progress achieved since the 1994 Mayer *et al*. paper [7]. The ELFSE process has also been inverted. Indeed, instead of using an uncharged label to provide extra friction for hard-to-separate polyelectrolytes like ssDNA, one can use ssDNA to pull uncharged labels and separate the latter by free-solution CE. This idea was used successfully to separate PEG molecules of different sizes [10], effectively achieving results comparable to MALDI-TOF by CE. This reverse process will also be reviewed in this article.

2 Free-solution electrophoresis of DNA

2.1 Polymers in solution

In order to understand the properties of semiflexible (or worm-like) polymers such as DNA, one needs to define several important molecular length scales: the polymer Kuhn length b_K, its radius of gyration R_g, its molecular weight M (number of monomers), the monomer size b, and finally the total contour length $L_D = Mb$ of the molecule. Classical results from polymer statistical mechanics [4, 11, 12] show that the mean square end-to-end distance of a freely jointed random-walk chain (for which $b_K = b$) is simply given by $\langle h^2 \rangle = b_K L_D = bL_D$. For a semiflexible chain, however, the Kuhn length b_K is larger than the monomer size b; in fact, b_K is related to the ensemble-average of the projection of an infinitely long chain along the direction of the first chain segment [11, 12]. In other words, b_K is a measure of the stiffness of the chain backbone, and segments of length b_K are independent from each other so that the relation $\langle h^2 \rangle = b_K L_D$ still applies for very long chains (*i.e.*, for $L_D \gg b_K$). When $b_K \geq L_D$, however, the chain is very stiff and behaves like a rod-like object with $h \cong L_D$. For dsDNA, the Kuhn length is typically $b_K \cong 100$ nm while $b \cong 0.34$ nm (these numbers are approximately 6 nm and 0.43 nm, respectively, for ssDNA).

More useful is the polymer radius of gyration R_g which characterizes its conformational size; it is given by the root mean square distance between the monomers and the center of mass of the molecule. For the freely jointed random-walk chain, one has simply $\langle {R_G}^2 \rangle = \frac{1}{6} Mb^2$. The semiflexible properties of DNA are better captured by the Kratky-Porod equation [11]:

$$\langle {R_G}^2(M) \rangle \approx \frac{b_K L_D}{6} \left[1 - 3\left(\frac{b_K}{2L_D}\right) + \right.$$

$$\left. + 6\left(\frac{b_K}{2L_D}\right)^2 - 6\left(\frac{b_K}{2L_D}\right)^3 \left(1 - e^{-2L/b_K}\right) \right] \quad (1)$$

It is easily verified that one gets $\langle {R_G}^2 \rangle = \frac{1}{6} L_D b_K$ for very long chains, and $\langle {R_G}^2 \rangle = \frac{1}{12} {L_D}^2$ for short, rod-like chains. For contour lengths L_D much larger than the Kuhn length, DNA is a semiflexible random coil with L_D/b_K Kuhn segments of length b_K undergoing Brownian motion. Short DNA fragments, on the other hand, can be approximated by straight, rigid cylinders.

The best model to describe the diffusion of a polymer in free solution is the Zimm model [4, 11]. The hydrodynamic interactions between the monomers are strong (this interaction decays only as the inverse of the distance) and dominate the dynamics. As a consequence, the coil behaves like an impermeable spherical object, and its friction coefficient ξ when diffusing in a solvent of viscosity η is given by Stokes' law:

$$\xi = 6\pi\eta R_H \quad (2)$$

This expression actually defines the hydrodynamic size R_H of the chain. In dilute solutions, the diffusion coefficient of the polymer chain is then given by the Stokes-Einstein relation:

$$D = \frac{k_B T}{\xi} \quad (3)$$

Note that we have roughly $R_H \cong \frac{2}{3} R_g$ for a random coil; in other words, these two lengths describing a polymer conformation are quite similar.

The theoretical description presented above neglects the excluded volume interactions between monomers of the same chain. The Flory theory for chains with excluded volume describes the balance between the repulsion energy between the monomers, which swells the chain, and the entropy loss upon swelling [4, 11, 12]. It predicts the following scaling law:

$$R_H \sim M^{\nu} \quad (4)$$

where Flory's exponent is $\nu = 1/2$ for an ideal chain (consistent with the discussion above) and $\nu = 3/5$ for a swollen linear polymer. Although the random-walk model and the Flory equation describe two different types of polymer chains, there is actually a critical DNA molecular weight for which both approaches predict the same mean end-

to-end distance *h.* Following the approach of Section 3.1 in [12], this critical value can be estimated using the scaling relation [3]:

$$h^* \cong \frac{b_K^2}{d} \tag{5}$$

where $d \approx 1.1$ nm is the diameter of an ssDNA strand. Below that size, excluded volume interactions are negligible, and one can use the Kratky-Porod expression (1); above that size, however, one must use Flory's expression (4) with $\nu = 3/5$. Using the fact that $h^2 \cong Mbb_K$ for a Gaussian coil, Eq. (5) gives the critical molecular size

$$M^* \cong \frac{b_K^3}{bd^2} \tag{6}$$

With $b_K \cong 6$ nm and $b \cong 0.43$ nm, this gives a critical molecular size of about $M^* \cong 400$ bases for ssDNA. Although this is a rough estimate (numerical factors of order unity were dropped from the calculation), it suggests that for typical DNA sequencing ($M = 100 - 1000$ bases), the excluded volume interactions are not strong and one can reasonably use the Kratky-Porod equation and the related expressions for describing molecular conformations and dynamics.

2.2 Distribution of ions in solution

If a charged object is immersed into an electrolyte solution, the freely moving ions will distribute themselves around the object in order to minimize the free energy of the system. Under normal circumstances, the local ionic concentration will follow the Boltzmann distribution, and the local electrical potential will then be given by the Poisson-Boltzmann equation:

$$\varepsilon \nabla_r^2 \Phi(\mathbf{r}) = -e \sum_j z_j C_j \exp\left(-\frac{e z_j \Phi(\mathbf{r})}{k_B T}\right) \tag{7}$$

where the sum is over the ionic species in the solution, ez_j and C_j are the charge and bulk concentration of ion species j, respectively, and ε is the permittivity of the fluid. Equation (7) is highly nonlinear and an analytical solution does not generally exist, except for particular geometries. In the weak field limit $|ez_j\Phi| \ll k_B T$, one can use a linearized version of Eq. (7):

$$\nabla^2 \Phi = \frac{\sum_k e^2 z_j^2 C_j}{\varepsilon k_B T} \Phi = \frac{1}{\lambda_D^2} \Phi \tag{8}$$

This expression defines the well-known Debye length:

$$\lambda_D = \left(\frac{\varepsilon k_B T}{e^2 \sum_k z_k^2 C_k}\right)^{1/2} \equiv \left(\frac{\varepsilon k_B T}{2e^2 I}\right)^{1/2} \tag{9}$$

where we introduced the ionic strength $I = \frac{1}{2}\sum_j z_j^2 C_j$ of the buffer solution. For most systems of interest in colloid and polyelectrolyte science, the Debye length ranges from a fraction of a nanometer to (at most) 100 nm. The solution of Eq. (8) is quite simply given by:

$$\Phi(x) = \Phi(0) \exp(-x/\lambda_D) \tag{10}$$

where $\Phi(0)$ is the potential at the surface $x = 0$. The potential, and hence the ion concentrations, thus decay exponentially near the charged object, with a length scale defined by λ_D.

2.3 Electrophoresis of spheres and infinite cylinders

The full theory of the electrophoresis of a spherical object of radius R is a complicated matter, which must include coupled electrostatic and hydrodynamic equations. Two different limits, however, lead to important and simple results. If the particle radius R is small compared to the extent λ_D of the Debye cloud of counterions, we find [3]:

$$\mu = \frac{2\sigma R}{3\eta} = \frac{Q}{6\pi\eta R} = \frac{Q}{\xi} \tag{11}$$

where σ is the surface charge density of the particle, Q is its total charge, and ξ is the friction coefficient of a sphere of radius R, as given by Eq. (2). In the opposite (Debye-Huckel) limit of a very large particle, we have [3]:

$$\mu = \frac{\sigma \lambda_D}{\eta} \tag{12}$$

We note that the radius of the particle does not appear in the last expression. This means that all particles with the same surface charge density σ will have the same electrophoretic mobility μ, a situation that does not allow separation. Moreover, the shape of the particle does not matter either. Electrophoresis of DNA is similar to this case, as we will see later.

2.4 Polyelectrolytes in solution

Obviously, the Kuhn length of a polyelectrolyte is made larger by the electrostatic repulsion between the charged monomers [13]. Although higher ionic strengths screen these repulsive interactions, a deeper understanding of this issue requires a look at the nature of screening. The distribution of the counterions around the polyelectrolyte backbone is generally divided into a Stern layer of counterions firmly attached to the molecule and a diffuse layer, which refers to the mobile counterions of the Debye ion cloud discussed previously. The thickness of the Stern layer is roughly given by the Bjerrum length $l_B = e^2 / \varepsilon k_B T$;

this length is the distance below which the electrostatic potential energy between two attracting charges in this medium is larger than the thermal energy that tends to shake them apart. For water at room temperature, the Bjerrum length ($l_B \cong 0.7$ nm) is of molecular dimensions.

The counterion condensation theory [14] describes the distribution of counterions near a polyelectrolyte. As discussed by Hoagland *et al.* [15], a key parameter is the dimensionless parameter $\rho = l_B/b$, the ratio of the Bjerrum length to the distance b between the charges along the chain backbone. When $\rho \ll 1$, the polyelectrolyte is not very charged and a simple linearization of the Poisson-Boltzmann equation will do fine. When $\rho > 1$, however, counterions moving close to the chain condense on the chain backbone to reduce the local charge density. The counterion condensation theory [14] predicts that the counterions will actually redistribute themselves so that the net charge becomes consistent with $\rho \cong 1$. This forms the Stern layer. This is also why many models of DNA electrophoresis use an effective (reduced) linear charge density. Since $\rho > 1$ for DNA, this effect reduces the mobility of the ssDNA (experimentally by a factor of about 3 – 6) over what one would expect for a polymer with about 1 charge per 0.43 nm. Note that this also means that all highly charged polyelectrolytes ($\rho > 1$) will be affected by counterion condensation in a similar way, and will thus end up having very similar mobilities [15]. This remarkable effect is the reason why ELFSE aims at changing the effective friction coefficient of the DNA molecules, not their charge.

Given the counterion condensation effect described above, there are four important length scales remaining for a polyelectrolyte such as DNA: the monomer size b, the Debye length λ_D, the Kuhn length b_K, and the hydrodynamic radius R_H. In the next section we shall discuss the four possible electrophoresis regimes in terms of these length scales. In the case of ssDNA sequencing, though, the most common situation is characterized by the inequality $R_G > b_K > \lambda_D > b$ because a ssDNA molecule is a moderately long chain with a stiff backbone, while the amount of salt generally present reduces λ_D to near-molecular dimensions.

2.5 Electrophoresis of polyelectrolytes

Since Hermans [16], many theoreticians have studied the electrophoresis of random-coil polyelectrolytes (or porous particles/macromolecules). In short, the results indicate that if λ_D is large compared to the radius of gyration R_g of the molecule, the frictional resistance to drift is mostly due to the fluid itself, and we recover essentially the situation of an impermeable sphere pulled through a liquid by a mechanical force. In the opposite limit $\lambda_D \ll R_g$, viscous dissipation is due to the local friction between the DNA chain and the counterions, because the two are quite close and move in opposite directions. In the latter case, the interplay between the different length scales may be tricky.

Desruisseaux *et al*. [17] considered four possible regimes depending on the relative value of λ_D and the other length scales mentioned previously. When the Debye length is larger than the radius of gyration of DNA, $\lambda_D > R_G(M)$ (see Fig. 1A), the electrophoretic friction coefficient is given by the Stokes relation $\zeta = 6\pi\eta R_H$, and the resulting electrophoretic mobility is given by:

$$\mu(M) = \frac{Me}{6\pi\eta R_H(M)} \tag{13}$$

where $Q = Me$ is the total charge of the polyelectrolyte. This is essentially a restatement of Eq. (11). In this case of extremely low ionic strengths and/or small molecules, it should be possible to separate DNA fragments.

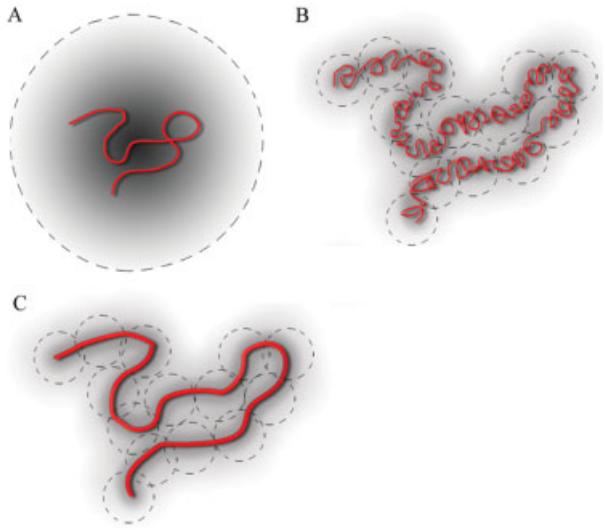

Figure 1. The three regimes of free-solution DNA electrophoresis (the DNA molecule is surrounded by a cloud of counterions). (A) The radius of gyration R_g of the DNA coil is smaller than the Debye length λ_D describing the extent of the cloud. In this case, the DNA is not free-draining, and we can separate the molecules. (B) The Debye length is smaller than R_g but remains longer than the Kuhn length b_K of the DNA molecule. Gaussian blobs of size λ_D are then hydrodynamically independent from each other, and the DNA molecule is free-draining. The net mobility is given by that of one of the blobs. (C) The situation is similar to the previous case, except that λ_D is comparable to b_K. The blobs are not Gaussian in conformation.

For the other three cases, a large DNA molecule has the same mobility as a DNA molecule of radius λ_D since viscous dissipation (and hence hydrodynamic interactions) do not extend beyond the range over which the counterions are found. For all practical purposes, the DNA chains are then made of independent subchains of hydrodynamic radius λ_D, each of which moves as if it were in "free-fall" in the electrostatic potential. Therefore, the mobility becomes independent of the DNA size *M*. When $R_G(M) > \lambda_D > b_K$, for example, we have large Gaussian blobs (each with a friction coefficient $\zeta_D \cong 6\pi\eta\lambda_D$ and a charge $Q_D \cong 6e\lambda_D^2/bb_K$; see Fig. 1B) and the mobility is given by:

$$\mu \cong \frac{e\lambda_D}{\pi\eta b b_K} \sim \frac{1}{\sqrt{I}} \tag{14}$$

where we now show the dependence upon the ionic strength *I* of the buffer solution. When the DNA is rigid on the length scale of a Debye length, *i.e.*, when $R_G(M) > b_K > \lambda_D > b$, the mobility is that of a cylinder of length λ_D and radius *b* (see Fig. 1C). The charge of those independent DNA cylindrical blobs is $Q_D \approx e\lambda_D/b$, while their electrophoretic friction is $\zeta_D \approx 3\pi\eta\lambda_D/\ln(\lambda_D/b)$; the corresponding electrophoretic mobility is thus:

$$\mu \cong \frac{e}{3\pi\eta b} \ln\left(\frac{\lambda_D}{b}\right) \sim \ln\left(\frac{1}{\sqrt{I}}\right) \tag{15}$$

Finally, the case $\lambda_D < b$ leads to a mobility that is independent of the ionic strength, since the Debye length does not play any role in this limit. It is important to note, however, that this analysis is not complete, since some end effects actually give rise to size-dependent mobilities below about 20 bases for ssDNA and below about 400 base pairs for dsDNA [15, 18].

3 Electrophoresis of composite molecules: theory

As discussed in the previous sections, free-solution electrophoresis of DNA fragments does not normally allow size-based fractionation because of the free-draining nature of such highly charged polyelectrolytes (although separation is sometimes possible under strong confinement [19, 20]).

Mayer *et al*. [7] were the first to try to estimate the potential of ELFSE for ssDNA sequencing. Their theoretical model was based on two main assumptions: (i) that one can use the Nernst-Einstein equation to relate the diffusion coefficient *D*(*M*) and the electrophoretic mobility $\mu(M)$ of a ssDNA molecule of size *M* in free solution (this is their Eq. 2); (ii) that the velocity of a hybrid ELFSE molecule (made of a charged component, the ssDNA fragment, and an uncharged component, the drag-tag) is simply given by the total electrical force applied to the molecule divided by the total friction coefficient of a free-draining molecule (this is their Eq. 1). Using these two assumptions, it is rather straightforward to estimate the elution time and the maximum read length (in bases) for a given set of experimental conditions (field intensity *E*, capillary length, injection zone width, and the effective drag produced by the drag-tag). Conservative numbers predicted that 500 bases could easily be sequenced in less than 45 min, while extreme conditions (*e.g.*, very narrow injection zone, high electric fields) could potentially lead to 1500–2000 bases called in about the same time. The size of the label necessary to achieve such results corresponded to having between α = 100 and α = 250 additional uncharged ssDNA monomer equivalents of drag attached to the ssDNA molecules to be sequenced.

These predictions attracted a lot of attention, but it took several years before an experiment demonstrated that the general idea was indeed valid (see Section 4). However, we now know that the two assumptions of the Mayer model are wrong. It was recently shown that the Nernst-Einstein relation is not valid for free-solution electrophoresis of DNA molecules [21]. In fact, the diffusion coefficient is not affected by the electric field and the DNA does not behave as a free-draining polyelectrolyte as far as diffusion is concerned. This means that the Mayer *et al*. paper [7] actually underestimated diffusional band broadening and overestimated ELFSE's potential performance. Moreover, the simultaneous presence of electrical and mechanical forces (the latter coming from the effect of the drag-tag) on the hybrid molecule leads to a situation where forces and frictions are not necessarily additive. Long and co-workers [22–26] have recently revised the theory of electrophoresis for composite molecules, and they have shown that the second assumption of the Mayer *et al*. paper is only a special case. Luckily, it turns out that this is the most likely case to occur in practice, so we may continue to use this hypothesis with some prudence. Clearly, these factors indicate that a new theory of ELFSE was required. The rest of this subsection describes the new ELFSE model that we have developed over the last several years.

3.1 Standard theory of ELFSE

In principle, an end-labeled ssDNA composite molecule can, during free-solution electrophoresis, adopt four possible hydrodynamic conformations (see Fig. 2). All of these behave differently when it comes to estimating the electrophoretic mobility μ of the DNA-drag-tag conjugate. In Fig. 2A, we see a simple case where the drag forces are not large enough to disturb the random coil conformation

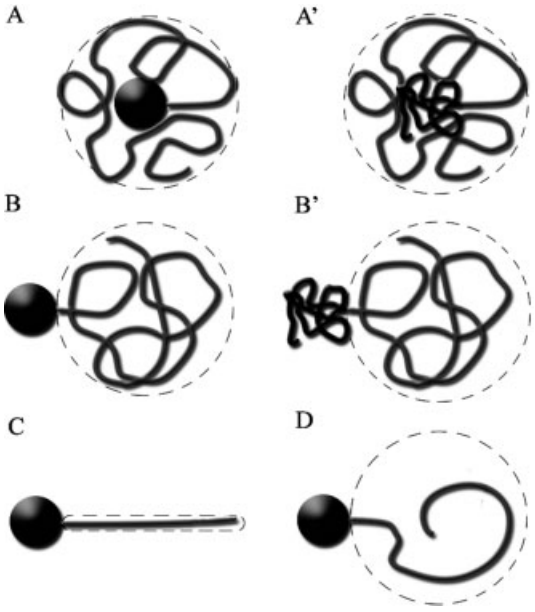

Figure 2. Schematic diagram showing various ELFSE regimes for a spherical drag-tag (A, B, C, D) and an extended polymeric drag-tag (A' + B'); the DNA is shown in red and the label in black. (A+A') The drag-tag is part of the random coil conformation of the DNA molecule. (B+B') The drag-tag and the DNA are hydrodynamically segregated, but the random-coil conformation is not disturbed. (C) The DNA stretches out in response to the drag forces. (D) Short DNA molecules can be sterically segregated from a bulky drag-tag.

of the hybrid molecule. Once these forces reach a certain threshold, however, the drag-tag separates physically from the undisturbed ssDNA component and we have hydrodynamic segregation (Fig. 2B). At large field, the drag force slowing down the ssDNA is so large that the latter stretches out (Fig. 2C). Finally, it should be mentioned that the ideal case of Fig. 2A cannot occur if the ssDNA molecule is not large enough to accommodate the label; for example, a short ssDNA molecule attached to a large bulky label would actually be segregated from it even at low velocity, as shown schematically in Fig. 2D.

Which case is most relevant for ssDNA sequencing? In order to answer this question, we must estimate the value of the critical electric field E_o that one needs to move out of the random coil conformation regime (Fig. 2A) and into the hydrodynamic segregation regime (Fig. 2B). However, to do so we need to compute the mobility of a labeled ssDNA molecule in the random coil regime. So this will be our starting point.

Let us first examine the case of a polymeric label made of M_u monomers (see Fig. 2, A' and B') that have the same physical properties (hydrodynamic radius and Kuhn length) and thus the same friction coefficient as ssDNA monomers. Since the resulting block copolymer forms an unsegregated random coil, the composite electrophoretic mobility μ can be approximated, according to the Long *et al.* [26] theory of polyampholyte electrophoresis, by a uniformly weighted average of the free-solution mobilities of the components (μ_o for ssDNA and μ = 0 for the uncharged label):

$$\mu(M) = \mu_o \frac{M_c}{M_c + M_u} \tag{16}$$

where M_c is the number of charged monomers and $M = M_c + M_u$ is the total number of monomers. For an elution length L, the elution time of these ssDNA fragments is given by

$$t(M) = \frac{L}{\mu(M)E} = \frac{L}{\mu_o E}\left(1 + \frac{M_u}{M_c}\right) \tag{17}$$

We clearly see that the shorter the ssDNA fragment, the longer its elution time. This predicts that separation is possible, but the peak spacing would be rather poor for $M_c \gg M_u$.

In most cases, however, uncharged label monomers would not be hydrodynamically equivalent to ssDNA monomers. In this case, a uniformly weighted average of the monomers' electrophoretic properties is not valid. We then have to revise the way we compute the mobility; to do so, we must use the so-called blob theory. To use this powerful theoretical concept, we must regroup the monomers into "blobs" (or subgroups) that have equivalent hydrodynamic properties [26, 27]. These blobs will then act as "supermonomers" with the same hydrodynamic friction coefficient but possibly different charges (see Fig. 3A). Proceeding in this way, we can construct an equation similar to Eq. (16) where M_c and M_u are now replaced by M_{Bc} and M_{Bu}, the number of charged and uncharged blobs, respectively. For Gaussian blobs of type i (with $i = u,c$), these new molecular weights are given by the total number of Kuhn lengths per molecule, M_{Ki}, divided by the number m_{Ki} of Kuhn lengths per blob:

$$M_{Bi} = \frac{M_{Ki}}{m_{Ki}} \tag{18a}$$

$$M_{Ki} = \frac{M_i b_i}{b_{Ki}} \tag{18b}$$

where b_i is the size of monomer type i. Since the mobility of naked ssDNA is molecular size-independent, the charged blobs have the mobility μ_o. The electrophoretic mobility of the composite molecule then becomes:

$$\mu = \mu_o \frac{M_{Bc}}{M_{Bc} + M_{Bu}} = \mu_o \frac{M_c}{M_c + \alpha_1 M_u} \tag{19}$$

$$\alpha_1 = \frac{m_{Kc} b_u b_{Kc}}{m_{Ku} b_c b_{Ku}} \tag{20}$$

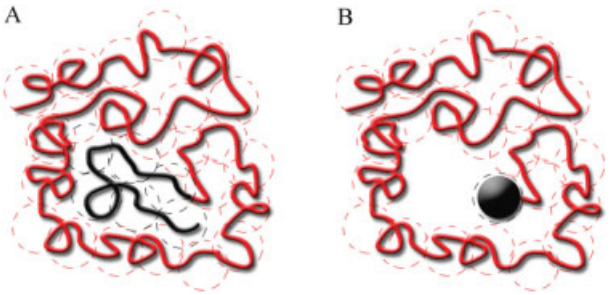

Figure 3. Schematic diagram showing the blob theory as applied to a ssDNA molecule (in red) attached to an extended polymeric label (A) or a spherical label (B). In both cases, we form blobs having identical hydrodynamic radii. The hybrid molecule is then seen as a chain of blobs; the latter thus play the role of effective monomers.

Clearly, the dimensionless parameter α_1 plays the role of the relative friction coefficient per uncharged monomer. Since the hydrodynamic radius of the charged and uncharged blobs must be equivalent ($R_{Hc} = R_{Hu}$), the expression for α_1 (which contains only molecular parameters) can be simplified. For Gaussian blobs, the relation $R_H \tilde{} m_K^{1/2} b_K$ reduces the equality $R_{Hc} = R_{Hu}$ to

$$\frac{m_{Kc}}{m_{Ku}} = \frac{b_{Ku}^2}{b_{Kc}^2} \tag{21}$$

Equation (20) then gives

$$\alpha_1 \equiv \frac{b_u b_{Ku}}{b_c b_{Kc}} \tag{22}$$

Since blobs are theoretical entities designed to help us solve the problem, the blob parameters are not present in the final result (Eqs. 19 and 22). Instead, Eq. (22) shows that the relative friction coefficient α_1 is the product of the relative monomer (b_u/b_c) and Kuhn (b_{Ku}/b_{Kc}) lengths. We also remark that Eq. (16) is only a special case of Eq. (19) valid when $\alpha_1 = 1$.

With this generalized ELFSE theory for random coil molecules, we have all the tools to calculate the critical electric field intensity E_o required for segregation to happen. To segregate the label from a long ssDNA, we need to apply a counter-flow (drag) force $F_o \geq 2k_BT / 3R_H$ on the composite molecule, where $R_H \cong \frac{2}{3}R_g$ is the hydrodynamic radius of the composite molecule in its coiled (un-deformed) state and $F_o = 6\pi\eta R_{Hc}\,(\mu_o - \mu)E$ [17, 26]. Using Eq. (19) we obtain:

$$E_o \cong \frac{k_BT}{9\pi\eta R_{H,u}^2 \mu_o} \cong \left(\frac{7}{R_{H,u}}\right)^2 \frac{\text{kV}}{\text{cm}} \tag{23}$$

where we used the typical values $T = 30°C$, $\eta = 1$ cP, $\mu_o \cong 3 \times 10^{-4}$ cm^2/Vs, and $R_{H,u}$ is the hydrodynamic radius of the label in nm. In [8], the streptavidin label had a radius of $R = 2.5$ nm and a read length of $M_c = 110$, thus leading to a critical field of $E_o \cong 8$ kV/cm, well beyond fields that are achievable on standard CE instruments. Interestingly, the value of the critical field does not depend, to first order, on the size of the DNA molecule. This means that either the DNA molecules are all segregated, or they are all in a single random coil conformation. With a label large enough to provide competitive ELFSE sequencing (say, $M_c > 400$ monomers), we would possibly be able to reach this segregated regime. The impact of this situation on sequencing will be discussed in Section 6.1. Until now, however, segregation (Fig. 2B) has not been observed experimentally. Of course, the next regime (stretched DNA, Fig. 2C – note that in the presence of a polymeric drag-tag, the latter could also stretch during migration) would require even higher fields, and it is safe to say that this is unlikely to be relevant for ssDNA sequencing (although it may be relevant for other applications). We thus conclude that the random coil regime (Fig. 2A) is currently the only relevant regime for analyzing ELFSE experimental data, and that one should thus use Eqs. (19) and (22). Although the artificial segregation regime (Fig. 2D) is probably present for very short ssDNA molecules attached to bulky labels, this is not of great interest since our goal is to find conditions that would allow the sequencing of very large ssDNA molecules using ELFSE; furthermore, as we shall discuss later, the most likely labels for achieving this goal are uncharged, extended polymer chains and not globular labels, such as streptavidin.

With Eq. (19) the expression giving the elution time becomes:

$$t(M) = \frac{L}{\mu_o E}\left(1 + \alpha_1 \frac{M_u}{M_c}\right) \tag{24}$$

There is no fundamental difference between this result and Eq. (17), except in the definition of the effective friction coefficient of the drag-tag; therefore, the same general conclusions still apply. In fact, since α_1 is generally smaller than unity (because DNA is a very rigid polymer), this expression indicates that very long ($M_u \gg M_c$) polymeric labels would be required to achieve long ssDNA read lengths.

In the case of a spherical label of radius R (like a folded protein or a colloidal particle), the effective friction coefficient would be given by $\zeta_u(R) = \alpha_1 M_u$; this means that ζ_u would then be interpreted as the number of ssDNA uncharged monomers that would form a random-coil blob with a hydrodynamic radius given by the label size R (see Fig. 3B). In principle, this allows one to predict the value of $\zeta_u(R)$ [17]; however, since R increases only as the cube root of the molecular weight for a compact object, it would be quite difficult to achieve large friction coeffi-

cients with such drag-tags. Hence, the blob theory suggests that extended polymeric labels are highly preferable.

Equation (24) predicts that a plot of t/t_0 (or, equivalently, μ_0/μ) *vs.* $1/M_c$, where $t_0 = \mu_0 E/L$ is the elution time of an unlabeled ssDNA molecule ($M_u = 0$), should be a straight line with a slope $\alpha_1 M_u$. This is the critical test of the theory and the easiest way to measure the drag-tag's frictional properties. Note that from the value of $\alpha_1 M_u$ one can estimate various molecular properties by using Eq. (22).

Finally, we can also predict the peak spacing for DNA fragments differing in size by one charged monomer by taking the derivative of Eq. (24) with respect to M_c:

$$t(M_c) - t(M_c + 1) \cong -\frac{\partial t(M_c)}{\partial M_c} = \frac{L}{\mu_o E}\frac{\alpha_1 M_u}{M_c^2} \tag{25}$$

According to Eq. (25), as the length M_c of the charged polyelectrolyte (*i.e.*, ssDNA) increases, the peaks get closer to each other, thus making sequencing more difficult. However, since diffusional band-broadening also decreases with size M_c, it is important to look at resolution and not only peak spacing. This is what we do next.

3.2 Diffusion and resolution

As ssDNA fragments migrate along the capillary, their motion is subject to thermally induced fluctuations that we will assume to be exclusively Brownian in nature (*e.g.*, we will neglect interactions with the walls). These fluctuations lead to diffusional band-broadening, governed by Eqs. (2) and (3) because the diffusion process is unaffected by electrophoresis [21]. A labeled ssDNA fragment, attached to an uncharged polymer of size M_u, behaves, as far as hydrodynamics is concerned, like a polymer with $M = M_c + \alpha_1 M_u$ effective monomers. Therefore, it is possible to write the size dependence of the diffusion coefficient of this Gaussian block copolymer as:

$$D(M) = \frac{D_1}{\sqrt{M_c + \alpha_1 M_u}} \tag{26}$$

where D_1 is essentially the diffusion coefficient of one ssDNA monomer in free solution. In order to estimate the maximum performance of ELFSE, we will assume that diffusion is the only source of band-broadening.

The size resolution factor $R(M = M_c + \alpha_1 M_u)$ will be defined as the ratio of the final full (temporal) width at half maximum of two consecutive peaks, $\sigma_{t\text{-FWHM}}(M) \cong \sigma_{t\text{-FWHM}}(M + 1)$, to the difference between their elution times $t(M)$:

$$R(M) = \frac{\sigma_{t-FWHM}(M)}{|t(M) - t(M-1)|} \cong \frac{\sqrt{8\ln(2)}\sigma_t(M)}{|\partial t/\partial M|} \tag{27}$$

where we assumed that we have Gaussian bands. This definition implies that R has units of bases: R simply gives the minimum difference in molecular size that can be resolved. For sequencing, we need $R \leq 1$ base. For normal diffusion, the spatial peak variance is given by

$$\sigma_x^2(M) = 2D(M)t(M) \tag{28}$$

if we assume that the initial injection zone width σ_o is negligible. With a peak velocity $v(M) = \mu(M)E$, the time width of the peaks is related to Eq. (28) *via* the expression:

$$\sigma_t(M) = \frac{\sigma_x(M)}{v(M)} = \frac{\sqrt{2D(M)t(M)}}{\mu(M)E} \tag{29}$$

We can now use these expressions to obtain the final result:

$$R = R_o \frac{M_c^{1/2}(M_c + \alpha_1 M_u)^{5/4}}{\alpha_1 M_u} \tag{30}$$

where $R_o = 4(\ln(2)D_1/\mu_o EL)^{1/2}$ is an experimental parameter, and the partial derivative in Eq. (27) was taken *vs.* M_c, the size of ssDNA molecule to be separated.

Equation (30) can be used to make some rough predictions regarding ELFSE's maximum read length M_c^{max} as a function of the size M_u of the label and the microscopic parameter α_1. Solving Eq. (30) for $R = 1$ base, with typical experimental values $D_1 = 7 \times 10^{-6}$ cm^2/s, $\mu_o = 3 \times 10^{-4}$ cm^2/Vs, $E = 300$ V/cm, and $L = 40$ cm, reduces the expression to the numerical evaluation of the nonlinear equation

$$M_c^{1/2}(M_c + \alpha_1 M_u)^{5/4} \cong 215\alpha_1 M_u \tag{31}$$

If we use the value $\alpha = \alpha_1 M_u \cong 30$, reported for streptavidin [17], the maximum predicted read length M_c^{max} is 129 bases, a number very close to the observed value of 110 $\pm$ 10 bases [9]. With a larger end-label, $\alpha = \alpha_1 M_u \cong 200$ for example, this predicts $M_c^{max} = 312$ bases; raising the field strength to $E = 1050$ V/cm would then give $M_c^{max} = 500$, an interesting read length for practical sequencing purposes. Clearly, although the situation is promising, significant experimental advances are required to make this prediction a reality.

4 ELFSE: experimental results

4.1 dsDNA separations: proof of concept

Heller *et al*. [8] were the first to report that ELFSE does indeed allow the separation of dsDNA fragments by free-solution CE. To add a drag-tag to the DNA fragments, these authors used the same method that had been pre-

viously used to study DNA migration and trapping electrophoresis in polyacrylamide gels and polymer solutions [28–30]. A DNA polymerase was used to fill in the overhanging ends of DNA restriction fragments with a biotinylated nucleotide, and a molar excess of streptavidin was then added [8]. The fact that the sequences at the two ends of the restriction fragments were different allowed the authors to label either one or both ends with streptavidin. The biotin-streptavidin conjugation is easily implemented and streptavidin is essentially neutral under standard CE conditions, thus providing an adequate drag-tag for proof-of-concept studies.

The results shown in Fig. 4 confirmed the predictions [7] that, in the absence of EOF, larger DNA fragments would migrate faster than smaller ones, and that higher resolution would be observed with a larger label or, in this case, with two labels instead of one (Figs. 4B and A, respectively). Using the original Mayer *et al*. ELFSE theory [7] to fit the data, it was estimated that adding a single streptavidin drag-tag generates a friction equivalent to about 23 uncharged nucleotides, whereas adding two streptavidins generates a friction equivalent to about 54 uncharged nucleotides (insert of Fig. 4A). Other results presented in [8] also confirmed the prediction that higher resolution is obtained at higher electric fields.

4.2 ssDNA sequencing using streptavidin: proof of concept

Although the experimental study of Heller *et al*. [8] on dsDNA fragments confirmed some of the predictions made by Mayer *et al*. [7], the separations obtained were relatively poor and certainly far from providing single-base resolution. This poor resolution was largely the result of the polydispersity of the streptavidin that was used (insert of Fig. 4B) and the imperfect coating of the capillaries used. The poor wall coating not only complicated the injection of the samples but also failed to suppress all interactions of the analytes (especially the streptavidin) with the capillary walls. A later study published by Ren *et al*. [9] showed that single-base resolution of ssDNA sequencing reactions could be achieved using both a gel-purified (relatively monodisperse) streptavidin label and an adequate dynamic, polydimethylacrylamide (pDMA) adsorption wall-coating process. These relatively simple improvements allowed these authors to read the first 100–110 bases of a sequencing sample, using 4-color LIF detection, in 18 min (Fig. 5). An analysis of the results showed that the system was essentially diffusion-limited. Consequently, this study also demonstrated that the size and properties of the drag-tag are the main limiting factors in ELFSE separations.

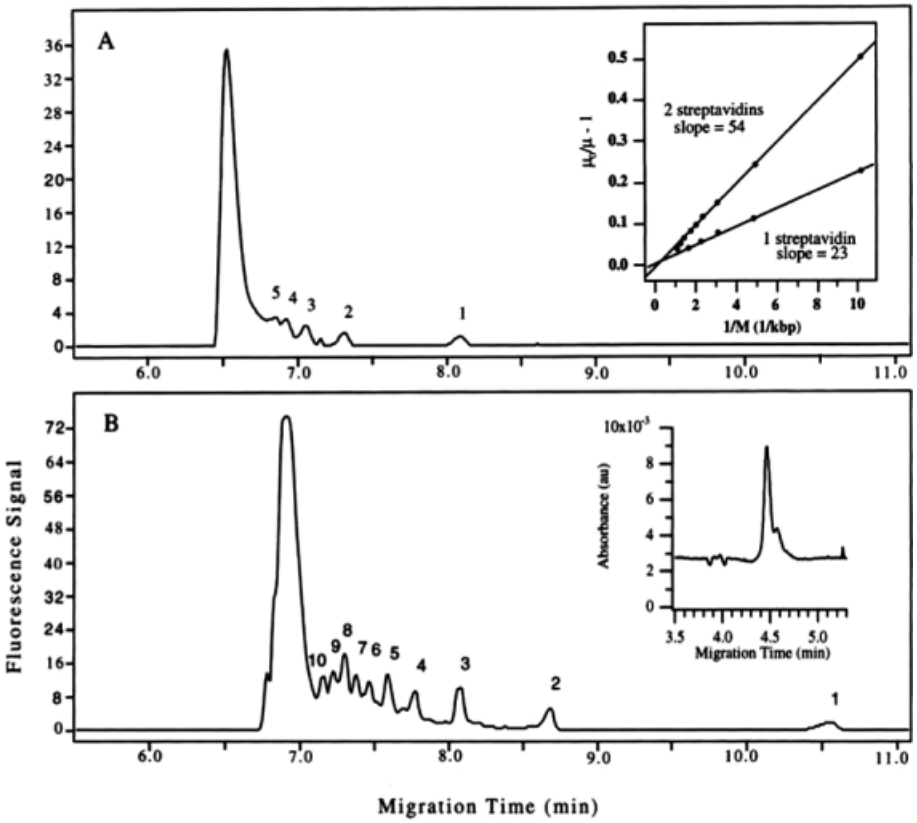

Figure 4. Separation of a 100 bp dsDNA ladder with (A) one or (B) two streptavidin molecules used as ELFSE drag-tags. The peaks marked 1–10 represent the M = 100 to M = 1000 bp dsDNA fragments, respectively. Insert of A: plot of μ_0/μ–1 *vs.* 1/M. Insert of B: polydispersity of the streptavidin as measured by CE. Reprinted from [8], with permission.

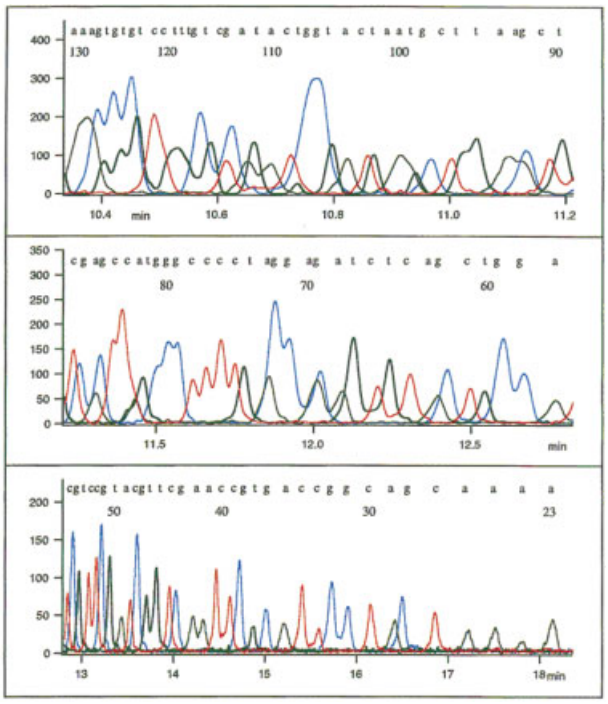

Figure 5. Separation of DNA sequencing reaction products using a purified streptavidin drag-tag and a dynamically coated capillary. The *Y*-axis represents the intensity of the fluorescence signal and the *X*-axis gives the elution time in minutes. The numbers above the peaks give the number of nucleotides (including the primer sequence). The weaker signal observed for the smaller fragments is due to their less efficient electrokinetic injection. Reprinted from [9], with permission.

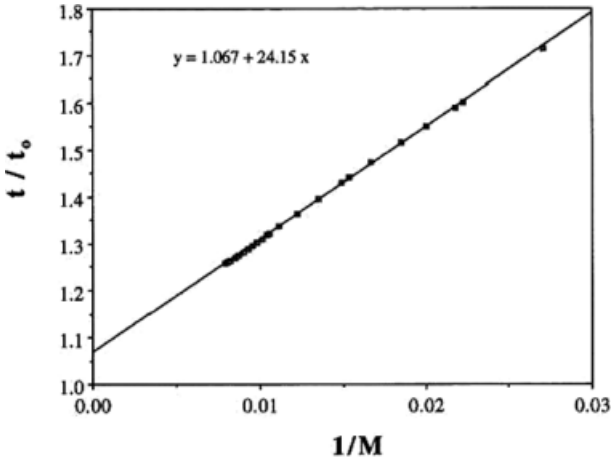

Figure 6. Ratio $t(M)/t_0$ *vs.* inverse ssDNA molecular size $1/M$, for the electropherogram presented in Fig. 5. Here, $t(M)$ is the elution time of the labeled ssDNA molecule of size M (in bases), while t_0 is the elution time of the unlabeled ssDNA molecules. The straight-line fit, predicted by theory, provides the effective friction coefficient of the streptavidin label, as $\alpha = \alpha_1 M_u = 24$ bases. Reprinted from [9], with permission.

Figure 6 shows the ratio $t(M)/t_0$ plotted *vs.* the inverse molecular size $1/M$ of the ssDNA molecules, where $t(M)$ is the elution time of a molecule of size M and t_0 is the CE elution time of unlabeled ssDNA molecules. According to the theory presented in Section 3.1, this figure should give a straight line with a slope $\alpha = \alpha_1 M_u$ according to the theory of ELFSE, where α is then the effective friction coefficient of the drag-tag (assuming that the latter is uncharged). The value obtained here, $\alpha = 24.15$ bases, is rather small. As discussed in Section 3.2, the number of bases sequenced is quite close to the maximum read length predicted for a label of this size in the diffusion-limited case. Based on their results, these authors further 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, being a natural (folded) protein, is not an ideal drag-tag because it is too small to generate hydrodynamic drag equivalent to more than about 25–40 nucleic acids. Since the friction coefficient of a compact sphere increases linearly with the radius of the sphere while the effective radius of a folded protein increases only as the 1/3 power of its molecular weight, a folded protein label with a friction coefficient equivalent to that of 300 uncharged ssDNA bases would be about 600 times heavier than streptavidin (more than 30 million Da), an unrealistic number. Therefore, [9] also demonstrated that folded uncharged proteins are not likely to be the drag-tags of choice for the development of a truly competitive sequencing device based on ELFSE.

4.3 Using polypeptoids and polypeptides

The sequencing of ~110 bases of DNA in free solution using streptavidin as an end-label provided an interesting proof of concept of ELFSE, but streptavidin is limited primarily by the modest amount of frictional drag it provides. Streptavidin is a moderately large protein (a total of 536 amino acids, or 52 kDa), but like most water-soluble proteins it is folded into a compact globular structure, which is inefficient for providing drag. Additionally, streptavidin shares several other properties common to natural proteins that limit its effectiveness for ELFSE. These include a tendency to interact nonspecifically with capillary or microchannel walls, leading to peak broadness, as well as an inherent electrical charge (except at the isoelectric point). A net positive charge, while in principle improving resolution in ELFSE, tends to exacerbate wall-analyte interactions (since the wall is negatively charged), and may also lead to deleterious ionic interactions between the drag-tag and the DNA, or between analytes (potentially leading to aggregation). A net negative charge

avoids these difficulties, but reduces the resolution of the separation by giving the label its own electrophoretic mobility in the same direction as that of the DNA. In addition, many natural proteins, with the exception of proteins from simple organisms such as *Escherichia coli*, have some degree of post-translational modification, and the resulting polydispersity leads to increased peak width (or multiple peaks for a single size of ssDNA) in ELFSE separations.

Avidin and streptavidin are relatively unique among proteins for their very strong affinity to biotinylated DNA, which is readily prepared and easy to work with. Other natural proteins would require more elaborate conjugation strategies, including the use of chemical cross-linking agents to link a reactive site on a protein (*e.g.*, an amino group) to a suitably modified DNA (*e.g.*, a 5'-thiol modification). Since most proteins have a multiplicity of reactive sites on their surfaces, guaranteeing unique end-on attachment of the protein to DNA is quite difficult. (Another possibility is the use of immunoglobulins to associate with a short peptide epitope appended to a DNA fragment. Although immunoglobulins have the apparent advantages of large size and relatively strong binding to the appropriate antigen, these proteins tend to be quite "sticky" and adsorptive in CE, with significant post-translational modification leading to hydrodynamic polydispersity).

Since it is difficult to identify large natural proteins to use as end-labels for ELFSE, alternative approaches have been explored by the Barron group. These approaches include (i) synthetic polymers, including conventional polymers such as PEG, as well as polymers synthesized using solid-phase techniques, and (ii) non-natural polypeptides genetically engineered specifically for use as end-labels for ELFSE.

Given the theoretical arguments above, synthetic polymers may appear initially to be promising candidates as frictional labels for ELFSE, although the requirement for complete monodispersity disqualifies the majority of large synthetic polymers. As will be discussed in Section 5.2, a PEG with a polydispersity index of only 1.01 (highly monodisperse, by synthetic polymer standards) yields a large family of ~110 closely-spaced peaks, or a broad smear, for each size of DNA [10]. This polydispersity renders any known synthetic polymer made by solution-phase methods unusable for high-resolution DNA sequencing separations.

In contrast to conventional solution-phase polymerization techniques, solid-phase synthesis has the potential to produce short polymers of precisely controlled length and sequence in high purity. Oligonucleotides and peptides are routinely synthesized by adding one monomer at a time to a growing chain covalently attached to a solid support. After adding all of the desired monomers in the desired sequence, the complete chain is chemically cleaved from the solid support, and purified to near-total monodispersity using reversed-phase HPLC or other chromatographic techniques. The drawback to using solid-phase techniques for synthesis of end-labels for ELFSE is that only relatively short chains can be synthesized with reasonable yield, *e.g.*, 30–40 amino acids for a peptide.

A novel class of polymers that has found application as frictional labels for ELFSE is the poly-*N*-substituted glycines, or polypeptoids [31]. Polypeptoids are a class of polyamides that resemble normal polypeptides, with a key difference: the side chains are appended to the backbone amide nitrogen rather than to the α-carbon, which eliminates the possibility of both intramolecular and intermolecular hydrogen bonding. Polypeptoids are easily produced by solid-phase synthesis using a "submonomer" protocol, meaning each monomer unit is built up in two separate steps. This submonomer approach allows the incorporation of a wide variety of side chain chemistries, including analogues of natural amino acid side chains, as well as other completely non-natural side chains. Polypeptoids bearing non-natural methoxyethyl side chains have been used for ELFSE [32, 33]. This side chain is uncharged, hydrophilic, and can act as a hydrogen bond acceptor but not a hydrogen bond donor. These properties tend to make polypeptoids bearing this side chain highly water soluble, but relatively inert to interactions with surfaces or other molecules [34] and thus ideally suited for frictional labels for ELFSE.

As with polypeptides, solid-phase synthesis of polypeptoids only allows the production of relatively short chains. Chains with up to 60 *N*-methoxyethylglycine monomers have been produced successfully [32, 35], although purification to monodispersity becomes more difficult, and synthetic yield drops for such long chains. As such, simple polypeptoids are unlikely to provide sufficient drag for long read-length sequencing. Analysis of polypeptoids ranging in length from 20 to 60 monomers suggests that, in a typical DNA sequencing buffer, the drag for a polypeptoid end-label increases linearly with its chain length, and that roughly 4–5 peptoid monomers provide the drag of a single base of ssDNA (which means that $\alpha_1 \cong 1/4$ to 1/5) [32]. This small ratio is due mostly to the fact that ssDNA is more rigid than polypeptoid chains. Thus, a M_u = 60-monomer long polypeptoid end-label provides drag equivalent to $\alpha_1 M_u \cong 13$ bases of ssDNA. Although DNA conjugated to polypeptoid end-labels provide sharp, clean peaks, the drag is simply not sufficient for high-resolution separation of large ssDNA. Note, however, that 120-monomer long polypeptoids would

provide an effect equivalent to that of streptavidin ($\alpha \approx$ 25–30), while the latter consists of 536 amino acids. This enormous difference in drag-per-monomer between the two compounds arises entirely from the fact that the polypeptoids are not folded, but instead are random coil. This demonstrates very clearly that the use of unfolded monodisperse polymer chains represent the best strategy to realize the promise of ELFSE.

Despite being too small for long-read length DNA sequencing by ELFSE, polypeptoid labels do have the potential for use in other ELFSE-based genetic analyses where single-base resolution is not required. Polypeptoids have found utility in single-base extension genotyping. Single-base extension or SBE (or "minisequencing") resembles conventional dye-terminator sequencing, except that only dye terminators (ddNTPs) are used, with no dNTPs. Thus, the SBE primer is extended by a single base only, and the fluorescent signature of the incorporated ddNTP reveals the identity of the base immediately adjacent to the 3'-end of the primer. A variety of approaches have been developed for the multiplexing of SBE reactions to allow rapid testing of multiple alleles. One approach, marketed by Applied Biosystems as "SNaPshot", has been to use several primers with different lengths of nonhybridizing "tail" regions, allowing the separation of multiplexed SBE reaction products in a single capillary filled with a conventional polymeric sieving matrix, such as POP4. Another approach employing injection from multiple samples separated by short intervals of electrophoresis has been marketed by Amersham as "SNuPe". An alternative technique using ELFSE was demonstrated by Vreeland *et al.* in 2002 [33]. In this approach, SBE primers were conjugated to polypeptoid end-labels of different sizes (M_u = 10, 20, and 30 *N*-methoxyethylglycine monomers). The polypeptoids functioned as drag-tags allowing separation and identification of the DNA products from a multiplexed SBE reaction by CE, without using a sieving matrix for the electrophoresis. The SBE-ELFSE approach was successfully used to genotype several mutations in "hot spot" loci in the *p53* gene amplified from cancer cell line samples. Using the polypeptoids as drag-tags eliminated the need to synthesize long nonhybridizing "tails" on primers for multiplexed SBE reactions, as done with the SNaPshot method, and also eliminated the need for the polymer sieving matrix. Analysis of SBE-ELFSE samples required less than 10 min on a MegaBACE DNA sequencing instrument, and could conceivably be performed in much less time on a microfluidic device. Additionally, the large peak spacing afforded by peptoid labels of 10, 20, and 30 monomers suggests that a larger number of more closely spaced peptoid labels could allow a greater degree of multiplexing in a single SBE-ELFSE reaction.

Since most synthetic polymers are unsuitable for DNA sequencing by ELFSE, either due to polydispersity (PEG and other synthetic polymers), or small size (polypeptoids and other solid-phase synthetic products), and most larger natural (folded) proteins are disqualified based on a variety of factors as discussed above, the Barron group has undertaken the development of non-natural polypeptides genetically engineered specifically for ELFSE. Desirable properties identified for such polypeptides include: (i) water solubility; (ii) charge neutrality (or very slight negative charge); (iii) absence of secondary or tertiary structure; (iv) minimal interactions with microchannel walls, (v) no post-translational modification, and (vi) amenability to unique end-on attachment to DNA.

In addition, "artificial" polypeptide sequences need to be expressed well by living organisms, and isolated in highly pure preparation from cell lysates. In order to facilitate rational design of a polymer with these properties, long polypeptides or "protein polymers" consisting of many repeats of simple amino acid sequences have been created. A novel "controlled cloning" technique was developed to allow the construction of large concatemers of repetitive DNA sequences with well-defined size [36]. These large, repetitive synthetic genes encoding the protein polymers are then transformed into *E. coli* and expressed as fusion proteins with an *N*-terminal polyhistidine tag, which allows for an affinity chromatography-based purification of the protein polymer from cell lysate. The polyhistidine tag can then be cleaved by chemical methods (treatment with cyanogen bromide and formic acid) or by enzymatic methods, with the chemical method being employed most commonly. Sequences designed thus far do not include lysine, leaving the amino terminus as a convenient reactive group for end-on conjugation to DNA. The bifunctional cross-linker Sulfo-SMCC has proven useful for conjugation of polypeptides *via* the amino terminus to DNA oligomers with 5'-thiol modifications.

Initial attempts to create protein polymer end-labels show the promise of this approach, although polypeptides produced thus far do have certain drawbacks. Several different sequences based on repeating sequences of seven amino acids were developed to explore the effect of varying the hydrophobicity and the intrinsic charge of the polypeptide. Two protein sequences that expressed well were tested extensively. These two sequences, referred to here as "P1" and "P2" have the following general structures, using the one-letter codes for the amino acids.

P1: [(GAGQG**S**A)$_n$G] (charge neutral, n = 12, 24, or 48)

P2: [(GAGQG**E**A)$_n$G] (negative charge, n = 18, 36, or 72)

Table 1. Effective friction (a) measured for two different protein polymer sequences: P1 = $(GAGQGSA)_nG$; P2 = $(GAGQGEA)_nG$, with different numbers of repeats (*n*) yielding the different lengths shown

Peptide sequence (number of amino acids)	Effective friction α (bases of ssDNA)
P1-85	15
P1-169	29
P1-337	70
P2-127	5.0
P2-253	7.1
P2-505	9.2

Data from Won [37]

Thus, P1 was produced in lengths of 85, 169, and 337 amino acids, while P2 was produced in lengths of 127, 253, and 505 amino acids. Each of these protein polymers was conjugated to fluorescently labeled, thiolated ssDNA oligonucleotides (22 bases in length) and analyzed by CE to determine the effective friction provided. Friction (a) was calculated using the simple model presented in Section 3.1 (Eq. 24) by comparing the mobility of conjugate peaks with the mobility of unlabeled DNA. Results are summarized in Table 1 [37]. As expected, the neutral sequence P1 was more effective than the negatively charged P2 in imparting drag to fluorescently labeled DNA oligos. Even with only one out of every seven residues in P2 carrying a negative charge, this polypeptide has a strong mobility of its own toward the anode in electrophoresis, of the same order of magnitude of the mobility of DNA. This can be explained in terms of counterion condensation and charge screening: DNA, with its highly charged backbone, has tightly bound counterions that screen much of the negative charge, leading to a greatly reduced apparent charge for the DNA backbone. P2 has a much lower charge density, and does not bind counterions so tightly, and thus has much less of a reduction in its apparent charge. Thus, P2 on its own migrates at roughly 60% the velocity of DNA, making P2 ineffective as a frictional label – even a 505-amino acid P2 protein polymer provides the apparent drag of only 9 bases of ssDNA (corresponding to a disastrous value of $\alpha_1 \cong 1/56$).

The charge-neutral P1 was much more effective as a frictional label. The 169-amino acid variant provided the friction equivalent of about 29 bases of DNA (or $\alpha_1 \cong 0.17$). This is of a similar magnitude to the friction obtained from streptavidin, a protein more than three times this size. Again, this indicates that an unstructured sequence, such as that of the P1 drag-tag, is much more efficient at providing drag than the globular streptavidin. The 337-amino acid variant provided the friction of approximately 70 bases of ssDNA, which could be sufficient for sequencing more than 200–250 bases of ssDNA. In addition, peak efficiencies for small DNA conjugated to P1 were found to be higher than those reported for streptavidin [37]. Although the CE instruments and experimental setups were quite different, this does suggest that the protein polymers produced thus far have minimal interaction with capillary walls, and thus satisfy that design criterion.

These P1 sequences do have an important drawback. The initial amino acid sequences that were proposed all included glutamine as a polar amino acid that would contribute to water solubility without contributing to charge. Glutamine is, however, prone to deamidation to form glutamic acid, and it was determined that deamidation of some glutamine residues is unavoidable, either during the expression of the protein or during the harsh chemical treatment to remove the *N*-terminal histidine tag. Thus, all samples of the P1 protein studied contained a certain fraction of molecules with one or more deamidated glutamines. Each of these different species showed up as separate peaks when conjugated to a monodisperse DNA, as illustrated in Fig. 7. This led to the use of ELFSE as a novel technique for quantitatively profiling the amount of deamidation in various P1 samples [38]; however, the complex pattern of peaks for each size of DNA makes this protein polymer unsuitable for DNA sequencing.

Efforts are currently underway to design new protein polymer sequences based on the lessons learned from the studies of the P1 and P2 sequences. These efforts will be described in Section 6.2.

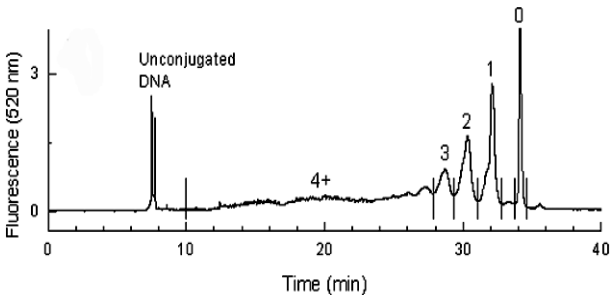

Figure 7. Capillary electrophoretic analysis of a 22-base fluorescently labeled DNA conjugated to a 337 amino acid protein polymer with the repeating "P1" sequence $(GAGQGSA)_{48}G$. The numbered peaks refer to conjugate species with 0, 1, 2, 3, 4 or more deamidated glutamines; this polydispersity renders the protein polymer unusable for DNA sequencing by ELFSE. Reprinted in part from [38], with permission.

5 Free-solution conjugate electrophoresis

The concept of ELFSE has also successfully lent itself to the creation of a complementary separation method called free-solution conjugate electrophoresis (FSCE). While ELFSE uses a uniformly uncharged drag-tag molecule to achieve the free-solution separation of a polydisperse polyelectrolyte sample, the complementary technique of FSCE uses a uniformly charged polyelectrolyte "engine" to separate a polydisperse, uncharged polymer sample [10, 27]. Since the polydispersity of a polymer sample can be difficult to determine accurately with conventional techniques, such as gel permeation chromatography [39] and mass spectrometry [40], this novel technique is quite interesting, and may find some useful applications. With this new method, each of the various sizes of uncharged polymers has the same charged polyelectrolyte engine attached to it, and hence the same electrical force; however, the effective drag of the uncharged polymers is directly proportional to their individual size, and hence the balance of the charge-to-friction ratio is indeed broken, thereby allowing for size separation of polymer by free-solution electrophoresis. Conjugates with shorter uncharged sections (*i.e.*, less frictional drag) have a higher electrophoretic velocity than those with longer uncharged sections; hence the resulting electropherogram yields the polydispersity of the sample.

This type of free-solution separation was employed by Stefansson and Novotny [41] to study uncharged oligosaccharides that have been complexed with charged tag molecules, and by Bullock [42] to separate uncharged poly(ethylene glycol) (PEG) polymers through derivatization with the charged small molecule phthalic anhydride. More recently, FSCE separations of PEG conjugated to ssDNA engines were performed by Vreeland *et al*. [10]; these well-controlled experiments, briefly mentioned above in Section 4.3, will be fully discussed in Section 5.2. FSCE is expected to be a useful means for separation of any water-soluble uncharged polymer that is amenable to uniform conjugation to a suitable charged engine, and which can be solubilized for CE analysis.

5.1 Theory of FSCE

The theory developed for FSCE separations [27], as with ELFSE, makes use of the polyampholyte theory developed by Long *et al*. [26]. Both components of the conjugates are assumed to be flexible, water-soluble linear polymers that can be adequately characterized as Gaussian coils (other situations, such as branched neutral polymers, could also be treated using a similar theoretical approach). The net mobility, arrival time and diffusion coefficients of these conjugates are also given by Eqs. (19), (24), and (26) in Section 3, where again the α_1 parameter is used to adjust for any differences in monomer size and stiffness between the charged and uncharged sections. Remarkably, Eq. (24) predicts that the peak spacing is constant in the FSCE case (*i.e.*, the elution time increases linearly with molecular size M_u), in agreement with the data reported in [10]. The FSCE size resolution factor $R(M)$ can be determined in the same way as with ELFSE, where the difference in arrival time of two adjacent peaks $t(M)$–$t(M–1)$, can be approximated as the derivative $\delta t(M)/\delta M \rightarrow \delta t(M)/\delta M_u$ of arrival time $t(M)$ with respect to the number of uncharged monomers (M_u) rather than charged monomers M_c, since FSCE conjugates only differ in the number of uncharged monomers. The calculations yield:

$$R(M) = R_0\sqrt{\frac{(M_c + \alpha_1 M_u)^{5/2}}{\alpha_1^2 M_c}} \tag{32}$$

Again, we have assumed optimal conditions where band-broadening is exclusively thermal-based, and the initial peak loading width is negligible (see [27] for a more general development). Note that the resolution depends only on the voltage drop, *i.e.*, the product $V = EL$, and not the length L of the capillary itself. Taking the derivative of Eq. (32) with respect to the number of charged monomers M_c, and setting it equal to zero gives the optimal engine size $M_c^* = \frac{2}{3}\alpha_1 M_u$ for a given polymer analyte of size M_u. The elution time for this optimal engine size would be:

$$t\left(M = M_c^* + \alpha_1 M_u\right) = \frac{5}{2} \times \frac{L}{\mu_0 E} \tag{33}$$

Remarkably the final elution time for conjugates with the optimal engine size (where we assume that the optimal engine size was set for the largest uncharged polymer to be separated) does not depend on the number of monomers or the nature of the polymers. As can be seen from Eq. (32), the size resolution factor increases with the size M_u of the uncharged polymer, which means that we lose the resolving power of FSCE for longer uncharged polymers. This loss of resolution is due to the fact that while the peak spacing remains constant, the elution time increases with molecular size M_u, thus leading to broader bands. In order to determine the longest uncharged chain size M_u^{max} that can be separated by FSCE, we simply set the resolution factor in Eq. (32) equal to one monomer $R(M_c + \alpha_1 M_u^{max}) = 1$. This gives the largest uncharged chain size that can be separated by FSCE as a function of and the engine size M_c:

$$M_u^{max} = \left[\frac{M_c^2}{R_0^4 \alpha_1}\right]^{1/5} - \frac{M_c}{\alpha_1} \tag{34}$$

This number M_u^{max} is effectively the "read-length" for a system under the experimentally ideal conditions of purely thermal band-broadening. If we use the typical values for CE experimental parameters as given in [27], we find that the read length for PEG conjugated to a 20-base ssDNA engine, as studied by Vreeland *et al*. [10], is roughly 10 kDa. The largest PEG samples studied in [10] were about 20 kDa, for which the resolution was indeed considerably less than for their 5 kDa sample.

5.2 PEG separation

The usefulness of FSCE for molar mass profiling has been demonstrated for synthetic polymers and solid-phase synthesis products. In the case of synthetic polymers, Vreeland *et al*. [10] conjugated monodisperse, fluorescently labeled DNA oligomers to commercially available "monodisperse" (PDI 1.01) PEG samples of various sizes by end-on covalent thiol-maleimide conjugation. The samples were then analyzed by free-solution CE in long (100 cm) capillaries with a 25 μm inner diameter. The "monodisperse" PEGs actually contain a distribution of sizes centered around the nominal molecular weight, and it was found that the FSCE technique was indeed able to separate the different sizes of PEG in a given sample, with single-monomer resolution. Because PEG is an uncharged molecule, the net charge for each conjugate was obviously identical (the DNA oligomer being uniformly charged); however, the molecular friction linearly varied with the length of the PEG tail, with each monomer unit contributing friction equivalent to $\alpha_1 = 0.138$ bases of ssDNA when the data were analyzed using the theory presented in Section 5.1. The degree of polymerization corresponding to each conjugate peak could thus be calculated, and the molar mass and polydispersity index of the PEG were determined from the areas of the conjugate peaks. Results from a typical FSCE molar mass profiling of PEG samples can be found in Fig. 8.

The FSCE method was also demonstrated on a microfluidic chip as well as by conventional CE. The microfluidic device has the potential for higher separation efficiencies and faster separations due to narrower injection zones. In the work presented by Vreeland *et al*. [10], the peak resolution on chips was significantly lower, but the authors predicted that, with optimization of the chip system for FSCE-like separations, single-monomer resolution can be achieved with shorter run times than by CE.

Several factors limit the effectiveness of the FSCE method for molar mass profiling. For example, the use of higher molar mass PEG ($>$ 5 kDa) or longer DNA oligomers (35 bases) decreased resolution by lowering the difference in mobility, in agreement with the theoretical

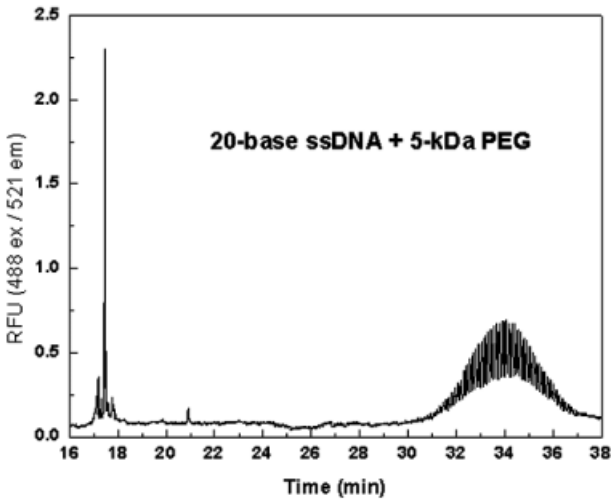

Figure 8. FSCE separation of a monodisperse, fluorescently labeled, 20-base ssDNA "engine" conjugated to a PEG with nominal molar mass of 5 kDa and a polydispersity index of 1.01. The series of peaks between 30 and 37 min represent PEG chains with different numbers of monomer units, conjugated to the ssDNA engine. Reprinted from [10], with permission.

predictions. In these cases, peak spacing is decreased, resulting in a nearly Gaussian global profile for the family of uncharged molecules, without individual peaks for each one of the monomers. It was shown that in the case of a higher molar mass PEG (20 kDa in average molar mass conjugated to 20-base DNA), deconvolution of the peaks to extract molar mass distribution information is possible using software such as PeakFit. Deconvolution is significantly more difficult in the case of longer DNA oligomers, where the spacing between peaks is smaller, so determining the optimal DNA oligomer size is critical for molar mass profiling by FSCE. Additionally, a bias exists when electrokinetic (EK) injection is used, since analytes with higher mobility enter the capillary disproportionately. In the studies with narrow-polydispersity PEG, the bias due to EK injection did not substantially alter the molar mass profile, but conjugate samples with higher polydispersity could display a significant bias, requiring a correction. Alternatively, a pressure injection can be used to introduce an unbiased sample. Finally, the FSCE technique is limited to cases where there is a method for unique, end-on attachment of the synthetic polymer to a monodisperse DNA, and for polymer-DNA conjugates with a solubility suitable for CE. Despite these limitations, the results of the FSCE experiments were found to be in agreement with molar masses and polydispersities for the PEGs determined by MALDI-TOF. In certain situations unsuitable for MALDI-TOF or gel permeation chromatography (GPC), FSCE provides a useful alternative method for molar mass profiling.

More recently, Vreeland *et al*. [35] used the same method to profile products from the solid-phase synthesis of polypeptoids, a task normally performed with RP-HPLC. Since each cycle of solid-phase synthesis is only 97–99% efficient, the synthesis of a long molecule inevitably results in a host of shorter "deletion" products, missing one or more monomer units. To profile this mixture of products by FSCE, the entire crude mixture of reaction products was conjugated end-on to fluorescently labeled ssDNA, and the resulting conjugates were separated based on differences in mobility due to the varying length of the polypeptoid. The polypeptoids in this work were also preparatively purified using RP-HPLC, and the resulting purified (and nominally monodisperse) polypeptoids were conjugated end-on to a monodisperse ssDNA oligomer. CE analysis of the resulting conjugates with LIF detection revealed the presence of significant levels of peptoid deletion fragments, which were undetectable by analytical RP-HPLC with UV detection. Using the FSCE technique for analysis rather than RP-HPLC, the average separation efficiency between adjacent peaks was 2.5 times higher and the average peak resolution was 5 times higher than that observed with RP-HPLC. These experiments highlight a potential high-impact application for FSCE, namely the characterization of therapeutic peptides or other pharmaceuticals synthesized by solid-phase synthesis, where precise characterization of sample purity is necessary for clinical or regulatory purposes.

6 The future of ELFSE and FSCE

6.1 Optimizing ssDNA sequencing by ELFSE

One possibility for optimizing the read length of ssDNA sequencing with ELFSE would be to take advantage of the dependence of the friction coefficient on drag-tag and/or DNA conformation. As discussed in section 3.1, under certain conditions, one or both of the two components of the conjugate molecule may become stretched. In a stretched conformation the hydrodynamic radius R_H increases, meaning that the friction is greater for a deformed molecule. Stretching ssDNA-streptavidin conjugates requires extremely high fields (see our estimates in Section 3.1). However, with a larger label providing more frictional drag, stretching would likely occur at attainable field strengths. A higher friction coefficient would increase the value of α_1, thereby improving the read length. One interesting feature of stretched conformations is that we have a positive feedback loop: if an ssDNA could pull against the frictional drag-tag with sufficient force, we would obtain a stretched drag-tag conformation with an increased friction coefficient, which in turn would lead to a greater extent of chain stretching, and so on. In this manner it may be possible to increase the separation achieved for longer ssDNA molecules, thereby extending the read length of ELFSE. However, there might be a secondary effect with a negative impact: larger DNA molecules may actually be slowed down more than smaller ones because they stretch the label more. The difference in drag-tag stretching generated by two ssDNA molecules differing in size by one base being quite small, we do not expect this effect to reduce the predicted gain in any substantial way. We are currently looking for evidence and studying the potential impact of this effect.

Another means of increasing the read length for ssDNA sequencing would be to optimize the ionic strength of the buffer. With an increase in the ionic strength there are more ions in the solution to screen the electrostatic repulsion between charges on the DNA backbone; hence the persistence length of DNA decreases. Since this would cause DNA to adopt a more compact shape, it would take more ssDNA bases to form a blob having the same hydrodynamic size as the end-label, with the result that the value of α_1 would increase. Hence, one way to increase α_1 is to increase the concentration of salt in the buffer (this is of course limited by practical considerations, such as the increased electrical current and Joule heating, and the possible impact of higher ionic strengths on the solubility of the DNA and the label).

A further route to optimizing the performance of ELFSE would be to employ a scanning LIF detection system, which can image the entire length of the microchannel in real time. Since the larger molecules are substantially harder to separate with ELFSE than smaller molecules with a given drag-tag, once these larger molecules arrive at the end of the capillary, all of the remaining smaller molecules are already separated; in fact, by the time they reach the end of the capillary they are needlessly overseparated [9]. Hence, as a means of decreasing the time required for the electropherogram to be produced, one could employ a scanning detection system rather than the finish-line detection system commonly utilized with CE. With this method, once the longest separated ssDNA arrive at the end of the capillary, the electric field could be turned off and the entire capillary scanned to produce the electropherogram immediately. In the case of [9], for example, such a device would have reduced the separation time by almost a factor of two, a non-negligible gain.

6.2 Designing new drag-tags

The primary obstacle preventing long read-length sequencing of ssDNA using ELFSE has thus far been the lack of a suitable, large, frictional end-label, or drag-tag. Gel-purified streptavidin provided only enough friction to sequence ~110 bases, and most other large natural pro-

teins are unsuitable for ELFSE or CE for a variety of reasons. Non-natural protein polymers are much more promising, but the repetitive polypeptide sequences produced thus far have suffered from unforeseen drawbacks, including the deamidation of glutamine leading to unacceptable polydispersity. Synthetic polymers made by solution-phase approaches, such as PEG, are also unacceptable due to their polydispersity, whereas monodisperse polypeptoids produced thus far are simply not large enough to provide the needed drag.

The most obvious next step in designing new labels useful for ELFSE sequencing is to produce large, non-natural polypeptides with different sequences. Neutral protein polymers with no unstable glutamine or asparagine residues are currently under development. Additional sequences with longer repeating units and a wider variety of amino acids, geared towards improved expressability, are also under development.

6.3 Charged oligosaccharides

In addition to separating ssDNA, ELFSE has also been successfully applied to the separation of highly charged oligosaccharides by Sudor and Novotny [43, 44]. They showed that, in the same manner as for separating ssDNA, these chains which normally have a length-independent charge-to-friction ratio, can be end-labeled in order to render the ratio length-dependent. The two types of labels they used, which serve the dual purpose of changing this ratio as well as providing the necessary fluorescence for detection, differ in nature in that one (8-aminonaphthalene-1,3,6-trisulfonic acid, or ANTS) increases the charge-to-friction ratio of the solute while the other (6-aminoquinoline) decreases it. With the former end-label, the migration order is from smaller to larger oligomers, while with the latter the order can be entirely reversed [43]. The resolution of small oligosaccharides, which is difficult to achieve with traditional CE, was dramatically improved with this ELFSE approach [43]. These authors also pointed out that ELFSE should be very useful for studying the affinity of highly charged oligosaccharide components with certain peptides, because it is free of any secondary equilibria of the oligosaccharides with "selective" counterions. ELFSE theory can be easily adjusted to account for the non-zero mobility of charged end-labels such as those employed by Sudor and Novotny.

7 Conclusions

ELFSE has already been shown to be useful for genotyping by single-base extension, an application for which it can offer easily multiplexing and fast analysis. ELFSE has also been used for the analysis of charged oligosaccharides, and the reverse process (FSCE) has been demonstrated for molar mass profiling of PEG and analysis of the products of a solid-phase synthesis reaction. ELFSE has also been demonstrated for separation of DNA sequencing fragments using a streptavidin end-label, although the read length was limited to about 110 bases.

The key to making ELFSE work is the drag-tag. For long-read sequencing, one must have a molecular "parachute" that is both large and monodisperse. By large, we mean that the selected object must generate enough friction to affect the net mobility of the DNA fragments in the range of molecular sizes under investigation. By monodisperse, we mean that the drag-tag must have a unique molecular weight and a unique friction coefficient (which is a function of the shape or conformation of the object in solution). Moreover, the selected drag-tag must have essentially no electric charge, be water-soluble, and have no affinity for the walls of the separation channel or for the DNA itself. Designing and producing such a molecule is not an easy task.

Our understanding of the physical mechanisms at play during ELFSE seems to be quite good at the present time. The theory described briefly in this review allows one to predict the dependence of the net mobility of an ELFSE conjugate upon the molecular size of the DNA. More importantly, however, it provides guidance for the development of new families of drag-tags. In particular, we showed how the available experimental results and our theoretical framework predict that the use of unfolded polymers is the preferred way to achieve large values of the effective total drag coefficient $\alpha = \alpha_1 \times M_u$. Equation (22) also indicates that stiff uncharged polymers would have very large values of the effective drag coefficient α_1. Hence, the ideal drag-tag would be a neutral polymer with a very large Kuhn length, one that exceeds the Kuhn length of ssDNA. One way to do this, employed in [17] and discussed briefly in Section 6.1, is to reduce the Kuhn length of ssDNA (the denominator of Eq. 22) by increasing the ionic strength of the solution (indeed, this paper showed that the α value of streptavidin can be increased by more than 30% simply by increasing the ionic strength of the buffer).

At this time, the University of Ottawa group focuses its attention on Molecular Dynamics computer simulations of the ELFSE process. Using this approach, we can follow the dynamics of the ssDNA, the drag-tag, the solvent and the ions in solution during electrophoresis at the molecular level. We are currently examining the design of new label geometries that would maximize the friction and hence the sequencing read length. We are also looking at issues like the behavior of rod-like or partially charged labels, the effects of the ends of the molecules on the net mobility, and the steric interactions between the ssDNA molecule and the drag-tag.

The Northwestern University group is currently focusing its attention on the design of new drag-tags, including new polypeptides and polypeptoids with sufficient friction and monodispersity for long read-length sequencing. We are also focusing on optimizing experimental aspects of ELFSE sequencing, including strategies for conjugating drag-tags to DNA, sequencing chemistry and reaction clean-up, as well as identifying appropriate conditions for capillary electrophoresis, including finding the best wall coating chemistry to minimize analyte-wall interactions.

The Northwestern group also plans to implement ELFSE on microfluidic chips. The narrow injection zones possible with microfluidic chips allow high-resolution separations over shorter distances than in conventional capillaries, allowing shorter analysis time. ELFSE seems particularly well-suited to implementation in microfluidic devices, as it eliminates the need for loading a viscous polymer matrix into narrow channels on chips that may not withstand high pressure. The short separation channels on microfluidic devices also allow the possibility of applying much higher electric fields than can be attained on commercially available capillary array instruments. Thus, microfluidic devices offer the possibility of testing or exploiting various theoretical predictions, including improved performance at very high electric fields, or the transition from the random coil regime to the hydrodynamic segregation regime. Microfluidic devices also offer improved flexibility for other experimental approaches, such as scanning detection, or integrating the sequencing reaction, clean-up, and separation on a single device. If the remaining technical challenges of ELFSE can be solved, it may have its greatest impact when implemented on microfluidic devices. DNA sequencing by ELFSE on microfluidic devices has the potential to replace sieving matrix-based sequencing by CE, if it can be fully optimized.

The work was supported, in part, by a University of Ottawa Admission Scholarship to LM, as well as the National Institutes of Health (NIH) of the USA (Grant No. NHGRI R01 HG002918-01) and Northwestern University. The findings, opinions, and recommendations expressed in this article are those of the authors and not necessarily those of Northwestern University, the University of Ottawa, or the NIH.

Received November 10, 2004

8 References

[1] International Human Genome Sequencing Consortium, *Nature* 2001, *409*, 860–921.

[2] Venter, J. C., *et al.*, *Science* 2001, *291*, 1304–1351.

[3] Viovy, J.-L., *Rev. Modern Phys.* 2000, *72*, 813–872.

[4] Doi, M., Edwards, S. F., *The Theory of Polymer Dynamics*, Oxford Science Publications, New York 1986.

[5] Slater, G. W., in: Heller, C. (Ed.), *Analysis of Nucleic Acids by Capillary Electrophoresis*, Vieweg & Sohn, Wiesbaden 1997.

[6] Noolandi, J., *Electrophoresis* 1992, *13*, 394–395.

[7] Mayer, P., Slater, G. W., Drouin, G., *Anal. Chem.* 1994, *66*, 1777–1780.

[8] Heller, C., Slater, G. W., Mayer, P., Dovichi, N., Pinto, D., Viovy, J.-L., Drouin, G., *J. Chromatogr. A* 1998, *806*, 113–121.

[9] Ren, H., Karger, A. E., Oaks, F., Menchen, S., Slater, G. W., Drouin, G., *Electrophoresis* 1999, *20*, 2501–2509.

[10] Vreeland, W. N., Desruisseaux, C., Karger, A. E., Drouin, G., Slater, G. W., Barron, A. E., *Anal. Chem.* 2001, *73*, 1795–1803.

[11] Teraoka, I., *Polymer Solutions*, John Wiley & Sons, New York 2002.

[12] Rubinstein, M., Colby, R. H., *Polymer Physics*, Oxford University Press, Oxford 2003.

[13] Mandel, M., Schouten, J., *Macromolecules* 1980, *13*, 1247–1251.

[14] Manning, G. S., *J. Phys. Chem.* 1981, *85*, 1506–1515.

[15] Hoagland, D. A., Arvanitidou, E., Welch, C., *Macromolecules* 1999, *32*, 6180–6190.

[16] Hermans, J. J., *J. Polymer Sci.* 1955, 527–534.

[17] Desruisseaux, C., Long, D., Drouin, G., Slater, G. W., *Macromolecules* 2001, *34*, 44–52.

[18] Stellwagen, N. C., Gelfi, C., Righetti, P. G., *Biopolymers* 1997, *42*, 687–703.

[19] Yeung, E. S., *Annu. Rev. Phys. Chem.* 2004, *55*, 97–126.

[20] Mathé, J., *PhD Thesis*, Louis Pasteur University, Strasbourg, 2003.

[21] 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.

[22] Long, D., Ajdari, A., *Electrophoresis* 1996, *17*, 1161–1166.

[23] Long, D., Viovy, J. L., Ajdari, A., *Phys. Rev. Lett.* 1996, *76*, 3858–3861.

[24] Long, D., Viovy, J. L., Ajdari, A., *Biopolymers* 1996, *39*, 755–759.

[25] Long, D., Viovy, J. L., Ajdari, A., *J. Phys. Condens. Matter* 1996, *8*, 9471–9475.

[26] Long, D., Dobrynin, A. V., Rubinstein, M., Ajdari, A., *J. Chem. Phys.* 1998, *108*, 1234–1244.

[27] McCormick, L. C., Slater, G. W., Krager, A. E., Vreeland, W. N., Barron, A. E., Desruisseaux, C., Drouin, G., *J. Chromatogr. A* 2001, *924*, 43–52.

[28] Desruisseaux, C., Slater, G. W., Drouin, G., *Macromolecules* 1998, *31*, 6499–6505.

[29] Ulanovsky, L., Drouin, G., Gilbert, W., *Nature* 1990, *343*, 190–192.

[30] Desruisseaux, C., Drouin, G., Slater, G. W., *Macromolecules* 2001, *34*, 5280–5286.

[31] Kirshenbaum, K., Barron, A. E., Goldsmith, R. A., Armand, P., Bradley, E. K., Truong, K. T. V., Dill, K. A., Cohen, F. E., Zuckermann, R. N., *Proc. Nat. Acad. Sci. USA* 1998, *95*, 4303–4308.

[32] Vreeland, W. N., Barron, A. E., *Abstr. Am. Chem. Soc.* 2000, *219*, 555–556.

[33] Vreeland, W. N., Meagher, R. J., Barron, A. E., *Anal. Chem.* 2002, *74*, 4328–4333.

[34] Ostuni, E., Chapman, R. G., Holmlin, R. E., Takayama, S., Whitesides, G. M., *Langmuir* 2001, *17*, 5605–5620.

[35] Vreeland, W. N., Slater, G. W., Barron, A. E., *Bioconj. Chem.* 2002, *13*, 663–670.

[36] Won, J. I., Barron, A. E., *Macromolecules* 2002, *35*, 8281–8287.

[37] Won, J. I., Meagher, R. J., Barron, A. E., *Electrophoresis* 2005, *26*, in press.

[38] Won, J. I., Meagher, R. J., Barron, A. E., *Biomacromolecules* 2004, *5*, 618–627.

[39] Barth, H. G., Boyes, B. E., Jackson, C., *Anal. Chem.* 1998, *70*, 251R.

[40] Marie, A., Fournier, F., Tabet, J. C., *Anal. Chem.* 2000, *72*, 5106–5114.

[41] Stefansson, M., Novotny, M., *Anal. Chem.* 1994, *66*, 1134–1140.

[42] Bullock, J., *J. Chromatogr.* 1993, *645*, 169–177.

[43] Sudor, J., Novotny, M. V., *Anal. Chem.* 1995, *67*, 4205–4209.

[44] Sudor, J., Novotny, M. V., *Anal. Chem.* 1997, *69*, 3199–3204.