uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Laurette McCormick
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Physics)
GRADE / DEGREE

Department of Physics
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Advancements to the Theory of Free Solution Electrophoresis of Polyelectrolytes

TITRE DE LA THÈSE / TITLE OF THESIS

Gary Slater
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

James Harden

Ruth Wilkins

Denis Rancourt

David Van Winkle

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Advancements to the theory of free solution electrophoresis of polyelectrolytes

Laurette McCormick

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN PHYSICS

UNIVERSITY OF OTTAWA
February 2006

Your file Votre référence
ISBN: 978-0-494-18597-1
Our file Notre référence
ISBN: 978-0-494-18597-1

The developmental process [of science] described in this essay has been a process of evolution *from* primitive beginnings–a process whose successive stages are characterized by an increasingly detailed and refined understanding of nature. But nothing that has been or will be said makes it a process of evolution *toward* anything. Inevitably that lacuna will have disturbed many readers. We are all deeply accustomed to seeing science as the one enterprise that draws constantly nearer to some goal set by nature in advance. But need there be any such goal? Can we not account for both science's existence and its success in terms of evolution from the community's state of knowledge at a given time? Does it really help to imagine that there is some one full, objective, true account of nature and that the proper measure of scientific achievement is the extent to which it brings us closer to that ultimate goal? [...] The entire process may have occurred, as we now suppose biological evolution did, without the benefit of a set goal, a permanent fixed scientific truth, of which each stage in the development of scientific knowledge is a better exemplar.

Thomas S. Kuhn, The Structure of Scientific Revolutions

Abstract

Capillary electrophoresis (CE) is the workhorse of countless analytical laboratories and is used routinely in various industries including pharmaceutical, forensic and clinical applications. Basically, CE is a method for separating charged molecular species in a buffer-filled capillary by the application of an electric field; the analytes move from one end of the capillary to the detector at the other end at speeds determined by their charge, size and shape. Generally, in free solution CE uniformly charged polyelectrolytes (such as DNA) are free-draining, meaning that their speed is independent of their size. Hence, until recently, a gel or other sieving medium has been necessary for the separation of polyelectrolytes; however, modifying uniformly charged polymers on the molecular level, via conjugation to uncharged polymers, allows for separation in free solution CE. In this thesis, advancements to the theory of free solution electrophoresis of polyelectrolytes, in particular, to the theories for two new free solution electrophoresis methods relying on conjugation, are presented. The first method, called End Labelled Free Solution Electrophoresis (ELFSE), can be used to sequence DNA, a negatively charged polymer in solution. Two different means of improving the resolution of ELFSE are predicted, one based on the molecular end effect, the other based on using a controlled electro-osmotic flow. In addition, a theory for the segregation of the DNA and label coils in ELFSE is presented. The second method is called Free Solution Conjugate Electrophoresis (FSCE); it allows for characterization of a sample of neutral polymers differing in length. The relevant theory, developed herein, elucidates how to accurately determine the molar mass distribution of the sample through FSCE measurements. In addition, supporting theories are developed that clarify the correct equation for the diffusion coefficient of molecules undergoing free solution electrophoresis, as well as illustrate that under ideal conditions, a viscosity gradient within the capillary serves only to decrease resolution and hence can not be used to improve performance. These theoretical studies constitute the six articles presented in this thesis. In addition, a comprehensive review article covering the development of ELFSE over the last decade, the theoretical concepts used to predict the ultimate performance of ELFSE for DNA sequencing, and the technological advances that are needed to speed the development of competitive ELFSE-based sequencing and separation technologies, is given in Appendix A. The predicted improvement in ELFSE resolution based on the end effect theory was also proven experimentally; the article with these findings is provided in Appendix B.

Sommaire

L'électrophorèse en capillaire, une méthode expérimentale d'analyse moléculaire fermement établie, s'avère vitale dans des domaines tels la pharmacologie, l'enquête médico-légale et la médecine. Il s'agit d'une technique qui permet de séparer des molécules chargées en solution au moyen d'un champ électrique externe; dans le capillaire, les analytes se déplacent d'un bout à l'autre du tube, vers le détecteur, à des vitesses déterminées par leur charge, leur grandeur et leur forme générale. Sous l'action du champ externe, un polymère uniformément chargé en solution (un polyélectrolyte, par exemple l'ADN) migre librement: sa vitesse est indépendante de sa taille. Par conséquent, l'utilisation d'un gel ou d'un autre type de tamis, pour contraindre la migration des molécules et ainsi arriver à les séparer, semble inévitable. Or, il se trouve qu'en modifiant ces polyélectrolytes, en particulier en les reliant à des polymères neutres, il devient possible de les séparer en solution libre, soit beaucoup plus rapidement. Nous poursuivons dans cette thèse les développements théoriques qui sous-tendent cette idée novatrice. D'abord, il est question de *End Labelled Free Solution Electrophoresis* (ELFSE), une méthode qui peut servir au séquençage de l'ADN. Nos prédictions indiquent qu'il est possible d'en augmenter la résolution d'au moins deux façons. La première repose sur l'importance accrue des bouts d'une chaîne de polymère quant à sa mobilité, alors que l'autre est fondée sur une utilisation contrôlée du débit électro-osmotique. De plus, nous présentons une théorie qui traite de la ségrégation de l'ADN et de son marqueur neutre durant la migration électrophorétique. Ensuite, nous portons notre attention sur une deuxième méthode, nommée *Free Solution Conjugate Electrophoresis* (FSCE), qui permet de caractériser un échantillon de polymères neutres comportant des chaînes de diverses longueurs. La théorie que nous proposons explique comment cette méthode permet de déterminer avec exactitude la distribution des masses moléculaires. De plus, nous présentons deux autres excursions théoriques connexes à cette recherche. La première clarifie l'équation du coefficient de diffusion de molécules soumises à l'électrophorèse en solution libre. La deuxième illustre qu'un gradient de viscosité, étant donné des conditions de séparation idéales, ne peut que nuire à la résolution et s'avère donc inutile pour améliorer la performance de la méthode FSCE. Outre les six articles scientifiques portant sur ces questions, la thèse comporte en annexe un article qui passe en revue le développement de la méthode ELFSE pour le séquençage de l'ADN, et qui détaille les progrès technologiques nécessaires au développement accéléré et compétitif de cette méthode. Finalement, nous incluons dans une deuxième annexe un article portant sur la démonstration expérimentale de notre prédiction théorique quant à l'amélioration de la résolution grâce aux effets de bout dans la méthode ELFSE.

Statement of originality

As far as I am aware, the research presented in this thesis is innovative. While I was primarily responsible for the research and writing of the first three articles (Chapters 2 through 4), with the critical support and guidance of my supervisor Gary Slater, the remaining three articles (chapters 5 through 7) as well as the two articles presented in the Appendices of this thesis were produced in collaboration with several co-writers, with my own involvement varying from writing small sections to being responsible for the bulk of the research. The article on Free Solution Conjugate Electrophoresis (Chapter 5), was researched and written together with my supervisor, and borrowed from an analogous theory written by Claude Desruisseaux for ELFSE. Experimentalists in the Annelise Barron group (Northwestern University) as well as Guy Drouin (University of Ottawa) and Achim Karger (Applied Biosystems) assisted with the comparison to experimental data. Claude Desruisseaux was responsible for the single stranded DNA (ssDNA) experimental data in the next paper, on the diffusion coefficient (Chapter 6), while I assisted my supervisor with the related analysis and writing. The double stranded DNA (dsDNA) sections were the responsibility of Bernard Tinland and coworkers. The following chapter on viscosity gradients, a project that I had started, is primarily the result of work undertaken by Steve Guillouzic. This paper was written by Steve; I derived the results presented in section 2. The review article provided in Appendix A was written in collaboration with Annelise Barron's experimental group at Northwestern University. The Ottawa group collaborated on the theoretical segments; I made contributions to sections 3.1, 3.2, 5.1 and 6.3. Appendix B presents the experimental confirmation of the end effects in ELFSE, predicted in the theoretical paper presented in Chapter 2. Robert Meagher of Northwestern University was primarily responsible for the experiments undertaken by the Northwestern University group led by Annelise Barron. Gary Slater and myself provided assistance with the data analysis.

Acknowledgements

This work would not have been possible without the financial support of the University of Ottawa, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the American National Institutes of Health (NIH). Throughout my doctoral work I have been fortunate to have the support of many wonderful fellow research group members, Jean-François, Fredéric, Michel, Tatek, Martin K, Sorin, Steve, Yannick, Martin B, Eric, Sébastien, Owen, Simona, Josée, Marc, Katerina, Justin, and Claude, who have been available for any scientific discussion, and furthermore, are great fun and made my time here enjoyable. I was fortunate to benefit from the expertise of Guy Drouin of the University of Ottawa Biology Department, Annelise Barron of the Department of Chemical and Biological Engineering at Northwestern University, and Bernard Tinland, while at the Centre National de la Recherche Scientifique, Institut Charles Sadron in Strasbourg, France. I am especially indebted to Fredéric Tessier whose expertise and generous offers of help have been invaluable. I am grateful for the unfailing support of my supervisor, Gary Slater, whose enthusiasm and depth of knowledge are second only to his dedication to the well-being of all his students. I am forever grateful to my parents, Dianna and Bert McCormick, who provided me with the foundation from which to take on anything I choose. Finally, I want to thank André Merizzi for his endless support, love and encouragement.

Table of contents

1

Introduction

Electrophoresis is the workhorse of countless analytical laboratories. In fact, it has been said that it is *the* major separation method in molecular biology [1]. Use of capillary electrophoresis (CE) in analytical chemistry research boomed in the late 1990's and its growth has continued through recent years, expanding in scope, both in terms of instrumentation and application; it is now being used routinely in various industries including pharmaceutical, forensic and clinical applications [2]. Basically, CE is a method for separating charged molecular species in a buffer-filled capillary by the application of an electric field; the analytes move from one end of the capillary to the detector at the other end at speeds determined by their charge, size and shape. Research into this technology is still very much an active field with many significant developments being unveiled at a rapid pace. While CE has traditionally relied on a gel or other sieving media, recent investigations into overcoming the need for such media via modifications at the molecular level have revealed an exciting new avenue for polymer characterization by electrophoresis. This new technique of free solution electrophoresis is the main subject of this thesis. This chapter provides a brief overview of polymer physics, an introduction to CE technology and free solution electrophoresis. A thorough review article on free solution electrophoresis of polymer conjugates (of which I was a co-author) is provided as a supplement to this introduction, in Appendix A. This chapter concludes with an outline of the article-based structure of the thesis.

Polymer Physics

A polymer molecule is a chain of repeating units, called monomers, that need not be identical. The number of monomers contained in a polymer is called the degree of polymerization; this number is usually quite large. In fact, it is not uncommon for it to be in the range of 10^4–10^5 [3]. The large number of monomers means that all polymers have very similar physical behaviour, as compared to small molecules, for their actual chemical properties play less of a role [3]. The polymers treated in this thesis are, by and large, all linear (as opposed to branched or cross-linked) in conformation. Since chemical properties take a secondary role in determining the behaviour of polymers, polymer physicists simplify them with models such as the bead-stick, wherein the chain is modelled by beads interconnected by sticks between which the angle can be completely free (called a freely jointed chain) or partially fixed, where the bond angle is commonly set but the dihedral angle is left free to rotate (called a freely rotating chain) [3].

Dispersed in solution, polymers take on a conformation far from extended and rod-like, but rather one more string-like[1]; if one were to drop a shoestring, the conformation it would assume on the floor would give an idea of a possible instantaneous conformation of a polymer in solution, albeit two-dimensional (a flexible thread model is another common polymer model). Clearly the conformation assumed by polymers in solution is quite random and can in fact be modelled by a *random walk* comprising a sequence of steps of fixed length b, where the orientation of each step is determined randomly. In this manner the conformation of a polymer is mapped out by random steps representing the positions of monomers or beads attached by sticks in the bead-stick model. (For the moment we neglect the fact that two monomers can not occupy the same place at the same time; this so-called excluded volume effect can be accounted for by the self-avoiding walk, to be discussed later.) The probability of the walk ending at a given position from the starting position is given by the binomial distribution [5]. For a very large number of total steps N, the final position x of a one dimensional random walk has an average of $\langle x \rangle = 0$ because a step in either direction occurs with equal probability, and a variance $\langle x^2 \rangle = Nb^2$. The central limit theorem of statistics guarantees that a sequence of random displacements that are small compared to the final mean displacement (or root-mean square displacement) will lead to a Gaussian distribution [5]. For a one dimensional random walk, the binomial distribution indeed converges to a Gaussian at large

[1] Solvents can be either good or poor; there is a strong attractive energy between good solvents and polymers, such that polymer segments prefer to be in contact with a good solvent over being in contact with other polymer segments and thus assume a large, random coil conformation, rather than collapse as they would in a poor solvent [4]. In this thesis we assume good solvent conditions.

N, with the probability for a final position of x being

$$P(x) = \frac{1}{(2\pi Nb^2)^{1/2}} \exp\left(-\frac{x^2}{2Nb^2}\right). \tag{1.1}$$

For the case of a three dimensional random walk, the probability for the final position being at vector position $\mathbf{h}$ from the starting point is

$$P(\mathbf{h}) = \frac{1}{\left(\frac{2}{3}\pi Nb^2\right)^{3/2}} \exp\left(-\frac{3\mathbf{h}^2}{2Nb^2}\right). \tag{1.2}$$

Hence, a linear polymer chain with a very large number of monomers N, each of size b, will have a distance between its two end monomers, called the end-to-end vector $\mathbf{h}$, that is on average zero,

$$\langle \mathbf{h} \rangle = 0, \tag{1.3}$$

and has a root-mean square magnitude of

$$\sqrt{\langle h^2 \rangle} = N^{1/2} b. \tag{1.4}$$

Along with the end-to-end distance h, the radius of gyration is a very useful physical property of polymers in solution, which measures the distribution of the monomers about the centre of mass. The radius of gyration can be defined as

$$R_g = \left\langle \frac{1}{N} \sum_i \mathbf{r}_i^2 \right\rangle^{1/2}, \tag{1.5}$$

where $\mathbf{r}_i$ is the position of the ith monomer relative to the centre of mass. For the ideal chain that we have been discussing so far, i.e. where there is no correlation between the steps, the radius of gyration has the same scaling as the end-to-end distance, and is given by the following:

$$R_g = \frac{N^{1/2} b}{\sqrt{6}}. \tag{1.6}$$

The ideal chain model, however, neglects the stiffness of a real polymer chain, a necessary result of bond angle limitations. The stiffness of a polymer chain is addressed by the worm-like chain model [6]. Also known as the Kratky-Porod model [6], it represents the polymer conformation by the vector position $\mathbf{r}$, which is a continuous function of the distance s along the contour of the chain from one end to the other. If we let the contour length of the polymer be $L_c = Nb$, then

s ranges from 0 to L_c. The orientation of a segment at contour distance s along the chain is given by the tangential unit vector $\mathbf{u}(s)$:

$$\mathbf{u}(s) = \frac{d\mathbf{r}(s)}{ds}. \tag{1.7}$$

In this model the stiffness of the chain results in the correlation between the orientation of two segments, say at s and s', which decreases exponentially with the curvilinear distance between them, $|s - s'|$:

$$\langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle = \exp\left(\frac{-|s - s'|}{L_p}\right), \tag{1.8}$$

where L_p is the persistence length, i.e. the curvilinear length over which correlations between the orientation of the segments persist [3]. The root-mean square of the end-to-end distance $\langle h^2 \rangle$, often written simply as h^2, (note that the simplified notation will be used interchangeably for the end-to-end distance average, i.e. h^2 for $\langle h^2 \rangle$) of a polymer represented by the worm-like chain model is [3]:

$$h^2 = 2L_p \left[L_c + L_p \left(\exp\left(-L_c/L_p\right) - 1\right)\right]. \tag{1.9}$$

For a very long and/or flexible chain $L_c \gg L_p$ and the end-to-end distance reduces to $h^2 = 2L_cL_p$. If the contour length is expressed as $L_c = Nb$ and the persistence length is rewritten as $L_p = b_K/2$, we find $h^2 = Nbb_K$, where b_K is known as the Kuhn segment length, another measure of the stiffness of the polymer. Simply re-expressing the contour length in terms of this Kuhn length, i.e. $L_c = N_K b_K$, where $N_K < N$ is the number of segments in the polymer of length b_K, recovers Eqs 1.3 and 1.4, i.e. the ideal chain statistics:

$$h^2 = N_K b_K^2. \tag{1.10}$$

Hence the stiffness of a real chain can easily be accommodated into the Gaussian, random coil model by a simple rescaling of the polymer into segments of the Kuhn length instead of monomer size, so long as its length is much greater than its stiffness, as measured by either L_p or b_K. As expected, in the other limit where the worm-like chain is short and/or very stiff, $L_c \ll L_p$, the end-to-end distance is simply $h = L_c$ [3].

The radius of gyration of the worm-like chain is given by the Kratky-Porod equation [3]:

$$R_g^2 = \frac{1}{3}L_pL_c - L_p^2 + 2\frac{L_p^3}{L_c}\left(1 - \frac{L_p}{L_c}\left[1 - \exp(-L_c/L_p)\right]\right). \tag{1.11}$$

For a very short and/or rigid chain ($L_c \ll L_p$), the radius of gyration can be approximated as $R_g = L_c/\sqrt{12}$, while for a very long and/or flexible chain ($L_p \ll L_c$), it can be approximated as

$R_g = \sqrt{L_p L_c/3}$. The latter can also be rewritten in terms of Kuhn segments using $L_c = N_K b_K$ and $L_p = b_K/2$:

$$R_g = \sqrt{N_K b_K^2/6}, \tag{1.12}$$

where again we recover the ideal Gaussian chain behaviour if the polymer is visualized in terms of Kuhn segment lengths instead of individual monomers (compare with Eq 1.6). Since the phenomena polymer physicists are concerned with commonly involve long, flexible polymers, Eqs 1.10 and 1.12 are often invoked.

A measure of the deviation from the freely jointed chain model due to the stiffness of a real chain can be obtained from the *characteristic ratio* defined as the ratio of the mean-square end-to-end distance of the actual chain to that of the corresponding freely jointed chain [7]:

$$C_N \equiv \frac{\langle h^2 \rangle_{\text{real}}}{\langle h^2 \rangle_{\text{f-j}}}. \tag{1.13}$$

For long chains, $C_N = C_\infty = N_K b_K^2/Nb^2$, and using $L_c = Nb = N_K b_K$, we find $C_\infty = b_K/b$, the expression used in Chapter 5. Polymers that are charged in solution, known as polyelectrolytes, are usually stiffer than uncharged polymers due to the extra long-range electrostatic contribution to their Kuhn length. An important natural polymer that is highly negatively charged in solution, and hence very stiff, is DNA. The stiffness of a polyelectrolyte depends on the ionic strength of the buffer, which is defined as

$$I \equiv \frac{1}{2}\sum_j c_j Z_j^2 \tag{1.14}$$

where c_j is the concentration of ion species j, of charge Z_j, and the summation extends over all the ions in the buffer solution [8]. A higher ionic strength (often given in mol/L) means a greater concentration of ions in the buffer and therefore more counter-ions to shield the charges of the polyelectrolyte from each other, such that the electrostatic repulsion between monomers of a charged polymer is screened and the electrostatic stiffness is reduced. Desruisseaux et al. [9] found a linear behaviour between the persistence length and ionic strength

$$L_p = L_{p0} + A/I, \tag{1.15}$$

where L_{p0} is the persistence length that a polymer would have were it uncharged, called the intrinsic persistence length, and A is a constant. For single stranded DNA (ssDNA) they found that $L_p \cong (1.6 + 0.28/I)$nm provided a good fit to their experimental data [9].

Real polymers do not only deviate from ideal polymers in terms of stiffness, but also in terms of excluded volume. With the ideal chain model, any conformation is available to the chain — the random walk can cross its own path. Clearly two sections of a real polymer cannot occupy the same space at the same time, a fact taken into account by the *self-avoiding walk*, where each step is chosen randomly, except that the path taken can not cross itself at any point. As could be expected, excluded volume interactions have the effect of swelling the conformation of the polymer chain. The swelling is more pronounced for longer polymer chains, for which the greater number of monomers exclude a greater volume from each other. Excluded volume can be sufficiently well accounted for by a further amendment to the ideal chain model wherein the exponent of 1/2 in the end-to-end distance, $\sqrt{\langle h^2 \rangle} = N^{1/2}b$ from Eq 1.4, is replaced by the so-called Flory exponent [3] of 3/5, giving, for a real chain (with non-negligible stiffness) that is sufficiently long and/or flexible ($L_\mathrm{p} \ll L_\mathrm{c}$)

$$\sqrt{\langle h^2 \rangle} \cong N_\mathrm{K}^{3/5} b_\mathrm{K}, \tag{1.16}$$

while the corresponding radius of gyration, from the ratio $R_\mathrm{g}^2/h^2 = 0.952/6$ given in [3], is approximately

$$R_\mathrm{g} \cong N_\mathrm{K}^{3/5} b_\mathrm{K}/6. \tag{1.17}$$

Whether or not excluded volume is negligible depends on both the polymer diameter, length and stiffness. A rough estimate for the critical size below which excluded volume is negligible, $N^* \approx b_\mathrm{K}^3/bd^2$, is presented in the review article in Appendix A. For single stranded DNA (ssDNA), this rough estimate yields a critical size of about 400 bases and it is hence suggested that for typical DNA sequencing (for which N = 100–1000 bases) excluded volume interactions are weak and can be neglected. As a result, for most of the calculations presented in this thesis, excluded volume is neglected and Eqs 1.10 and 1.12 are used in favour of Eqs 1.16 and 1.17.

The conformation of a polymer in solution is always changing due to collisions with solvent molecules, with the preferred conformation being that which maximizes the entropy. The entropy of a polymer with a given end-to-end distance h can be expressed in terms of its probability distribution from Eq 1.2 as $\mathrm{S}(h) = k_\mathrm{B} \ln P(h)$, giving

$$\mathrm{S}(h) = \text{const.} - \frac{3k_\mathrm{B}}{2Nb^2}h^2, \tag{1.18}$$

where the first term on the right hand side is a constant. The corresponding Helmholtz free energy

of the polymer, $A \equiv U - T\mathrm{S}$, where the internal energy U of the molecule is constant, is

$$A = \text{const.} + \frac{3k_\mathrm{B}T}{2Nb^2}h^2. \tag{1.19}$$

The chain conformations that minimize free energy are those that minimize h, hence the most favourable end-to-end distance for a polymer chain is zero. To maintain a non-zero end-to-end distance requires a force of $\delta A/\delta h = \frac{3k_\mathrm{B}T}{Nb^2}h$; this is equivalent to the force for a harmonic spring stretched to a distance h away from equilibrium, $F_\mathrm{sp} = k_\mathrm{sp}h$ where the spring constant is $k_\mathrm{sp} = 3k_\mathrm{B}T/Nb^2$:

$$F_\mathrm{e}(h) = \frac{3k_\mathrm{B}T}{Nb^2}h. \tag{1.20}$$

The subscript for the force is chosen to represent the entropic origins of the elasticity of the polymer chain, and the denominator is equal to the equilibrium square average of the end-to-end distance $\langle h^2 \rangle_0$. This entropic, spring-like force acting to resist deformation of a polymer from the random coil conformation with an end-to-end distance of zero is used in Chapter 4 to predict deformation during ELFSE. The spring-like nature of a polymer in solution led to the physical model called the bead-spring model, where the polymer is represented as a series of beads connected by springs [3].

Another important property of polymers is their diffusion. In free solution, the diffusion of a polymer is not affected by the presence of an electric field. This crucial point, which is studied in Chapter 6, means that the appropriate expression for the diffusion coefficient, under the conditions relevant to this thesis, is given by Einstein's expression for non-electrophoretic conditions [4]

$$D(N) = \frac{k_\mathrm{B}T}{\xi} \tag{1.21}$$

where the friction coefficient is given by Zimm's equation,

$$\xi = 6\pi\eta R_\mathrm{H}(N), \tag{1.22}$$

to yield

$$D(N) = \frac{k_\mathrm{B}T}{6\pi\eta R_\mathrm{H}(N)}. \tag{1.23}$$

The hydrodynamic radius of the polymer, R_H, is defined as the radius of a solid sphere that has the same diffusion coefficient as the polymer. When the polymer is moving, for example, under the influence of an electric field, all the monomers interact with, and partially drag the surrounding solvent molecules. These so-called hydrodynamic interactions build up so that the fluid inside the polymer coil essentially moves with the polymer [10]. As a result, the overall friction coefficient of

the polymer is less than it would be were all of its monomers unbonded and moving independently; rather than simply scaling as the total number of monomers N, it scales as $N^{1/2}$ (or $N^{3/5}$ when excluded volume interactions are important) [3]. The two characteristic radii (hydrodynamic and gyration) are similar, with the rough equation $R_{\mathrm{H}} \cong \frac{2}{3} R_{\mathrm{g}}$ being useful [4].

Before we look at capillary electrophoresis of polymers, it is interesting to note that most polymer mixtures are polydisperse, meaning that they are composed of polymers with different lengths. Although rare, some uniform molecular weight polymers do exist (often of biological origin) and are known as monodisperse. Polydisperse polymer samples can be characterized by the number average and the weight average molecular weight, defined respectively as

$$M_{\mathrm{n}} \equiv \frac{\sum_i M_i N_i}{\sum_i N_i} \tag{1.24}$$

and

$$M_{\mathrm{w}} \equiv \frac{\sum_i M_i^2 N_i}{\sum_i M_i N_i} \tag{1.25}$$

where N_i is the number of molecules of type i, with molecular mass M_i [11]. For truly monodisperse polymers $M_{\mathrm{n}} = M_{\mathrm{w}}$, however with polydisperse polymer samples $M_{\mathrm{n}} < M_{\mathrm{w}}$. The ratio of M_{w} to M_{n}, called the polydispersity index (PDI), is often used to characterize the polydispersity of a sample

$$PDI \equiv \frac{M_{\mathrm{w}}}{M_{\mathrm{n}}} = \frac{\sum_i N_i \sum_i M_i^2 N_i}{\left(\sum_i M_i N_i\right)^2}. \tag{1.26}$$

This ratio is equal to 1 for a completely monodisperse sample and increases from 1 with increasing polydispersity [3]. Even a PDI of 1.01, which is considered to be extremely low, represents significant polydispersity; for example, Figure 1.1 shows the range of molecular weights in a sample of poly(ethylene glycol) of nominal molecular weight 5000 g/mol (or, equivalently, 5000 Da in the common chemistry unit of Daltons). Although this sample has a polydispersity of only 1.01, over 50 different polymer lengths can be detected. This data was obtained by the powerful technique of free solution conjugate electrophoresis (FSCE), for which the relevant theory is presented in Chapter 5. The polydispersity present in most polymer samples, even those with low PDIs, presents a problem for the DNA sequencing technique End Labelled Free Solution Electrophoresis (the subject of Chapters 2, 3 and 4) which requires a near monodisperse polymer to act as a label, as discussed in the review article of Appendix A.

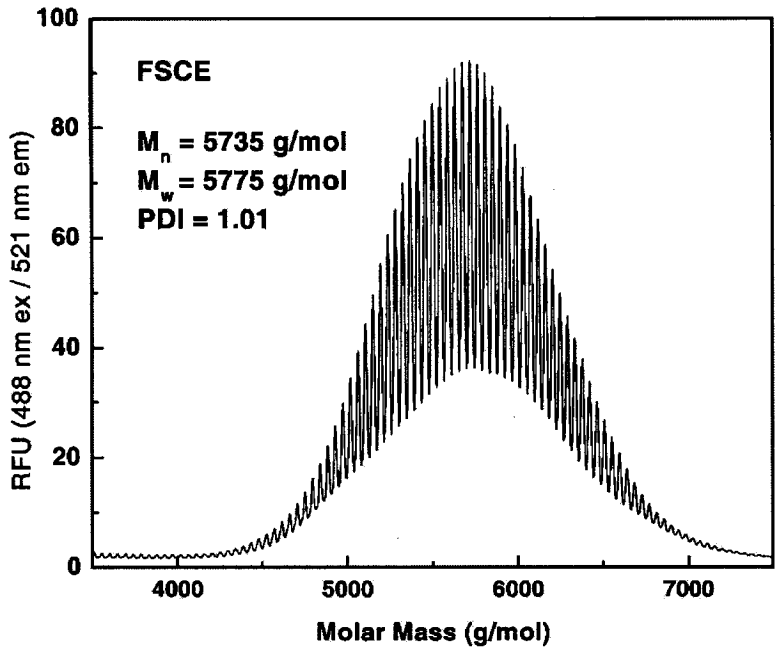


FIGURE 1.1 Electropherogram where the time axis has been transformed into a molar mass axis to show the range of molecular weights in a sample of poly(ethylene glycol) of nominal molecular weight 5000 g/mol. Although this sample has a polydispersity of only 1.01, over 50 different polymer lengths can be detected. This data was obtained by the powerful technique of free solution conjugate electrophoresis and is reproduced with permission from [12]. The relevant theory is presented in Chapter 5.

Capillary Electrophoresis

The main components of a capillary electrophoresis apparatus are the capillary, detector, data collection system, high-voltage power supply and means of temperature control, along with a personal computer to control the system [2] (see Figure 1.2). The fused-silica capillary has a length ranging from about 0.3 to 1 m, and an inner diameter of about 50 μm; with this inner diameter, a 50 cm long capillary would have a total volume of only 0.98 μL. An appropriate buffer solution[2] (so as to maintain the desired pH) fills the capillary and its ends are placed in reservoirs containing excess buffer as well as the electrodes, which are made of inert material such as platinum [14]. A sample volume, typically of 1–20 nL, is introduced at one end of the capillary by either pressure injection (hydrodynamic injection), or by the application of a voltage to move the charged species

[2] The buffer, which is simply an acid-base pair, is essential in capillary electrophoresis for maintaining the pH of the solution filling the capillary at the desired value. It functions by reacting with any extra acid or base; the equilibrium between the forward and back reactions, converting acid to base, or base to acid, is responsible for stabilizing the pH [13]. Two buffers commonly used in CE are Tris and Borate, which are effective at buffering in pH ranges 7.30–9.30 and 8.14–10.14 respectively [14].

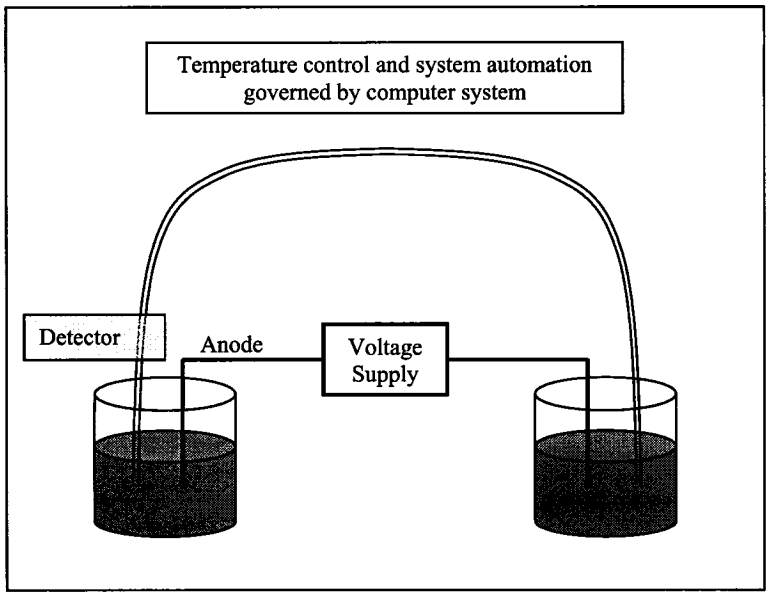


FIGURE 1.2 The main components of the capillary electrophoresis apparatus are the capillary, detector, data collection system, high-voltage power supply and means of temperature control, along with a personal computer to control the system.

into the capillary (electrokinetic injection), to form an analyte band with an initial width (or *loading width*) typically between 0.1 mm and 10 mm [2]. A voltage difference applied along the length of the capillary moves charged species from one end to the other. The speed of migration depends on the analyte's specific size, shape and charge. The difference in speed between the various species in the sample allows for separation so that they can be individually detected, in order of elution at, or rather just before, the other end of the capillary. The detector response is recorded to produce an electropherogram, which is simply a plot of detector response versus arrival time at the detector (see Figure 4 of the review article on page 79). Detection is commonly performed by laser-induced fluorescence (LIF); other methods include UV absorbance and electrochemical detection [14]. Typically, voltages range from 5–30 kV, resulting in currents ranging from 10–100 mA [14]. Given the substantial length of the typical capillary, the electric field strength is limited primarily by the power source; an applied voltage of 30 kV achieves a field strength of only 300 V/cm for a 1 m long capillary. (Microchips overcome this limitation with their shorter channels [15].)

The electrophoretic mobility, μ, of an analyte is defined as its speed, v, divided by the electric field strength E:

$$\mu = \frac{v}{E}, \tag{1.27}$$

and is commonly given in cm^2/Vs. The arrival time of an analyte at the detector is simply

$$t = \frac{L}{v}, \tag{1.28}$$

where L is the elution length, i.e. the distance from the injection end to the detector, which is thus slightly less than the total capillary length.

Clearly a greater difference in arrival time between consecutive bands (or peaks) corresponds to better resolution. This so-called peak spacing competes with band width in terms of resolution; all species within a sample are injected together in one thin band, and as each particular species moves along the length of the capillary, slowly pulling away from each other and forming individual bands, the bands widen. Various factors affect the final band width, including the initial band loading width, Joule heating (the production of heat due to friction between colliding ions and solvent moving under the applied field), interactions of analytes with the capillary wall, gradients in temperature or viscosity, any defects in the capillary, laminar flow due to fluid levels in the inlet and the outlet reservoirs differing significantly, and electrodispersion. (The latter can become important when the analyte concentration reaches that of the background electrolyte (BGE) and the conductivity of a band differs from that of the BGE; for example, if the mobility of the analyte is greater than that of the BGE, the band has a higher conductivity and lower field strength such that analytes near the edges of the band experience higher electric fields and accelerate, leading to peak asymmetry [14].)

The initial band width is usually negligible, in fact, great care is taken to reduce it as much as possible. Usually the sample volume is kept to a minimum, while still ensuring detection limits are met. It is also very common to use a phenomenon called stacking wherein the sample is in a low conductivity solution such that the local field strength is relatively high compared to the rest of the capillary. Due to the difference in the field strength, the analytes initially move fast and then slow down once they reach the background electrolyte filling the capillary, thereby concentrating, or stacking at the start of the capillary [14]. Stacking can similarly be achieved by having a jump in viscosity located at the beginning of the capillary such that the analytes move quickly in the lower viscosity medium and then slow down and concentrate at the beginning of the capillary where the viscosity is higher; the effect of viscosity gradients on resolution in capillary electrophoresis is investigated in Chapter 7.

Other factors contributing to the final band width are also minimized and can usually be controlled by careful experimental set up. The exception is diffusional band broadening; being

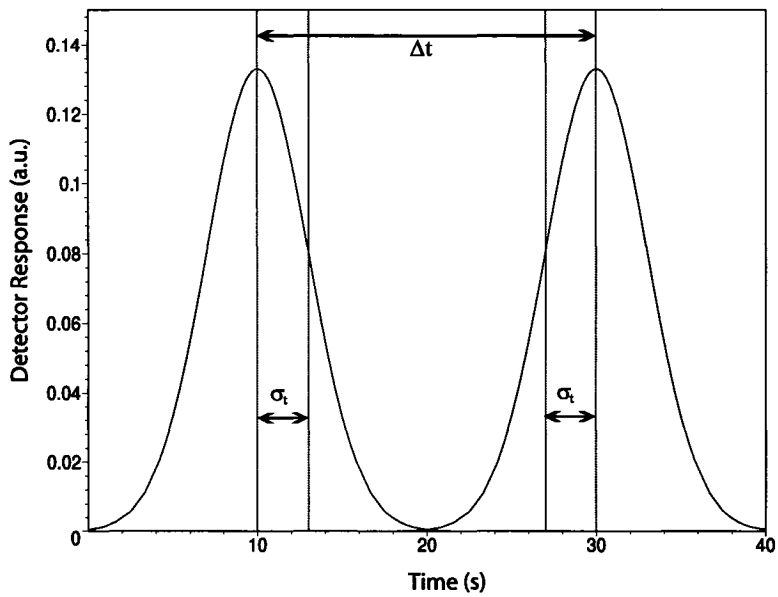


FIGURE 1.3 Two close-lying Gaussian peaks illustrating peak spacing Δt, and peak standard deviation σ_t. When the peak spacing is greater than or equal to the peak width as measured by the standard deviation, the peaks can be resolved.

due to thermal motion it is unavoidable. When diffusion limits the final band width and all other sources are negligible, the conditions are seen as optimal as far as band width is concerned.

As stated above, the peak spacing and band width compete in terms of resolution; we define the size-resolution factor as the temporal band width as measured by its temporal standard deviation σ_t (which can generally be taken to be essentially identical for two close-lying consecutive peaks since such peaks generally have similar properties and hence similar peak widths [5]), divided by the temporal spacing between the arrival of two consecutive peaks at the detector, Δt (see Figure 1.3). The temporal peak spacing is often taken as the derivative of the arrival time with respect to the number of monomers in the chain, N, giving the following expression for the size-resolution factor:

$$S = \frac{\sigma_t}{\frac{\delta t}{\delta N}}. \tag{1.29}$$

Defined in this way, the resolution factor has units of number of monomers and gives the smallest difference in the number of monomers that can be resolved; i.e. a factor of 1 is required for single-monomer resolution, which is needed for DNA sequencing. If one bears in mind that peak fitting software can be used to discern overlapping peaks, it is clear that even when the size-resolution factor somewhat exceeds the value 1, single-monomer resolution may still be obtained. Note

that since band width is usually measured in terms of the full width at half maximum, FWHM= $2\sqrt{2\ln 2}\sigma$, the size-resolution factor can be more conveniently defined by replacing the standard deviation by the FWHM. The temporal band standard deviation $\sigma_t = \sigma/v$, for the experimentally favourable diffusion-limited case can be calculated using the diffusion coefficient (given by Eq 1.23 for free solution conditions); the relationship between the spatial standard deviation of the peak and the diffusion coefficient is

$$\sigma = \sqrt{2Dt}. \tag{1.30}$$

Hence, the band width increases with the square root of time under these conditions while the peak spacing increases linearly with time, being equal to the difference in velocity of two consecutive bands, multiplied by the elapsed time. As a result, at the beginning of the electrophoretic separation, the band width increases more quickly than the peak spacing and there is a minimum amount of time required for the peak spacing to catch up to the band width, after which the peak spacing increases at a faster rate than the band width [5]. For this reason, a longer capillary achieves better resolution by allowing the analytes more time for their velocity differences to overcome diffusion. Care must be taken however to ensure that the wide peaks obtained with longer capillaries are still concentrated enough for the analytes to be detected.

With traditional electrophoresis, either a gel or other sieving medium such as an entangled polymer solution fills the capillary. The polyelectrolytes are then forced to snake around the gel fibres as they move under the influence of the electric field. Longer polyelectrolytes interact more with the fibres and are hence slowed down more than shorter polyelectrolytes; the result is an electropherogram with shorter polyelectrolytes being detected first, followed by longer and longer polyelectrolytes [16, 17]. This technique lends itself extraordinarily well to the sequencing of DNA, itself a negatively charged linear polymer in solution.

Briefly, the basic steps in sequencing a strand of DNA are the following. The strand is first separated from double-stranded form into single stranded form to allow for copies to be made. The desired starting location for the copying process is targeted by a small sequence of DNA known as a primer. Sets of copies are made in four different solutions, each solution containing normal nucleotides to build the copies, as well as terminating nucleotides that correspond to one of the four bases. These terminating nucleotides are labelled with a base-specific fluorescent dye to identify the chain that includes them as ending with either an A, T, C, or G base. Hence the four sets of solutions produce copies starting at the same location and having various lengths with an identifiable end base. When these copies are electrophoresed, they are detected in order from

shortest to longest, with the end monomer for each length being identifiable such that the sequence can simply be read directly from the electropherogram. Electrophoresis using gels or other sieving media has been responsible for the overwhelming majority of DNA sequencing. In this thesis we present some theoretical advancements made in our understanding of DNA sequencing, and other capillary electrophoresis separations, which do not require use of a sieving medium.

Free Solution Electrophoresis

As discussed above, generally a build up of hydrodynamic interactions between the monomers causes the fluid within the polymer to be dragged along with the polymer coil; this means that the overall friction coefficient ξ of the polymer scales like $N^{1/2}$ (or $N^{3/5}$ if excluded volume interactions are important). In free solution electrophoresis the situation is, however, somewhat more complicated, as detailed in the review article provided in Appendix A as a supplement to this introduction. The counter-ions in the buffer closely surround the charged polymer chain along its contour length. The electric field that pulls the charged chain in one direction, pulls equally hard on the counter-ions, but in the opposite direction. The moving counter-ions drag the fluid immediately surrounding the contour length of the chain with them, thereby screening the hydrodynamic interactions between the monomers[3]. As a result, the electrophoretic mobility in free solution is determined by a local balancing between the friction and electric forces. Since both the electrophoretic friction ξ_E, occurring along the contour length of the polymer, and the polymer charge Q (which determines the electric force F_E) scale linearly with the length, or the number of monomers N, the mobility is length-independent[4]: $\mu_0 \sim Q/\xi_E \sim N^1/N^1 \sim N^0$. Consequently the

[3] When an object of radius R_g moves in a fluid, the fluid is disturbed over a distance of about R_g, i.e. the momentum transfer to the fluid occurs over this distance. With an electric field affecting counter-ions however, the extent of the disturbance due to any electrophoretic motion is limited to that distance over which the concentrated counter-ions effectively balance the charge on the polymer chain, known as the Debye length, λ_D. If $\lambda_D > R_g$ then the disturbance still persists over R_g and the friction coefficient is unchanged by the counter-ions' response to the electric field. However, in electrophoresis, the Debye length is generally on the order of the monomer size b. The development presented in this thesis is hence for the case of the Debye length being small compared to the overall size of the molecule, $\lambda_D \ll R_g$. Details can be found in Appendix A.

[4] The approximate relation for the electrophoretic friction coefficient of a long and free-draining ($\lambda_D \ll R_g$) polyelectrolyte, $\xi_E \approx 3\pi\eta N_K b_K / |\ln(b_K/\lambda_D)|$, was suggested by Manning [18]. This overall electrophoretic friction coefficient was obtained by taking into account hydrodynamic interactions, which are screened by the counter-ions at distances greater than the Debye length, and by roughly estimating the friction coefficient of one Kuhn segment of the polymer chain, interacting with its counter-ion cloud, as $3\pi\eta b_K$. This indeed gives a length-independent mobility of $\mu \approx q_K |b_K/\lambda_D| / (3\pi\eta b_K)$, where q_K is the charge of each Kuhn segment.

electrophoretic mobility μ_0 of any long, flexible and uniformly charged polyelectrolyte in free solution is independent of its total length [17]. This so-called *free draining* property (i.e. fluid moves right through the coil such that the local friction determining the mobility scales with the contour length) is precisely why gels are used to sequence DNA, a long, flexible, negatively charged polymer. In a gel or other sieving medium, longer DNA collide more frequently with the gel fibres and are thereby selectively slowed down to provide separation by length [17].

While for mobility, the motion of the counter-ions effectively screens the hydrodynamic interactions such that the force due to the electric field and the corresponding force of friction, $F_{\mathrm{E}} \sim N$ and $\xi_{\mathrm{E}} \sim N$ respectively, balance each other, for any additional, non-electric force, (such as the random forces causing diffusion), the hydrodynamic interactions are not screened. This is because the counter-ions do not respond to non-electric forces and therefore do not affect the corresponding friction. Hence the friction coefficient for such an additional force would be given by that for non-electric conditions (Eq 1.22). This is indeed the case for diffusion, which is due only to fluctuations in the flow of the solvent. These fluctuations, which can be considered as an additional, random flow superimposed on the field-driven flow, are not affected by the electric field. Hence, the fluctuations in the flow, and the diffusion coefficient which depends on them, are subject to the hydrodynamic interactions between monomers which still exist for any non field-driven flow. Since the electrophoretic mobility depends on the flow of solvent, which depends on the electric field, while the diffusion depends only on the fluctuations in this flow, which are independent of the applied field, we can say that the polymer coil is free draining for the former, but not for the latter, as discussed in Chapter 6. This paper showed for the first time that the diffusion coefficient is not affected by the electric field in free solution, and hence the correct equation for the diffusion coefficient is given by Zimm's equation for non-electrophoretic conditions (Eq 1.23).

As a result of the free draining aspect of the electrophoretic mobility, in the absence of a gel (i.e., in free solution), DNA of all lengths co-migrate and no separation can be achieved (except for very short DNA, up to about 20 bases for ssDNA and 170 bp for dsDNA [19, 20]). Gels and other sieving media overcome this unfortunate phenomenon, however, avoiding the need for gels or other sieving media in DNA sequencing would be advantageous in several respects; for example, it would reduce the consumption of chemicals, as well as simplify the process and speed it up by removing the need for loading the sieving media into the capillary, a difficult and time consuming step.

There is another means of overcoming the free draining nature of DNA during electrophoresis to enable separation by length, one that does not require a sieving medium. When an uncharged molecule is attached at the end of the DNA chain, the identical scaling of friction and charge is undone. The uncharged molecule, called a label or *drag-tag*, adds a set amount of friction to all DNA chain lengths. Larger DNA, having more charges, are better able to pull the drag-tag, and hence should go faster than smaller DNA in free solution. This concept appeared in the literature in the early 1990's [21–23]; Mayer, Slater and Drouin [23] were the first to make quantitative predictions about this exciting possibility, which they called End Labeled Free Solution Electrophoresis, or ELFSE for short. Many researchers found this possibility very promising, however the complications such as finding and attaching an appropriate label, were considered intractable [24]. However, this very mechanism was proven valid in 1998 with the separation of a 100 bp ladder of dsDNA [25], see Figure 4 in Appendix A. The label molecule was streptavidin, a nearly neutral, globular protein, which can be easily attached to DNA. These initial results, while proving the concept of ELFSE, were inadequate for sequencing, since the resolution was limited to about 10 bases. The following year, as a result of purifying the streptavidin label to make it much closer to being monodisperse, and dynamically coating the wall to avoid streptavidin-wall interactions, dramatic results were achieved: about 100 ssDNA bases were sequenced in only 18 minutes [24] (See Figure 5 in Appendix A).

The review article in Appendix A gives the most comprehensive overview of ELFSE technology available, covering the development of ELFSE over the last decade, as well as 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. 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, is also reviewed. This review is provided as a supplement to the introduction presented here.

The biggest challenge to improving ELFSE performance is due to the fact that very long DNA chains have so much charge that their speed is barely affected by a small label and hence the peaks corresponding to longer DNA chains overlap badly and resolution is lost. This limits the maximum size of DNA that can be sequenced, known as the *read length*. Extending the read length of ELFSE is critical to enable it to be competitive with gel-based techniques; this is the main goal of current

ELFSE research. A larger label, which would result in a greater force of friction, would have a bigger effect on longer DNA and increase their resolution, thereby extending the read length. However, as discussed in Appendix A, it has been difficult to create a large, monodisperse label. Another means of extending the read length is discussed in Chapter 2. Before this paper, it had not been recognized that labelling both ends would lead to a greater friction than if the size of the label at one end were doubled. This is a result of the end effect predicted by the theory of Long and co-workers [26] for polyelectrolytes; our detailed analysis of this effect on ELFSE illuminated the potential of labelling both ends of the DNA instead of just one. The predicted improvement in ELFSE resolution based on the end effect theory was also proven experimentally; the article with these findings is provided in Appendix B.

Yet another possible means of extending the read length of ELFSE is presented in Chapter 3. Much like the moving counter-ions dragging the fluid surrounding the contour length of the chain with them, the counter-ions attracted to the negative charges of the capillary wall move in the opposite direction to the DNA, and drag the bulk of the fluid filling the capillary with them[5]. This process, called electro-osmotic flow (EOF) is commonly prevented by coating the capillary walls, however, in this paper we show how allowing a well-controlled amount of EOF could positively impact on ELFSE performance. It is predicted that this process could dramatically extend the read length, as well as decrease the time required for sequencing.

With next-generation labels providing greater friction, current ELFSE theory would need to be updated to take into account any possible deformation due to the strong drag forces of the label. Stronger electric field strengths would also be expected to encourage deformation. In Chapter 4 we predict the conditions under which the DNA and label may separate hydrodynamically from each other, and the resulting impact on ELFSE performance.

The following chapter develops the theory for the complementary separation technique known as Free Solution Conjugate Electrophoresis (FSCE), wherein a set of uniform DNA molecules is used as an electrophoretic engine to pull a set of uncharged polymers that vary in length. With this technique the polydispersity of the uncharged polymer can be measured. This chapter predates

---

[5] The fused silica capillary walls are negatively charged in solution and hence attract a cloud of counter-ions from the buffer. Under the influence of an electric field these charges, which are not bound to the wall as the negative surface charges are, move in the opposite direction as negatively charged DNA. Due to viscous drag forces, the bulk of the solution filling the capillary is dragged along with them in a plug-like flow. If the capillary is uniformly charged and both ends are at the same pressure, this counter-flow (which can drag DNA along with it) is constant [27, 28].

Chapter 2; the latter provides a critical revision to FSCE theory that takes the end effect into account.

Chapter 6, as mentioned above, demonstrates unequivocally that the Nernst-Einstein relation, despite its common use for free solution conditions, fails under these conditions. The proper relation (Eq 1.23) for free solution electrophoresis, which is critical for ELFSE and FSCE theories, is given.

Finally, Chapter 7 shows that a gradient in viscosity along the length of the capillary would not provide a means of improving resolution in free solution and hence could not be used to improve ELFSE resolution. However, under non-ideal conditions, when the initial band loading width is large, a viscosity jump located at the start of the capillary may be used (in addition to a jump in solution conductivity) to increase sample stacking.

Presentation of the thesis

The core of this thesis is composed of six articles published, accepted for publication, or submitted for publication (pending peer review) in scientific journals during the course of my Ph.D. degree. In addition, a review article (number 7 below) is provided in Appendix A to supplement the background information presented above, and the paper providing experimental support for the end effect in ELFSE, essentially the counterpart to Chapter 2, is given in Appendix B (number 8 below). The following list of articles constitute the next six chapters of this thesis, as well as the Appendices.

1) LC McCormick, GW Slater, *The molecular end effect and its critical impact on the behavior of charged-uncharged polymer conjugates during free-solution electrophoresis*, **Electrophoresis 26**, 1659–1667 (2005)

   End Labelled Free Solution Electrophoresis (ELFSE) and Free Solution Conjugate Electrophoresis (FSCE) theories were based on the predictions of Long and co-workers [26] for the electrophoresis of polyampholytes. In this paper we examine in detail the effect of one aspect of Long's theory that was previously neglected, that of the so-called end effect wherein monomers near the end of a polymer have a greater weight in the determination of the electrophoretic mobility. We found that this effect is more important than was previously assumed and impacts

on ELFSE performance; this led to Canadian and US patent applications for a new approach to ELFSE technology. The predicted improvement in ELFSE resolution based on the end effect theory was also proven experimentally; the article with these findings is provided in Appendix B. We also highlight the critical impact of the end effect on FSCE separations and show that caution must be used in interpreting FSCE results.

2) LC McCormick, GW Slater, *A Theoretical study of the possible use of electroosmotic flow to extend the read length of DNA sequencing by end labeled free solution electrophoresis*, **Electrophoresis**, in press.

In this paper we examine the effect on ELFSE performance of having a non-negligible electroosmotic flow (EOF), a flow which is normally suppressed in CE. We predict that the resolution can be dramatically improved with use of an optimal speed of EOF. This paper led to an invention disclosure filed with the University of Ottawa Technology Transfer and Business Enterprise Office. The EOF discovery, along with the new approach based on the end effect from Chapter 2, were combined in an application for the 1995 NSERC Innovation Challenge, an award aimed at fostering an appreciation of real-world applications of the research of masters or Ph.D. students who are in their final year or have recently graduated. My application was awarded an honourable mention.

3) LC McCormick, GW Slater, *Molecular deformation and performance of free solution electrophoresis of DNA-uncharged molecule conjugates: Theoretical predictions*, **to be submitted to Electrophoresis**

In this paper we investigate the behaviour of ELFSE molecules with next-generation labels and/or high electric field strengths. We make predictions as to the conditions leading to segregation of the DNA and label coils, as well as the expected impact on ELFSE performance. Optimal label architecture design and experimental conditions are suggested. This work also highlighted the potential gains in read length possible with high voltage use.

4) LC McCormick, GW Slater, AE Karger, WN Vreeland, AE Barron, C Desruisseaux, G Drouin, *Capillary electrophoretic separation of uncharged polymers using polyelectrolyte engines, a theoretical model*, **Journal of Chromatography A 924**, 43–52 (2001)

In this article we present the general theory for Free Solution Conjugate Electrophoresis (FSCE), a technique analogous to End Labelled Free Solution Electrophoresis (ELFSE) where instead of using a uniform set of uncharged drag-tag molecules to separate various lengths of DNA, a uniform set of DNA molecules are used as "engines" to separate various lengths of uncharged polymers in a sample.

5) AE Nkodo, JM Garnier, B Tinland, H Ren, C Desruisseaux, LC McCormick, G Drouin, GW Slater, *Diffusion coefficient of DNA molecules during free solution electrophoresis*, **Electrophoresis 22**, 2424–2432 (2001)

This seminal paper showed that the Nernst-Einstein equation relating the mobility and the diffusion coefficient during electrophoresis, while appropriate for most gel-based systems, is invalid for free solution electrophoresis. Since the diffusion coefficient is not affected by the electric field in free solution, the correct equation for the diffusion coefficient is given by Zimm's equation for non-electrophoretic conditions. This is a crucial result, impacting on predictions of ELFSE performance.

6) S Guillouzic, LC McCormick, GW Slater, *Electrophoresis in the presence of gradients: I. Viscosity gradients*, **Electrophoresis 23**, 1822–1832 (2002)

The effect of viscosity gradients within the capillary during electrophoresis are analysed in this paper. It is shown that such gradients only serve to decrease resolution under optimal conditions where the band width is limited by diffusion. When conditions are less than optimal and the initial band loading width is non-negligible, a gradient can be used to increase resolution, but is ideally located at the beginning of the capillary thereby serving only to increase sample stacking, a process commonly used to sharpen bands upon loading. Thus a viscosity gradient within the capillary can not be used to improve ELFSE performance.

7) RJ Meagher, J-I Won, LC McCormick, S Nedelcu, MM Bertrand, JL Bertram, G Drouin, AE Barron, GW Slater, *End-labeled free-solution electrophoresis of DNA*, **Electrophoresis 26**, 331–350 (2005)

Written in collaboration with Annelise Barron's experimental group at Northwestern University, this review article covers both ELFSE and FSCE technology, from a theoretical and experimental stand point. It is the most comprehensive introduction to ELFSE available.

8) RJ Meagher, LC McCormick, RD Haynes, J-I Won, JS Lin, GW Slater, AE Barron, *Free-solution electrophoresis of DNA modified with drag-tags at both ends*, **Electrophoresis**, in press.

This paper presents the experimental support for the end effect in ELFSE, predicted in the theoretical paper in Chapter 2. The prediction of enhanced drag arising from labelling both ends is confirmed, with 6-9% additional drag for the ssDNA and 10-23% additional drag for the dsDNA arising from labelling both ends than would be expected from simply doubling the size of the drag-tag at one end. The experimental results for ssDNA labelled at both ends are compared to the predictions of the theory (Chapter 2) of the end effect, with reasonably good quantitative agreement.

Other contributions

During the course of my Ph.D. studies I contributed to other articles published by the group, presented at various conferences, as well as filed patent applications for an innovation related to the end effect presented in Chapter 2.

Publications

1) GW Slater, M Kenward, LC McCormick, MG Gauthier, *The theory of DNA separation by capillary electrophoresis,* ***Current Opinion in Biotechnology 14*** (2003) 58-64

2) GW Slater, S Guillouzic, MG Gauthier, JF Mercier, M Kenward, LC McCormick, F Tessier, *Theory of DNA electrophoresis (~1999–2002 1/2): A Review,* ***Electrophoresis 23*** (2002) 3791-3816

3) GW Slater, Y Gratton, M Kenward, LC McCormick, F Tessier, *Deformation, stretching and relaxation of single polymer chains: fundamentals and examples,* ***Soft Materials 1*** (2003) 365–391. Also published as a chapter in *Soft Materials: Structure and Dynamics*, Marangoni and Dutcher Eds., Marcel Dekker, New York (2004) pp. 73-105

Conference presentations

1) GW Slater, LC McCormick, *Deformation of DNA and Polymer Labels during End-Labelled Free-Solution Electrophoresis,* ***to be presented at the American Physical Society March Meeting***, Baltimore, USA, 2006

2) LC McCormick, GW Slater, *A Theoretical Study of the Use of electro-osmotic Flow to Extend the Read-Length of DNA Sequencing by End Labelled Free Solution Electrophoresis,* ***American Physical Society March Meeting***, Los Angeles, USA, 2005

3) LC McCormick, GW Slater, *Novel DNA Sequencing Technique, Using Polymer Science and Microfluidics to Make it Competitive,* ***University of Ottawa Research Showcase, Advanced Technologies for Better Health***, Ottawa, ON, 2005

4) S Nedelcu, M Kenward, LC McCormick, GW Slater, *Theory of End-Labelled Free Solution Electrophoresis: Using Branched Polymeric Labels with ssDNA,* ***American Physical Society March Meeting***, Los Angeles, USA, 2005

5) LC McCormick, GW Slater, *Diffusion of DNA, a General Function to Describe all Regimes,* ***American Physical Society March Meeting***, Montreal, QC, 2004

6) LC McCormick, GW Slater, B Tinland, *Experimental Investigation of the Ogston Model for Electrophoresis Using Calibrated Matrices,* ***American Physical Society March Meeting***, Indianapolis, IN, USA, 2002

7) GW Slater, LC McCormick, S Guillouzic, *Viscosity Gradients and their Effect on Capillary Electrophoresis Resolution,* ***American Physical Society March Meeting***, Indianapolis, USA, 2002

8) GW Slater, B Tinland, A Ekani Nkodo, J-M Garnier, LC McCormick, C Desruisseaux, G Drouin, *The Diffusion Coefficient of DNA Fragments During Free-Flow Electrophoresis,* ***International Council of Electrophoresis Societies (ICES) Congress***, Verona, Italy, 2001

9) LC McCormick, GW Slater, AE Karger, WN Vreeland, AE Barron, C Desruisseaux, G Drouin, *Free Solution Conjugate Electrophoresis for the Determination of Polymer Solution Polydispersity,* ***Canadian Association of Physicists Annual Congress***, Victoria, B.C., 2001

10) LC McCormick, GW Slater, AE Karger, WN Vreeland, AE Barron, C Desruisseaux, G Drouin, *Capillary Electrophoresis Separation of Uncharged Polymers, a Theoretical Model,* ***University and Industry Opportunities in Polymer Physics***, Guelph, ON, 2001

Patent Applications

Methods for separation of polymeric compounds, Gary W Slater, Laurette C McCormick, Annelise E Barron, Robert J Meagher. US Provisional Application No. 60/615,600 filed October 5, 2004, US Application No. 11/241,990 and Canadian Application No. 2,523,089 filed October 4, 2005.

References

[1] C Heller. *Principles of DNA separation with capillary electrophoresis*, ***Electrophoresis* 22**, 629–643 (2001).

[2] KD Altria, D Elder. *Overview of the status and applications of capillary electrophoresis to the analysis of small molecules*, ***Journal of Chromatography A* 1023**, 1–14 (2004).

[3] I Teraoka. *Polymer Solutions, An Introduction to Physical Properties*. John Wiley & Sons, Inc., New York (2002).

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

[5] JC Giddings. *Unified Separation Science*. Wiley-Interscience, Toronto (1991).

[6] H Elias. *An Introduction to Polymer Science*. VCH, New York (1997).

[7] *Physical Properties of Polymers Handbook*, Edited by JE Mark. American Institute of Physics, New York, USA (1996).

[8] GM Barrow. *Physical Chemistry, fifth edition*. McGraw-Hill, Inc., Montreal (1988).

[9] C Desruisseaux, D Long, G Drouin, GW Slater. *Electrophoresis of Composite Molecular Objects. 1. Relation between Friction, Charge, and Ionic Strength in Free Solution*, ***Macromolecules* 34**, 44–52 (2001).

[10] M Kenward, GW Slater. *Molecular-dynamics simulations with explicit hydrodynamics – I: On the friction coefficients of deformed polymers*, ***European Physical Journal E* 14**, 55–65 (2004).

[11] H Elias. *Mega Molecules*. Springer-Verlag, New York (1987).

[12] WN Vreeland, C Desruisseaux, AE Karger, G Drouin, GW Slater, AE Barron. *Molar Mass Profiling of Synthetic Polymers by Free-Solution Capillary Electrophoresis of DNA-Polymer Conjugates*, ***Analytical Chemistry* 73**, 1795–1803 (2001).

[13] PH Raven, GB Johnson. *Biology*. Mosby Year Book, Toronto (1992).

[14] *Handbook of Capillary Electrophoresis*, Edited by JP Landers. CRC Press, Ann Arbor, USA (1994).

[15] RJ Meagher, JI Won, LC McCormick, S Nedelcu, MM Bertrand, JL Bertram, G Drouin, AE Barron, GW Slater. *End-labeled free-solution electrophoresis of DNA*, ***Electrophoresis* 26**, 331–350 (2005).

[16] GW Slater. *Electrophoresis Theories, pages 24–66 in: Analysis of Nucleic Acids by Capillary Electrophoresis*. C. Heller ed., Vieweg Verlag Press, Chromatographia CE Library, Volume 1, Wiesbaden, Germany (1997).

[17] JL Viovy. *Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms*, ***Reviews of Modern Physics* 72**, 813–872 (2000).

[18] GS Manning. *Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions. 7. Electrophoretic Mobility and Conductance*, ***Journal of Physical Chemistry* 85**, 1506–1515 (1981).

[19] NC Stellwagen, C Gelfi, PG Righetti. ***Biopolymers* 42**, 687–703 (1997).

[20] GW Slater, S Guillouzic, MG Gauthier, JF Mercier, M Kenward, LC McCormick, F Tessier. *Theory of DNA electrophoresis (~1999–2002 $\frac{1}{2}$)*, ***Electrophoresis* 23**, 3791–3816 (2002).

[21] J Noolandi. ***Electrophoresis*** **13**, 394–395 (1992).

[22] J Noolandi. ***Electrophoresis*** **14**, 680–681 (1993).

[23] P Mayer, GW Slater, G Drouin. ***Analytical Chemistry*** **66**, 1777–1780 (1994).

[24] H Ren, AE Karger, F Oaks, S Menchen, GW Slater, G Drouin. ***Electrophoresis*** **20**, 2501–2509 (1999).

[25] C Heller, GW Slater, P Mayer, N Dovichi, D Pinto, JL Viovy, G Drouin. ***Journal of Chromatography A*** **806**, 113–121 (1998).

[26] D Long, AV Dobrynin, M Rubinstein, A Ajdari. ***Journal of Chemical Physics*** **108**, 1234–1244 (1998).

[27] S Ghosal. ***Electrophoresis*** **25**, 214–228 (2004).

[28] D Sinton, C Escobedo-Canseco, L Ren, D Li. ***Journal of Colloid and Interface Science*** **254**, 184–189 (2002).

2

Molecular End Effect in Free Solution Electrophoresis

L.C. McCormick, G.W. Slater, *Electrophoresis* **26**, 1659–1667 (2005)

End Labelled Free Solution Electrophoresis (ELFSE) and Free Solution Conjugate Electrophoresis (FSCE) theories were based on the predictions of Long and co-workers for the electrophoresis of polyampholytes. In this paper we examine in detail the effect of one aspect of Long's theory that was previously neglected (and hence not taken into account elsewhere in this thesis), that of the so-called end effect wherein monomers near the end of a polymer have a greater weight in the determination of the electrophoretic mobility. We found that this effect is more important than was previously assumed and impacts on ELFSE performance; this led to Canadian and US patent applications for a new approach to ELFSE technology. The predicted improvement in ELFSE resolution based on the end effect theory was also proven experimentally; the article with these findings is provided in Appendix B. We also highlight the critical impact of the end effect on FSCE separations and show that caution must be used in interpreting FSCE results.

I was primarily responsible for the research and writing of this article, with the critical support and guidance of my supervisor.

Electrophoresis 2005, *26*, 1659–1667

Laurette C. McCormick
Gary W. Slater

Département de Physique,
University of Ottawa,
Ottawa, ON, Canada

The molecular end effect and its critical impact on the behavior of charged-uncharged polymer conjugates during free-solution electrophoresis

Recently two novel techniques using free-solution electrophoresis to separate charged-uncharged polymer conjugates have proven successful: end-labeled free-solution electrophoresis (ELFSE) for DNA sequencing, and free-solution conjugate electrophoresis (FSCE) for molar mass profiling of uncharged polymers. The approach taken to analyze the experimental data was an extension of the theory of Long and co-workers (Long, D., Dobrynin, A. V., Rubinstein, M., Ajdari, A., *J. Chem. Phys.* 1998, *108*, 1234–1244) for the electrophoresis of molecules with varying charge distributions. This theory also predicts that the ends of the polymers play a large role in determining the polymer's overall mobility; however, this aspect of the theory was neglected in previous work. Until now this "end effect" has, to the knowledge of the authors, not been recognized in experimental data. Through a careful investigation of the predicted end effect and a reanalysis of the experimental data, we demonstrate that indeed this effect critically impacts on the behavior of charged-uncharged polymer conjugates during electrophoresis. This work indicates that not only does the end effect need to be taken into account to avoid significant errors in data analysis, but also it provides novel system optimization approaches.

Keywords: DNA sequencing / Electrophoretic mobility / End-labeled free-solution electrophoresis / Free-solution electrophoresis / Hydrodynamic interactions DOI 10.1002/elps.200410276

1 Introduction

Much attention has been focused recently on the free-solution electrophoresis of charged-uncharged polymer conjugates, an approach that overcomes the need for gels or entangled polymer solutions for the electrophoretic separation of polyelectrolytes, while offering a means of molar mass profiling for uncharged polymers. In addition, this technique allows for a deepening of our fundamental understanding of the electrophoresis of polyelectrolytes and polyampholytes. End-labeled free-solution electrophoresis (ELFSE), for instance, was successfully used to sequence ssDNA up to 110 bases in less than 20 min [1]. This technique cleverly uses an uncharged "label" or "drag" molecule attached to each ssDNA chain in order to break the local balancing between friction and electric force [2–6] which normally leads to comigration of all ssDNA lengths [7, 8] (excepting very small fragments [9, 10]) in free solution. More recently, a complementary technique called free-solution conjugate electrophoresis (FSCE) has been used to characterize uncharged, water-soluble polymers that can be uniquely conjugated to ssDNA [11–13]. Here, the ssDNA chains are of uniform length, and act as engines to pull the varying lengths of uncharged polymers for electrophoresis leading to single-monomer resolution over a wide range of molecular sizes. In fact, the resolution obtained was approximately five times higher, and the separation efficiencies were increased by 150% compared to the more traditional RP-HPLC [12].

For both FSCE and ELFSE, the theoretical equation utilized for the overall mobility μ of the charged-uncharged block copolymer was until now a uniformly weighted average [5, 6, 11, 13]:

$$\mu = \mu_0 \frac{M_c}{N} = \mu_0 \frac{M_c}{M_c + \alpha M_u} \quad (1)$$

where M_c is the number of charged monomers each of mobility μ_0, and M_u is the number of uncharged monomers. This equation comes from a pioneer investigation of Long and co-workers [14] into the electrophoresis of polymers containing both charged and uncharged monomers. The factor α rescales M_u to account for the

Correspondence: Professor Gary W. Slater, Département de Physique, Université d'Ottawa, 150 Louis-Pasteur, Ottawa, Ontario K1N 6N5, Canada
E-mail: gslater@science.uottawa.ca
Fax: +613-562-5190

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

difference in hydrodynamic properties arising, for example, from the different persistence lengths (a measure of flexibility) of the charged and uncharged polymers. Hence, the α value depends on the chemistry of the molecules and varies with both temperature and buffer ionic strength (which affect the molecules' flexibilities). In fact, α enables a counting of uncharged units which have the same friction as one ssDNA monomer, such that the total number of effective monomers is $N = M_c + \alpha M_u$. The α value is an important determinant of the mobility since the frictional drag of the uncharged polymer is what selectively slows down longer conjugates in FSCE, and determines the read length of ELFSE.

The uniformly weighted average utilized for the mobility was an approximation that neglected certain second-order effects derived in [14], and in particular the so-called end effect wherein the mobilities of the monomers near the ends of a polyelectrolyte contribute with a greater weighting to its overall mobility. While the qualitative results for the range of data treated with the above approach were fairly good for certain molecular sizes, the inclusion of the end effect into the theory makes significant changes for the quantitative results, and how the theory can be utilized. In particular, we will demonstrate below that the previously utilized approximation would have resulted in unrealistic molar mass profiles had it been applied to a different range of polymer sizes in [11]. Hence the end effect must be carefully accounted for when using FSCE for molar mass profiling of synthetic uncharged polymers. We investigate in this paper the addition of the end effect to the theories of FSCE and ELFSE.

2 Theory of electrophoresis of polyampholytes

The electrophoretic behavior of polymers with inhomogeneous charge distributions was investigated by Long and co-workers [14]. The mobility of such chains was calculated as a function of charge distribution, taking into account both hydrodynamic interactions and the elasticity of the chain. They investigated the linear regime of small electric fields where the polymer chains remain in approximately Gaussian conformation, and assumed excluded volume effects to be negligible. For uniformly charged polymers, the counter-ions effectively cancel the long range hydrodynamic interactions between monomers, such that hydrodynamic and electric forces are balanced locally, leading to the well known "free-draining" phenomenon where uniformly charged polymers migrate at the same electrophoretic velocity despite their varying lengths [7, 8]. However, with nonuniformly charged polymers, it was shown that hydrodynamic interactions can play a large role. The general expression for the electrophoretic mobility of a polymer with a variable charge distribution was given as

$$\mu = \int_0^N \psi(n)\mu(n)dn \tag{2}$$

where $\mu(n)$ is the electrophoretic mobility of the n^{th} monomer, and N is the total number of monomers. The weighting function $\psi(n)$ is universal for sufficiently long polymers, *i.e.*, it looks the same for all sizes N beyond about ten persistence lengths in that $\psi(n) = 1/N\ \psi(n/N)$. We found that the numerical function $\psi(n/N)$ given in [14] is represented quite well by the following normalized interpolation function, shown in Fig. 1:

$$\Psi(n/N) = -0.65 + 0.62/(n/N)^{1/4} + 0.62/(1 - n/N)^{1/4} \tag{3}$$

We note that $\psi(n/N)$ increases substantially for monomers within the first and last ~8% of the chain (*e.g.*, these sections would account for 24% of the total weighting of the molecule, compared to the 16% expected by the uniformly weighted average approximation). This is a consequence of monomers located close to the ends of the chain spending more time, on average, closer to the surface of the coil, and hence affecting the overall mobility more than the middle monomers. As a result, the mobility is a weighted average of all individual monomer mobilities, where monomers in the middle have approximately the same weighting, but monomers near the end have a much greater weighting. This is the end effect which was neglected in previous ELFSE [5, 6] and FSCE [11, 13] analyses, where a uniform weighting, the dotted line in Fig. 1, was taken as an approximation (see Eq. 1).

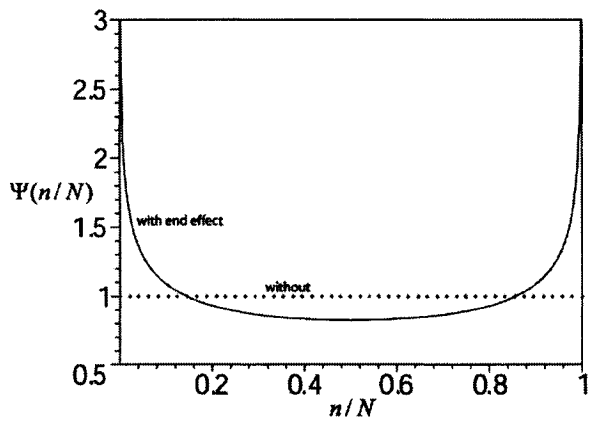


Figure 1. End effect weighting function (Eq. 3), an interpolating function that provides a good fit of the numerical curve presented in Fig. 2 of [14]. The dotted line is the uniform weighting approximation that was used previously [5, 6, 11, 13].

This effect may indeed be of importance when analyzing data for charged-uncharged block copolymers, especially if one of the blocks is relatively small (*e.g.*, less than 10% of the total polymer length) and hence has an unexpectedly large weighting determined solely by one of the "ends" of the curve in Fig. 1.

3 The end effect and FSCE

For the case of FSCE, where only the M_c charged monomers have a non-zero mobility, we can rewrite Eq. (2) as follows:

$$\mu = \int_0^{M_c} \frac{\mu(n)}{N} \Psi\left(\frac{n}{N}\right) dn \tag{4}$$

where we have started labeling the monomers from the charged end of the chain. The mobility of the n^{th} monomer $\mu(n)$, is simply the length-independent free solution ssDNA mobility μ_0, and the effective total number of monomers N is $M_c + \alpha M_u$ as before in the uniformly weighted average, such that

$$\mu = \mu_0 \frac{\int_0^{M_c} \Psi\left(\frac{n}{M_c + \alpha M_u}\right) dn}{M_c + \alpha M_u} \tag{5}$$

On comparison with Eq. (1) it is clear that taking the end effect into account involves replacing the numerator (M_c) with the integral of ψ over all the charged monomers (*i.e.*, replacing the uniform weighting $\psi = 1$ of which would give ${}^{M}\int_0 \psi dn = M_c$, with the ψ function of Fig. 1). As we can expect from the form of the ψ function, in going from a molecule that is completely charged to one that is attached to an uncharged chain, we lose the higher relative weighting of one of the charged ends and hence the end effect is manifested by an initial drop in the integral of ψ as M_u increases. However, as the uncharged segment grows quite large, the proportion of the conjugate molecule which is charged (M_c/N) decreases significantly, and the weighting for each of the charged monomer mobilities is determined solely by the remaining higher end weighting. Consequently, as the uncharged segment becomes much larger than the charged segment, the latter is given a higher weighting in the average determining the total mobility, thereby increasing the mobility over that expected by neglecting the end effect. This is indeed what we see in Fig. 2 when we plot the integral of ψ for the specific case studied by Vreeland *et al.* [11] of a 20-base ssDNA fragment (M_c = 20) attached to various lengths of poly (ethylene glycol) (PEG), for which α was estimated to be approximately 0.138 (to be discussed later). The integral in the mobility equation initially decreases for small PEG

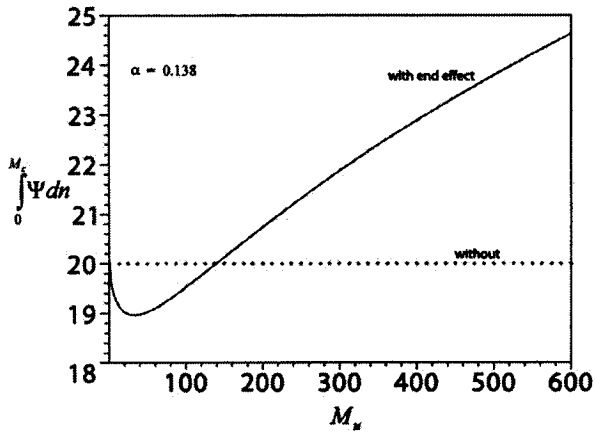


Figure 2. Integral of end effect weighting function ψ, from $n = 0$ to M_c, for FSCE with a charged ssDNA segment of $M_c = 20$ bases plotted as a function of the number M_u of monomers of PEG ($\alpha = 0.138$). Neglecting the end effect would give a constant value of 20, indicated here by the horizontal dashed line.

molecules, and then increases for the larger molecules. This factor grows well beyond the value of 20 previously taken as an approximation (neglecting end effects). For the longest PEG chains examined by Vreeland *et al*, which have a molecular mass of about 24 kDa (corresponding to about 550 monomers), we estimate that the integral of ψ is about 24, significantly higher than the previous approximation of 20.

The mobility of the conjugates varies not only with the weighting of the engine, but also with the total size: clearly molecules with larger uncharged segments move more slowly (this is the very means of separation). We take the mobility from Eq. (5) to find the arrival time of the molecule at the detector:

$$t = \frac{L}{\mu_0 E} \times \frac{M_c + \alpha M_u}{\int_0^{M_c} \Psi\left(\frac{n}{M_c + \alpha M_u}\right) dn} \tag{6}$$

where L is the length to the detector and E is the electric field intensity. Figure 3 shows how the arrival time (scaled by the constant $\frac{L}{\mu_0 E}$ which is the elution time of naked ssDNA, *i.e.*, for molecules with $M_u = 0$) depends on the end effect. When the end effect is neglected, we see a straight line (as reported by Vreeland *et al.* [11] for narrow ranges of PEG molecular size). However, taking into account the end effect results in a slightly higher slope for very small PEG segments, which decreases as the size of the PEG grows, becoming significantly less than it would be were the end effect not at play. As expected, the end effect gives a higher weighting to the charged engine

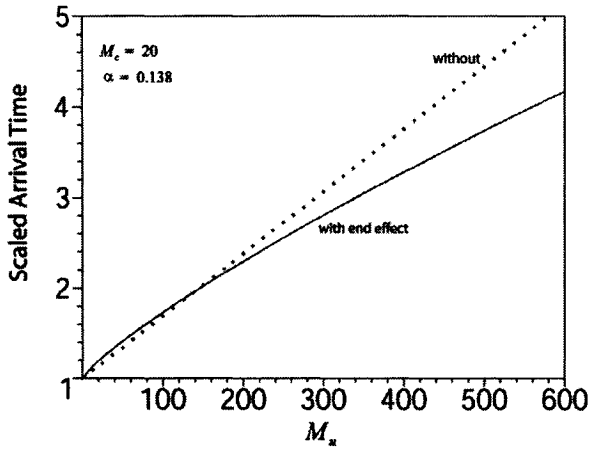


Figure 3. Predicted arrival time at detector (scaled by the constant $L/\mu_0 E$) for FSCE with an M_c = 20 base engine plotted as a function of the number M_u of monomers of PEG (α = 0.138). The solid curve is the case with the end effect taken into account, the dotted line would be expected were there no end effect. The lines cross at M_u = 140 PEG monomers in this example.

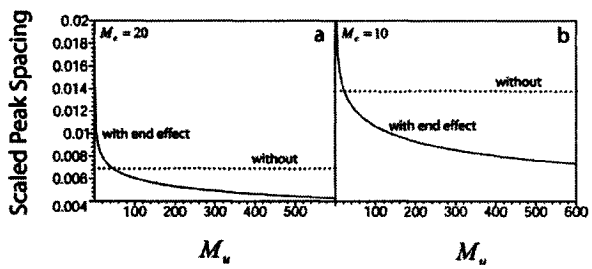


Figure 4. Predicted peak spacing (scaled by the constant $L/\mu_0 E$) for FSCE with (a) an M_c = 20-base engine, and (b) an M_c = 10-base engine, as a function of the number M_u of monomers of PEG (α = 0.138). The solid curve is for the case with the end effect taken into account, the dotted line would be expected were there no end effect.

such that molecules (having more than 140 PEG monomers in this example) go faster than if the end effect is neglected, and increasingly so for larger conjugates where the engine weighting is pushed further to the left on Fig. 1.

Unfortunately this increased speed has a negative impact on separation: for the same separation length L and field intensity E, the molecules have less time for their differences in speed to slow one relative to another. The predicted temporal peak spacing $\left|\frac{\partial t}{\partial M_u}\right|$ is shown in Fig. 4 for both an ssDNA engine size of M_c = 20, and one of size of 10, which was previously predicted to be the optimal engine size [13]. Without end effects we would expect a horizontal line (one of the most interesting features of FSCE); however, with end effects we see that peak spacing decreases with increasing conjugate size. For the larger molecules studied by Vreeland *et al.* (around 550 monomers conjugated to a 20-base DNA engine), we estimate that the end effect reduces the peak spacing to only 63% of that expected were there no end effect (see Fig. 4a). This decrease is even more pronounced for shorter charged segments; for an ssDNA engine size of M_c = 10 (Fig. 4b), we predict that the peak spacing for the larger molecules would drop to a mere 54% of that previously expected. Note that even though the end effect plays a more detrimental role for the shorter engine, the overall peak spacing is still higher, as expected by our previous work [13]. As well, it should be noted that for conjugates with small uncharged segments, the end effect could be exploited as it actually leads to an important increase in separation under these conditions (see Figs. 4a and b).

Now we illustrate the manifestation of the end effect in the published FSCE experimental data [11], which previously went unnoticed. The decrease in the slope of arrival time (Fig. 3) is slow, hence over a small range of sizes the size-dependence of the arrival time could easily appear to be linear; this was indeed what Vreeland *et al.* reported [11]. The measured arrival times were linear for both PEG molecular size ranges, the smaller sizes ranging from approximately 4.5 kDa through 7 kDa (corresponding to about M_u = 100 through 160 PEG monomers) and the larger ranging from about 20 kDa through 24 kDa (about M_u = 450 through 550 PEG monomers). As previously mentioned, the approach taken for the data analysis was to neglect the end effect by assuming the ψ weighting function to be uniform (see Eq. 1). Hence by neglecting the ψ dependence on αM_u, this term could be isolated from the mobility expression, $\alpha M_u = M_c\left[\mu_0/\mu - 1\right]$ and plotted as a function of peak number (which varies linearly with the number of PEG monomers M_u since FSCE yields single monomer resolution). The slope of this plot, which is basically a scaled arrival time, was then simply taken to be α. This value was then used to calculate the molar masses of both samples since it should not depend on the length of the polymers, rather just their individual monomer lengths and flexibilities. As we can see from Fig. 3, while the slopes of the arrival times with and without the end effect taken into account diverge for larger PEG sizes, they are fairly similar for $M_u \approx$ 100–160 monomers. Hence the approximation used to determine α from the data by neglecting the end effect may be reasonable for these small sizes; however, one would expect it to be poor for the larger sizes for which the end effect has a more critical impact. Fortunately, when the approx-

imation of neglecting the end effect was used [11], $\alpha \approx 0.138$ was in fact determined using only the small sizes range ($M_u \approx 100$–160 monomers) and then this value was used to calculate the molar mass profiles for all sizes. As a result of this somewhat lucky choice for the size range to determine α, very good agreement with MALDI-TOF analyses of molar masses was achieved; for example, FSCE gave a (number) average molar mass of M_n = 5735 g/mol, while MALDI-TOF, the industry standard, gave M_u = 5728 g/mol for the small sizes range [11]. In fact, for all PEG sizes conjugated to an engine of 20 bases, FSCE molar masses agreed with MALDI-TOF results to within a 3.2% difference, supporting the use of the α value of 0.138 determined from the small sizes. If, however, one had used the FSCE data for the larger PEG sizes ($M_u \approx 450$–550 monomers) to determine α under the approximation of neglected end effects, good results would not have been achieved. Figure 3 suggests that the slope of the arrival time for these larger sizes is significantly less than expected by neglecting the end effect. Hence the α value obtained by this approximation, *i.e.*, from the scaled arrival time slope, would be expected to be less than that of the small sizes. Indeed experimental data for the large PEG sizes yields an α value of 0.082 (unpublished result, Vreeland and Barron, personal communication), much less than that from the smaller sizes of 0.138. Using the approximation of neglected end effect to determine α from the larger sizes would have lead to erroneous molar mass calculations from FSCE data, *i.e.*, M_n = 9652 g/mol instead of M_n = 5728 g/mol from MALDI-TOF, for the small sizes range. This means a 69% difference, compared to the mere 0.12% difference from using the α value determined from the smaller PEG sizes for which the end effect plays a lesser role. Clearly the end effect has a critical impact on the electrophoretic behavior of charged-uncharged polymer complexes and must be taken into account to ensure accurate determinations of molar mass from FSCE analysis.

In the preceding development we chose to use α = 0.138 due to the good agreement achieved between FSCE and MALDI-TOF results; however, we can also determine a value for both α and M_u simultaneously by solving the equation for arrival time (Eq. 6) and its derivative with respect to M_u. By this approach, we take the end effect into account and use only the arrival time of the conjugates at the detector and the derivative of this time with respect to peak number. (Note that the peak number varies linearly with PEG size M_u, as mentioned previously.) This system of two equations and two unknowns was solved numerically to yield values of α = 0.168 and M_u = 111 monomers for the middle peak of the small PEG sizes (5 kDa nominal average molar mass). The results for the midpoint of the larger PEG sizes (20 kDa nominal average molar mass) were also fairly reasonable at α = 0.129 and M_u = 560 monomers. The α values determined by this technique have a percent difference of 23% (as opposed to 69% using the previous approach). One possible reason for the remaining discrepancy is that experimental conditions may have changed either between runs with the shorter and larger PEGs or even during a single run. The larger PEGs take about 3 times longer to elute and hence it is possible that the electric current may drop and/or the temperature may change slightly during the course of the experiment, for example. A change in temperature would change the value of α directly since this value depends on the flexibility of the polymers, which in turn depend on temperature. If there were a drop in current between the time when the mobility of the unconjugated engine, μ_0, is measured and when the mobilites of the conjugates, μ, are measured then these values would not correspond to the same conditions as is assumed by Eq. (5). While the end effect is clearly manifested in the FSCE data, there is still some discrepancy between prediction and that which is observed experimentally; this may be due to changes in experimental conditions such as those mentioned above, or to second-order effects not yet taken into account which will be discussed later.

4 The end effect and ELFSE

With ELFSE, variable engine (ssDNA) lengths M_c are conjugated to uncharged molecules of a set size M_u. In previous experimental work [1, 5, 6], the uncharged drag molecule was streptavidin, which was estimated (by neglecting the end effect) to have an effective number of monomers αM_u = 36 under the specific experimental conditions. Through conjugation with the uniform drag molecules, the various lengths of ssDNA, up to about 110 bases, were successfully sequenced in free solution [1]. Since ELFSE is used for sequencing of DNA, an exact value for α is not as crucial for data analysis, *i.e.*, one need only be concerned with the sequence of arrival times, which is not changed by the end effect. However, to fully understand ELFSE data, and to make predictions for optimal sequencing conditions, the role that the end effect plays should be addressed.

The arrival time at the detector for ELFSE is given by Eq. (6), as with FSCE; here, however, the engine size M_c is no longer constant, rather it is the uncharged segment that remains fixed. As the engine grows relative to the drag molecule, the region of the ψ curve determining its weighting expands beyond the "end" weighting to encompass more of the lower weighting of the "middle" (see Fig. 1). In Fig. 5 it can be seen that the end effect speeds up smaller molecules, while it slows down larger

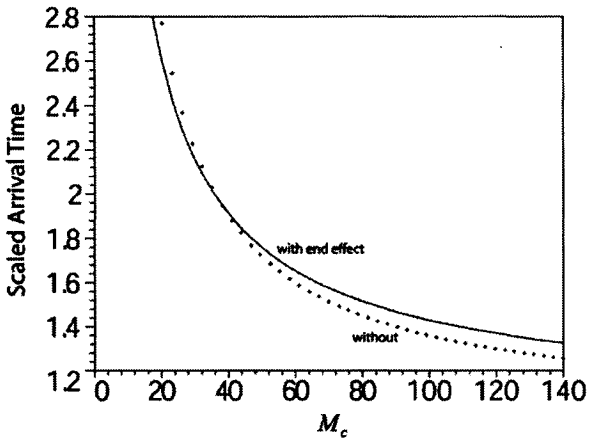


Figure 5. Predicted arrival time at the detector for ELFSE, scaled by the constant $L/\mu_0 E$, as a function of the number M_c of uncharged monomers. The uncharged drag molecule is of effective size αM_u = 36. The solid line represents the case with the end effect taken into account, the dotted line would be expected were there no end effect.

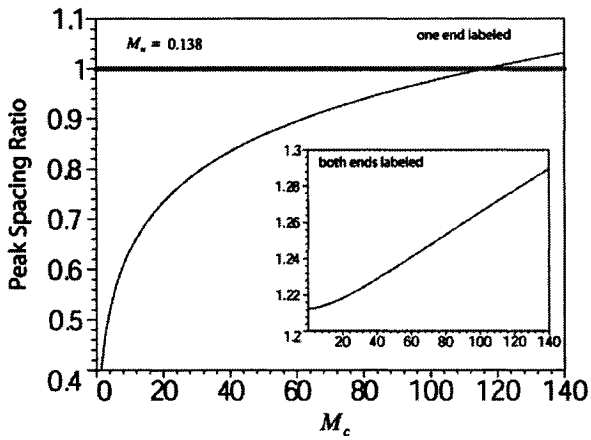


Figure 6. Predicted ratio of ELFSE peak spacing with the end effect to that expected without, for an uncharged drag molecule of effective size αM_u = 36 as used in [5, 6], as a function of the number M_c of charged monomers. Inset: predicted ratio of ELFSE peak spacing with the end effect to that expected without, for an uncharged drag molecule of effective size αM_u = 36 attached at both ends of the ssDNA, as a function of the number M_c of charged monomers.

molecules. Again we are mostly concerned with the resolution, which depends in part on peak spacing. The end effect is expected to decrease peak spacing for the range of data previously investigated (below 110 bases); however, it should start to increase peak spacing at about 115 monomers for these conditions. This crossover from a negative impact on peak spacing to a positive one is shown by the ratio of predicted peak spacing with the end effect taken into account to that without; see Fig. 6. For 110 bases, the largest molecules yet sequenced with this system, there is a slight decrease in peak spacing expected due to end effects, which will quickly be replaced by a positive effect for larger sizes. Hence this examination of the end effect bodes well for ELFSE as this technique matures, *i.e.*, by increasing separating capacity for larger molecules over what could be expected based on data for shorter molecules, where end effects had a more pronounced negative effect.

One of the goals of current ELFSE work is to increase the size of the uncharged segment of the conjugate so as to increase the frictional drag it induces and hopefully extend the read-length, *i.e.*, the number of ssDNA bases which can be sequenced. Unfortunately, for a larger "drag" molecule of αM_u = 100 (rather than 36) effective monomers, the end effect would be expected to decrease peak spacing up until about 320 monomers, *i.e.*, a crossover from a negative to positive effect at about 320 instead of 115 monomers. However, despite the farther reaching negative impact of the end effect, the greater friction of a larger drag molecule would nevertheless result in better separation. The predicted ratio of peak spacing for the hypothetical drag molecule of 100 effective monomers to that of 36 effective monomers is shown in the inset of Fig. 7. The peak spacing is significantly higher for the larger label, at least two times higher throughout the range of DNA sizes shown.

Another means of increasing the resolution of ELFSE would be to label both ends of the ssDNA chain with the drag molecule. This would give each conjugate two drag molecules, thus increasing the total friction. However, in contrast to simply doubling the size of a single drag molecule, the key feature of this configuration is that the drag molecules would be given the highest weighting, that of both ends, leaving the charged section only the lower "middle" weighting of the ψ function. Hence by placing the uncharged sections, with their null mobility, at each end, the resulting frictional drag of the conjugate would be optimized; adding one label to each end of the ssDNA chain would have more impact than doubling the size of a single end label. Figure 7 (main figure) shows the expected peak spacing improvement were both ends labeled with the drag molecule of 36 effective monomers rather than just one. Clearly, having two drag molecules instead of one would not simply double the effective friction coefficient of the uncharged sections, as would be expected were there no end effect, rather it would be increased by a factor greater than two due to the end effect. One important finding is that, unlike the situation with only one end label, the end effect would increase

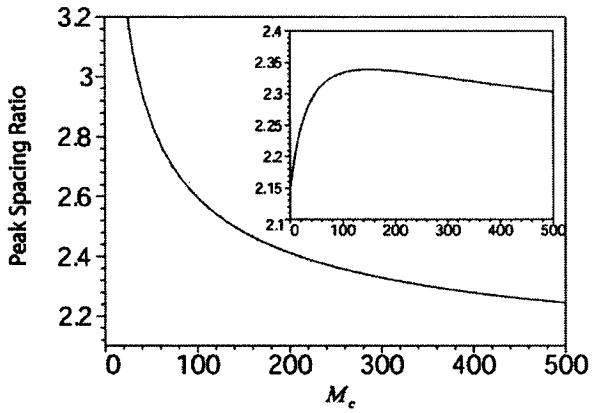


Figure 7. Predicted ratio of ELFSE peak spacing for both ends of the ssDNA chain labeled with a drag of effective size αM_u = 36 to that with only one end labeled. Inset: predicted ratio of ELFSE peak spacing, taking into account the end effect, for a hypothetical uncharged drag molecule of effective size αM_u = 100 to that of effective size αM_u = 36, showing the higher peak spacing of the larger drag molecule. Both curves were calculated by taking into account the end effect.

peak spacing for all sizes if both ends were labeled (see the inset of Fig. 6). For smaller ssDNA chains, having both ends labeled with a drag molecule of αM_u = 36 effective monomers would result in better peak spacing than having a single drag molecule of 100 effective monomers, whereas for larger chains, beyond 308 bases, the inverse would be true (compare Fig. 7 with its inset). Since it may be difficult to find a larger drag molecule which is suitable (*i.e.*, it would have to be water-soluble and amenable to uniform conjugation to ssDNA), it may be preferable to attach two of the smaller labels as a means of improving ELFSE separation; this is one of the main findings of this work. Previously, Heller *et al.* [4] labeled double-stranded DNA with a streptavidin molecule on one end as well as both ends. Without knowledge of the end effect, they calculated the α value for these conjugates to be 23 for a single drag molecule, but 54, rather than 46, for two drag molecules. While these results were misunderstood at the time, our re-analysis of their data indicates that the end effect played a significant role in determining the overall mobility of the conjugates: labeling both ends more than doubled their effective friction coefficient, a result that could not be explained until now.

5 Discussion

It is important to note that the end effect theory of Long and co-workers [14] is for undeformed Gaussian coils. The end effect arises due to the effective "shielding" of monomers located inside the coil (on average) which leaves the ends (located closer to the outside of the coil on average) to interact more with the surrounding fluid, and thereby to have a greater effect on the overall mobility. Hence one must be careful in applying the results presented in this paper to very short molecules whose conformation may not yield this end effect. Also, for very large molecules, there is an excluded volume effect that is not accounted for by the Gaussian coil approximation, which could change the predictions somewhat for these larger molecules.

There is also a small effect due to the hydrodynamic interactions between adjacent monomers on the chain which was not taken into account in previous theories. Although long-range hydrodynamic interactions are screened by the counter-ions, there is some coupling on a local scale between adjacent monomers [14]. As a result, uncharged monomers neighbouring charged monomers are pulled along by the hydrodynamic flow created by the electrophoretic pull on the charged monomers. This effect is highly localized and drops off exponentially with distance; however, it gives an effective non-zero mobility to nearby uncharged monomers. This highly localized effect also means that the end monomers of a charged section have a slightly lesser effective mobility than those in the middle of the charged section since they do not have the additional mobility due to the hydrodynamic flow created by the electrophoretic movement of the nearby charged monomers on both sides. Hence for the mobility in FSCE and ELFSE, the most highly weighted monomers, the ones at the end, have a slightly lesser effective mobility, while the first few uncharged monomers near the joint with the charged chain section have a slight, non-zero mobility. Hence this local hydrodynamic effect could play a role in determining the overall mobility of conjugates; for example, it could decrease the end effect slightly by decreasing the mobility of the most heavily weighted monomers, those charged monomers at the end of the molecule. However, this would be in an absolute fashion in that it would not depend on the relative sizes of the different components of the molecule, unlike the end effect. For ssDNA under the conditions of ELFSE and FSCE, the extra mobility given to the uncharged segment neighbouring the ssDNA monomers, and that taken away from the first few ssDNA monomers on each end of the ssDNA segment, are expected to be negligible. However, for more flexible molecules this local hydrodynamic coupling extends over more monomers and hence this effect could be important and should be taken into consideration for the mobility of such conjugate molecules.

Our re-analysis of the FSCE results, in light of the end effect predicted by Long and co-workers [14], has shown that this effect is indeed significant; it is readily visible in the data and must be taken into account when calculating

the molecular mass. As the size of the uncharged polymers increases, the relative size of the engine decreases so that it receives a much greater weighting in the average determining the overall mobility. As a result, for larger molecules the predicted mobility is greater than would be expected were there no end effect. There is a corresponding decrease in peak spacing, originally assumed to be constant [11, 13], which must be taken into account when analyzing the data, especially when the peak spacing is used to determine the α value of the uncharged polymer. In our previous work [11, 13] we were fortunate that we used the peak spacing for the smaller PEG molecules to determine the value of α that we then used to determine the molecular masses, because the end effect had less of an impact for the smaller sizes, such that the approximation of negligible end effects was acceptable. The value of α used in the determination of the molecular masses from FSCE data is crucial and unfortunately can not be obtained as simply as previously thought. It can be calculated from the persistence lengths and monomer sizes of the two sections of the conjugate [13], although one would need to be careful to take the experimental conditions (temperature and ionic strength) into account. Another means of determining the α value would be to compare the FSCE results to MALDI-TOF results for the same polymer and find the α value that allows for agreement between the two molecular mass estimates (similar to the approach taken in this paper for assessing the accuracy of the value for α). This value need only be determined once for each conjugate type and then FSCE calculations can be made independently. In addition, the simultaneous solution of the equations for the arrival time and the derivative of the arrival time provides another means of estimating α. For this technique to yield accurate results, a very precise measurement must be made of the length-independent free solution ssDNA mobility μ_0, as the results obtained depend quite sensitively on it.

Although the end effect explains the decrease in peak spacing observed in FSCE data, it does not appear to completely account for the decrease. This effect is predicted (based on an α value of 0.138) to decrease the peak spacing of the larger PEG sizes (about 500 monomers) to 77% of that of the smaller PEG sizes (about 130 monomers), whereas the data shows an even greater decrease: the peak spacing of the larger PEG sizes is only 59% of that for the smaller PEG sizes. This discrepancy may be due to excluded volume effects for the larger PEG sizes which were neglected by Long and co-workers when they determined the function governing the end effect [14]. Also any variation in temperature or electric current during or between experiments would change the mobility, and the former would also lead to a change in persistence length, thereby changing the α value itself. A very clear demonstration of the decrease in peak spacing for larger molecules is provided by Bullock [15], where PEG with two end labels were electrophoresed in free solution. The end labeling was achieved by reacting the terminal hydroxyl groups of PEG with phthalic anhydride, thereby tagging a phthalate ester onto each end. The separation was performed under conditions of electroosmotic flow (EOF) which caused the molecules to migrate backwards in the electric field such that the slowest became the fastest and *vice versa*. The change in peak spacing with molecule size is readily visible in the electropherogram, Fig. 11 in [15]; the larger PEG molecules (about 70 monomers) have a peak spacing that is less than one-fifth of that of the smaller PEG molecules (about 20 monomers). Not only does this show a very clear, single data set expression of decreased peak spacing for larger conjugates, but it also confirms that the decrease in peak spacing is not due to a systematic change in experimental conditions during electrophoresis causing a decrease in peak spacing because here the EOF makes it such that the larger molecules elute first.

The end effect is also very important for ELFSE since it can greatly increase, or reduce peak spacing depending on the conditions of the experiment. Once the desired sequencing length is chosen, the end effect can be taken into account in order to determine the necessary label configuration. The end effect is predicted to increase peak spacing for molecules just beyond the range of current experimental data [1], and hence affects predictions of optimal performance. Having a precise value for α is not as much of an issue as it is for FSCE because with ELFSE the ssDNA is being sequenced and hence the length is known. This value may be important, however, for system optimization and other theoretical analyses; for example, we have found that attaching the small label (of effective size $\alpha M_u = 36$) that has been used experimentally thus far, to both ends of the ssDNA would result in better peak spacing than could be achieved through one single larger label (of effective size $\alpha M_u = 100$), under certain conditions. This remarkable result could not have been expected without taking the end effect into account.

The end effect not only has a critical impact on the electrophoretic behavior of charged-uncharged polymer complexes, but it also affects polymers with variable charge distributions. Due to the end effect, a polymer having more of its charges located near the end(s) would have a higher electrophoretic mobility than if its charges were located at the middle of the chain. Recently, a technique similar to FSCE was used to study glutamine deamidation in a long polypeptide [16]. The extent to which glutamine deamidation occurs varies with the extent of

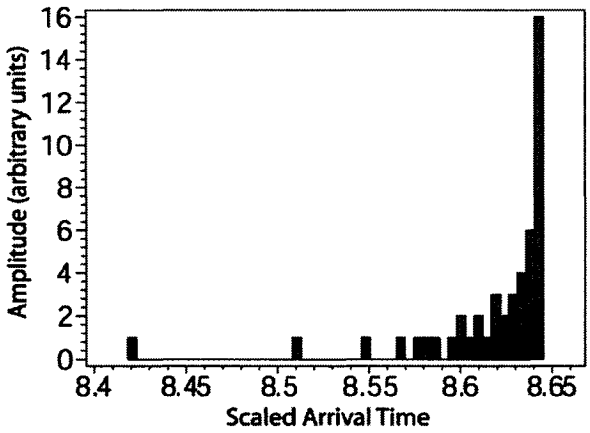


Figure 8. Histogram of predicted arrival times (scaled by the constant $L/\mu_0 E$) to roughly show the expected peak shape (without diffusion) due to the various possible locations for a single deamidation of the ssDNA-protein polymer complexes (for which $M_u = 337$ and $M_c = 23$ before any deamidation), investigated in [16]. We used $\alpha = 1$.

exposure to cyanogen bromide cleavage reaction mixture. In order to assess the degree of deamidation, a uniform DNA engine was conjugated to the protein polymer for electrophoresis. The latter, however, was also of a set length, but it had a varying charge distribution due to the negative charge of the deamidated glutamic acid residue(s). In this study, there were 48 potential sites for deamidation spaced evenly throughout the protein polymer and it was assumed that deamidation occurred randomly over these sites. The electrophoretic separation revealed varying electrophoretic mobilities even though the complexes were all of the same length, because of the varying extents of deamidation: the greater the extent of deamidation, the greater the charge and hence the higher the mobility. However, for each degree of deamidation the end effect would also result in a spread in mobilities based on the location of the deamidation site along the chain. Even for a single negative charge resulting from a single deamidation, the 48 possible locations for the charge, some near the end, others near the middle of the conjugate, would allow for a spread in mobilities. This spread is due to a constant velocity difference between the molecules with different deamidation locations, and hence the peaks would be expected to broaden linearly with time even in the absence of diffusion. The peak shape for a single deamidation is roughly predicted to be that presented in Fig. 8. This rough peak shape was obtained by approximating $\alpha = 1$, and taking the mobility of a deamidated glutamic acid residue to be about that of single-stranded DNA. Each location for the negative charge due to deamidation is expected to have equal probability. To obtain the expected peak shape we used a histogram that would collect the number of conjugates arriving at the detector within a set amount of time. Clearly there are some conjugates that have a much higher mobility (and hence shorter arrival time); these faster molecules have their deamidation induced negative charge located near the end of the chain and hence the end effect gives it a greater weighting in the mobility. These faster molecules may even be lost in the peak corresponding to the next level of deamidation. This may explain some of the peak shapes observed in [16]. Hence the end effect may also be of interest in analyzing electropherograms of uniform length molecules with varying charge distributions.

The authors would like to thank Wyatt Vreeland for experimental data and, along with Didier Long, Annelise E. Barron, and Robert J. Meagher, for useful discussions. The work was supported in part by a Natural Science and Engineering Research Council of Canada Discovery Grant to GWS, and 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 or the NIH.

Received October 26, 2004

6 References

[1] Ren, H., Karger, A. E., Oaks, F., Menchen, S., *et al.*, *Electrophoresis* 1999, *20*, 2501–2509.

[2] Völkel, A. R., Noolandi, J., *Macromolecules* 1995, *28*, 8182–8189.

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

[4] Heller, C., Slater, G. W., Mayer, P., Dovichi, N., *et al.*, Drouin, G., *J. Chromatogr. A* 1998, *806*, 113–121.

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

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

[7] Viovy, J. L., *Rev. Mod. Phys.* 2000, *72*, 813–872.

[8] Olivera, B. M., Baine, P., Davidson, N., *Biopolymers* 1964, *2*, 245–257.

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

[10] Stellwagen, N. C., Stellwagen, E., *Electrophoresis* 2002, *23*, 1935–1941.

[11] Vreeland, W. N., Desruisseaux, C., Karger, A. E., Drouin, G., *et al.*, *Anal. Chem.* 2001, *73*, 1795–1803.

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

[13] McCormick, L. C., Slater, G. W., Karger, A. E., Vreeland, W. N., *et al.*, *Electrophoresis* 2001, *924*, 43–52.

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

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

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

3

Using the Electroosmotic Flow in ELFSE

L.C. McCormick, G.W. Slater, *Electrophoresis* **27**, 1693–1701 (2006)

In this paper we examine the effect on ELFSE performance of having a non-negligible electroosmotic flow (EOF), a flow which is normally suppressed in CE. We predict that the resolution can be dramatically improved with use of an optimal speed of EOF.

I was primarily responsible for the research and writing of this article, with the critical support and guidance of my supervisor.

Electrophoresis 2006, *27*, 1693–1701

Laurette C. McCormick
Gary W. Slater

Department of Physics,
University of Ottawa,
Ontario, Canada

Received August 4, 2005
Revised November 7, 2005
Accepted November 8, 2005

Research Article

A theoretical study of the possible use of electroosmotic flow to extend the read length of DNA sequencing by end-labeled free solution electrophoresis

End-labeled free solution electrophoresis (ELFSE) provides a means of separating DNA with free-solution CE, eliminating the need for gels and polymer solutions which increase the run time and can be difficult to load into a capillary. In free-solution electrophoresis, DNA is normally free-draining and all fragments reach the detector at the same time, whereas ELFSE uses an uncharged label molecule attached to each DNA fragment in order to render the electrophoretic mobility size-dependent. With ELFSE, however, the larger molecules are not separated enough (limiting the read length in the case of ssDNA sequencing) while the smaller ones are overseparated; the larger ones are too fast while the shorter ones are too slow, which is the opposite of traditional gel-based methods. In this article, we show how an EOF could be used to overcome these problems and extend the DNA sequencing read length of ELFSE. This counterflow would allow the larger, previously unresolved molecules more time to separate and thereby increase the read length. Through our theoretical investigation, we predict that an EOF mobility of approximately the same magnitude as that of unlabeled DNA would provide the best results for the regime where all molecules move in the same direction. Even better resolution would be possible for smaller values of EOF which allow different directions of migration; however, the migration times then would become too large. The flow would need to be well controlled since the gain in read length decreases as the magnitude of the counterflow increases; an EOF mobility double that of unlabeled DNA would no longer increase the read length, although ELFSE would still benefit from a reduction in migration time.

Keywords: DNA sequencing / Electroosmotic flow / End-labeled free solution electrophoresis / Electrophoretic mobility / Free-solution electrophoresis
DOI 10.1002/elps.200500573

1 Introduction

End-labeled free solution electrophoresis (ELFSE) is a relatively new technique that achieves separation of various lengths of DNA in free solution [1–4]. This is accomplished by attaching an uncharged (or nearly so) end label called a drag molecule (or drag-tag) of a set size to each DNA fragment in order to render the resulting conjugate's electrophoretic mobility length-dependent, and overcome the free-draining phenomenon which normally leads to comigration of all lengths of DNA in free solution (except very small fragments [5, 6]) [7–10]. This phenomenon is the reason why most DNA separations are performed in a gel which selectively slows down longer polymers more by forcing them to collide more frequently with gel fibers [11]). The key to separation by the ELFSE technique lies in the drag-tag adding a set resistance (friction) to the motion of each DNA fragment, meaning that the more charged monomers a conjugate has (*i.e.*, the longer the DNA component), the more force it has to pull the drag-tag. Hence, larger conjugates go faster and *vice versa*, leading to size-based separation in free solution. Ren *et al.* [1] have

Correspondence: Professor Gary W. Slater, Département de Physique, Université d'Ottawa, 150 Louis-Pasteur, Ottawa, Ontario K1N 6N5, Canada
E-mail: gslater@science.uottawa.ca
Fax: +1-613-562-5190

Abbreviation: ELFSE, end-labeled free solution electrophoresis

successfully used this technique to sequence up to about 100-base long ssDNA molecules in about 18 min in a 34 cm long capillary; their drag-tag was the globular protein streptavidin.

The theory generally used to analyze ELFSE data indicates that the electrophoretic mobility μ_e of an undeformed conjugate molecule comprising M_c charged monomers (*e.g.*, the number of ssDNA bases in the case of DNA sequencing) and M_u uncharged monomers (the drag-tag) is given in references [2–4, 12, 13]

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

where μ_0 is the length-independent free solution mobility of unconjugated ssDNA. This equation, based on the theoretical work of Long *et al.* [14], has been shown to provide good fits to experimental data [2]. The α_1 value is a microscopic constant which accounts for the difference in monomer size and stiffness between the uncharged and charged monomers such that the product $\alpha = \alpha_1 M_u$ is the number of charged ssDNA monomers that have the same friction coefficient as the drag-tag, yielding a total number of effective monomers (each having the same friction coefficient) in the conjugate of $M = M_c + \alpha_1 M_u$. For example, the streptavidin drag-tag tested for ssDNA sequencing with ELFSE has an effective friction parameter $\alpha = \alpha_1 M_u \cong 24$–40, depending on the ionic strength of the buffer [1]. (Note that the calculations in [1] need to be adjusted to take into account recent improvements to ELFSE theory [2, 4]; however, the $\alpha = \alpha_1 M_u$ value can be taken directly from the slope of their fit in Fig. 7 of [1]. The net mobility of the conjugate given by Eq. (1) is simply a uniformly weighted average of the individual effective monomer mobilities. The migration time $t = \frac{L}{\mu E}$, *i.e.*, the time taken by the analyte to travel the distance L to the detector, is thus given by

$$t = \frac{L}{\mu_0 E} \times \frac{M_c + \alpha_1 M_u}{M_c} = t_0 \times \left(1 + \frac{\alpha_1 M_u}{M_c}\right) \tag{2}$$

where E is the electric field strength and $t_0 = \frac{L}{\mu_0 E}$ is the migration time of an unlabeled ssDNA fragment. The temporal peak spacing can be obtained by taking the derivative of the migration time with respect to the number of charged monomers since there is one peak *per* charged segment length

$$\left|\frac{\partial t}{\partial M_c}\right| = \frac{|t_0 - t|}{M_c} \sim \frac{1}{M_c^2} \tag{3}$$

We can see that the peak spacing decreases very quickly with M_c; hence, conjugates with larger ssDNA fragments, the fastest ones, have very small peak spacing (although they also form very narrow peaks because their short migration times and large molecular weights minimize diffusional peak broadening). As a result, longer ssDNA have peaks that overlap and are less resolved with ELFSE; this process is what limits the read length (currently, about 100 bases can be sequenced with streptavidin without any special base calling software [1]). This is the major issue to overcome in order for ELFSE to become competitive with other DNA-sequencing techniques. The read length would obviously increase if the peak spacing (Eq. 3) could be increased for the longer ssDNA.

Remarkably, unlike most electrophoresis systems, once the fastest resolved molecules reach the detector with ELFSE, all of the slower conjugates are already separated in the channel. In the case of [1] for instance, the smallest molecules (starting at about 23 bases long, including the primer size) took about 18 min to reach the detector but they were already resolved by the time the largest resolved molecule (about 100 ssDNA bases) reached the detector at $t \approx 10$ min. (We will use the results presented in the experimental article of [1] throughout this paper in order to illustrate our theoretical predictions.) The predicted peak spacing of all the smaller ssDNA molecules still in the capillary when the largest resolved conjugate (M_c^*) reaches the detector is shown in Fig. 1. The position of all the smaller molecules when the largest DNA resolved by [1] (M_c^*) =100 ssDNA bases reached the detector at $t(M_c^*) \approx 10$ min, $x(M_c) = \mu_e(M_c) \times E \times t(M_c^*)$ was calculated through use of Eq. (1) and the values of $\alpha = 24$ bases, $\mu_0 = 1.95 \times 10^{-4}$ cm^2/V·s and $E = 333$ V/cm given by these authors. The derivative with respect to M_c of the position gives the spatial peak spacing. From Fig. 1, it is clear that all of the smaller ssDNA have a much greater peak spacing at the time of detection of the fastest resolved ssDNA; this is the reason why [1] observed smaller molecule peaks that were needlessly overseparated by the time they reached the detector. The inset of Fig. 1 shows the corresponding predicted size resolution factor S_m (defined in the next section, Eq. 8), which is the smallest difference in the number of monomers which can be resolved from one another. Clearly we have better than single-monomer resolution ($S_m \leq 1$ monomer, as needed for sequencing) for all the remaining peaks once the largest (100 ssDNA bases) resolved conjugate reaches the detector; in fact the resolution is even better for smaller conjugates. Since the size resolution factor S_m is lower for all the conjugates still in the capillary once the largest resolved molecule reaches the detector, a whole-capillary snapshot detection mode would immediately yield an electropherogram with single-monomer resolution (or better) for $M_c = 0$ through $M_c = M_c^* = 100$ bases in this case. However, with the usual finish line detection mode employed with CE, one must wait for the

Electrophoresis 2006, *27*, 1693–1701

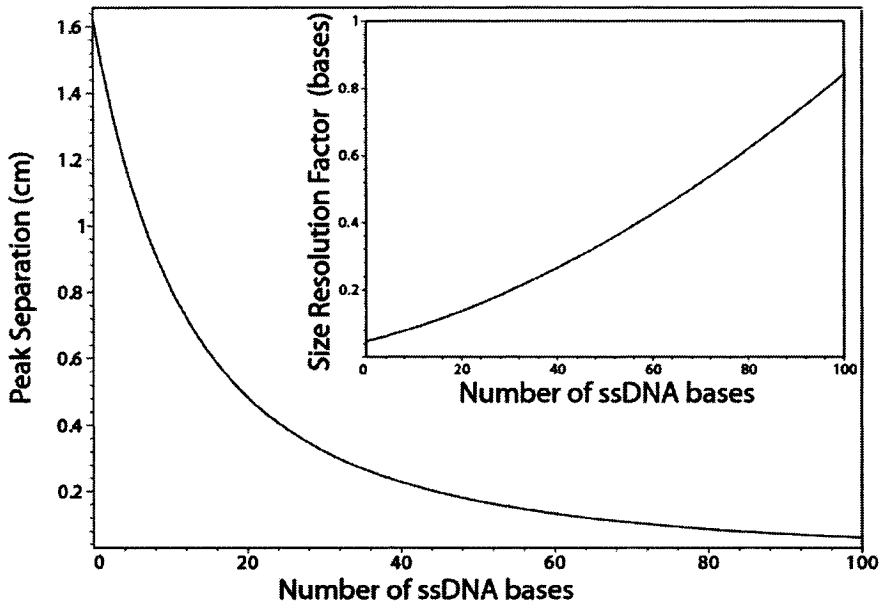


Figure 1. Peak separation (cm) once the fastest resolved ssDNA (of size $M_c = 100$ bases [1]) reaches the detector at $L = 34$ cm, as a function of the number of ssDNA bases (no EOF). Conditions are that of [1]: the effective number of uncharged monomers $\alpha = \alpha_1 M_u$ is 24, $\mu_0 = 1.95 \times 10^{-4}$ cm^2/Vs and $E = 333$ V/cm. Inset: the size resolution factor S_m for the peaks still inside the capillary when the fastest resolved ssDNA of 100 bases reaches the end of the capillary. This factor, defined in Section 2 (Eq. 8), is the smallest difference in the number of monomers which can be resolved from one another; hence, once the 100 base ssDNA reaches the detector with single-monomer resolution (or slightly better), all the smaller molecules inside the capillary are already fully resolved.

slower (smaller) molecules to reach the detector at the end of the capillary by which time they are needlessly overseparated.

With traditional ELFSE the longest conjugates are not separated enough to be resolved, while the shorter ones are overseparated; the longer ones are too fast while the shorter ones are too slow, the opposite of the situation with regular electrophoresis performed in a gel or polymer solution. In order to slow down the longer conjugates and allow them more time to separate, and to speed up the smaller conjugates, we consider performing ELFSE in the presence of an EOF. This counterflow, which is constant [15] (assuming that the capillary is uniformly charged and both ends are at the same pressure [16]), arises as a consequence of the negative charges of the uncoated inner capillary wall surface, and results in the analyte motion proceeding in the reverse direction. In the presence of EOF, the conjugates would be carried along by the opposing flow, resisting the motion to an extent determined by their own electrophoretic mobility μ_e. Hence, the fastest (longest) conjugates in traditional ELFSE would become the slowest in the presence of EOF since they could fight this flow the most, and vice versa.

In order to increase the read length, the peak spacing given by Eq. (3) needs to be increased for larger molecules, for which the numerator $|t_0 - t|$ (*i.e.*, the absolute difference in migration time between unlabeled and labeled DNA) is almost zero because very large ssDNA fragments can pull the drag-tag with ease and approach the speed of unlabeled ssDNA. There are four ways to increase the numerator. Most simply (a) a longer capillary and/or (b) a lower electric field strength could be used to increase both the migration times t and t_0, and thereby increase their absolute difference. (Actually the former will increase the peak spacing for most electrophoretic systems, including gel-based methods; however, with the latter the gain in peak spacing may unfortunately be accompanied by an insurmountable increase in diffusion.) Another means of increasing the numerator is to (c) use a drag-tag capable of exerting greater frictional drag which would decrease t while leaving t_0 unaffected (in fact increasing the frictional properties of the drag-tag is a main goal of current ELFSE research; however, it is extremely challenging experimentally [4]). Finally, while Eq. (3) would need to be adjusted for the presence of EOF, one would expect intuitively that if (d) the EOF were properly chosen it could increase both t and t_0, leading to an increase in peak spacing by slowing down both unlabeled and labeled ssDNA. Thus, the EOF may indeed increase the read length of ELFSE; furthermore, it may also reduce the unnecessary overseparation of small conjugates. In this paper we build on pre-existing ELFSE theory to predict the effect of EOF on ssDNA sequencing.

2 ELFSE in the presence of EOF: theory

In this section we develop detailed equations governing ELFSE in the presence of EOF, and investigate the predicted electrophoretic behavior. As previously mentioned, the EOF is assumed to simply add a constant term μ_{EOF} to the electrophoretic mobility of the analyte. The EOF results from the negative charges on the inner surface of uncoated fused-silica capillary walls which attract positive ions from solution. While the negative charges of the wall are immobile, the positive charges of the thin Debye layer (typically 1–10 nm [16]) neighboring the surface are free to move and hence once an electric field is applied, they move toward the cathode. Their motion drags the fluid from the bulk solution along with them, creating the plug-like EOF. This flow is generally constant and in the opposite direction to the ssDNA conjugate's own mobility μ_e, such that the net mobility of the analyte is the difference of these two mobilities [16]

$$\mu = \mu_{EOF} - \mu_e \tag{4}$$

where μ_e is the mobility of the analyte under conditions of no EOF, as given in Eq. (1). The magnitude of the EOF mobility μ_{EOF} depends on the extent and character of the capillary wall coating; a bare wall exhibits the highest EOF mobility. Whenever the proper mobility μ_e of the analyte is exceeded by the mobility due to the EOF μ_{EOF}, the migration proceeds in the opposite direction, with the conjugate moving toward the cathode instead of the anode. The net migration time in the presence of EOF, $t = \frac{L}{|\mu|E}$, is thus given by

$$t = \frac{L}{\mu_0 E}\left(\frac{1}{|\tilde{\mu}_{EOF} - \tilde{\mu}_e|}\right) \tag{5}$$

where the dimensionless mobility ratios $\tilde{\mu}_{EOF}$ and $\tilde{\mu}_e$ are defined as follows:

$$\tilde{\mu}_{EOF} \equiv \frac{\mu_{EOF}}{\mu_0} \tag{6}$$

$$\tilde{\mu}_e \equiv \frac{\mu_e}{\mu_0} = \frac{M_c}{M_c + \alpha_1 M_u} \tag{7}$$

Since the conjugate's proper mobility decreases due to the drag molecule of effective hydrodynamic size $\alpha = \alpha_1 M_u$ (*i.e.*, $\mu_e \leq \mu_0$), the maximum proper mobility of a conjugate is μ_0, and a scaled EOF mobility $\tilde{\mu}_{EOF}$ exceeding 1 means that all of the conjugates migrate in the opposite direction in the presence of the EOF. We will first investigate this case where all conjugates travel in the same direction, *i.e.*, scaled EOF mobilities in the range $\tilde{\mu}_{EOF} \geq 1$, and then the case for $\tilde{\mu}_{EOF} \leq 1$. Under the former conditions, the conjugates which were the fastest in the traditional EOF-free direction become the slowest in the opposite direction because they can fight the flow the hardest, and *vice versa*, as previously mentioned. Remarkably, we note that for $\tilde{\mu}_{EOF} = 1$, the temporal peak spacing $|\partial t/\partial M_c|$ is constant (as can be verified by taking the derivative of Eq. (5) with respect to M_c), whereas it decreases with increasing ssDNA size M_c (similar to all other separation methods) for any other value of $\tilde{\mu}_{EOF} \geq 1$.

The viability of ELFSE separations in the presence of EOF was shown by Heller *et al.* [10] for dsDNA, although with apparently less success than without the EOF. In the following we investigate how ELFSE separations are affected by the EOF, and in particular how they depend on the scaled EOF mobility $\tilde{\mu}_{EOF}$. We define the size resolution factor as the ratio of the temporal full width at half maximum ($FWHM_t$), to the temporal peak spacing $|\partial t/\partial M_c|$ as the bands pass in front of the detector

$$S_m(M_c, \tilde{\mu}_{EOF}) \equiv \frac{FWHM_t}{|\partial t/\partial M_c|} \tag{8}$$

where the units of S_m are number of monomers. This factor represents the smallest difference in the number of monomers which can be resolved from one another. An $S_m(M_c, \tilde{\mu}_{EOF})$ factor of 1 (*i.e.*, single-monomer resolution) or less is hence necessary for sequencing; clearly, smaller values of this factor correspond to an increase in the resolution power of the system. Following the development in [4, 13], this factor can be expressed as follows for the electrophoretic system of reference [1] (see Addendum A for a brief derivation)

$$S_m(M_c, \tilde{\mu}_{EOF}) \approx \frac{(|\tilde{\mu}_{EOF} - \tilde{\mu}_e|)^{1/2}(M_c + 24)^{7/4}}{5088} \tag{9}$$

The development of this equation assumes that the conjugates are in a Gaussian coil conformation, that the drag-tags are completely monodisperse, and that the band loading width is negligible. We take into account only thermal (diffusion) band broadening (as is the case for experimentally optimal conditions), and neglect any additional band broadening which may arise due to the EOF (for nonideal effects, see [16–18]). Our predictions compare well with the experimental results of [1]. For instance, we note that the predicted size resolution factor for the largest resolved ssDNA as shown in the inset of Fig. 1 is slightly less than 1, indicating that even better resolution could be expected were their experimental conditions ideal; however, the initial loading width was not completely negligible, there may have been some other sources of nonthermal band broadening, and the label was slightly polydisperse. Nonetheless, their experimental results are very close to what we predict based on ideal conditions, indicating that their conditions were close to optimized, and reinforcing the validity of our model.

Electrophoresis 2006, *27*, 1693–1701

3 Results

3.1 EOF mobility exceeding the mobility of all conjugates: single direction of migration

Here we investigate ELFSE in the presence of an EOF mobility μ_{EOF} that exceeds the DNA conjugates own proper mobility μ_e (*i.e.*, the mobility that it would have in the absence of EOF) which has a maximal value of the mobility of unlabeled DNA μ_0; hence, we are looking at the situation $\tilde{\mu}_{EOF} \geq 1$ where all conjugates travel backwards, carried by the EOF. The predicted size resolution factor $S_m(M_c, \tilde{\mu}_{EOF})$ using the experimental parameters of [1] is plotted in Fig. 2 for various values of $\tilde{\mu}_{EOF} \geq 1$. We can see clearly that scaled EOF mobilities close to 1 would provide increasingly better size resolution with DNA size (*i.e.*, smaller and smaller differences in the number of monomers could be resolved) than the same experimental system with no EOF; for $\tilde{\mu}_{EOF} \geq 1$, this improvement is expected to begin at about 24 ssDNA bases. As the ratio $\tilde{\mu}_{EOF}$ increases from 1 to 2, the expected gain in resolution quickly decreases; indeed for $\tilde{\mu}_{EOF} = 2$ the curve lies above that for negligible EOF, indicating that slightly poorer resolution would be achieved than with negligible-EOF conditions.

For each curve in Fig. 2 the predicted read length is shown, and corresponds to the intersection of the curve with the horizontal line at $S_m = 1$ which represents the cut-off for single-monomer resolution. These predictions were obtained by setting Eq. (9) equal to one and solving numerically for M_c; *i.e.*, since $S_m(M_c, \tilde{\mu}_{EOF})$ is the number of monomers that can be resolved, and since this value strictly increases with the number of ssDNA bases M_c (for all values of $\tilde{\mu}_{EOF} \geq 1$), setting it equal to one gives the largest ssDNA that can be resolved on a single-monomer basis. Note that ssDNA sequences can often be determined even when peaks are not completely resolved, especially with the aid of sophisticated base calling software, and hence the predicted read lengths are essentially lower bounds for experimentally ideal (diffusion limited) conditions. While for negligible EOF, the read length for the experimental system of [1] is predicted to be about 114 bases (as discussed above, this is similar to the value obtained experimentally of about 100 bases [1], and likely differs due to somewhat nonideal experimental conditions such as non-negligible initial loading width), for $\tilde{\mu}_{EOF} = 1$ it is expected to be much higher, at about 235 bases. Hence, an EOF with $\tilde{\mu}_{EOF} = 1$ would provide substantially better performance than conditions of negligible EOF, extending the read length by over 200%. The read length predicted for $\tilde{\mu}_{EOF} = 1.1$ is still a good improvement over the EOF-free case, at about 179 bases. Although the improvement to the resolution quickly drops off as the mobility $\tilde{\mu}_{EOF}$ increases from 1, the read length for $\tilde{\mu}_{EOF} = 1.5$ is still expected to be better than that of the EOF-free case, at 124 bases. Once the mobility $\tilde{\mu}_{EOF}$ reaches 1.65, the read length returns to that obtained without EOF; however, as we will demonstrate next, using the EOF still offers the advantage of lower total run times.

Figure 3 shows the predicted read length as a function $\tilde{\mu}_{EOF}$ of ≥ 1 for the experimental system of [1]; the corresponding migration time for the largest resolvable molecules is also shown. Under ideal conditions we predict that without the EOF the read length would be 114 bases; this could also be achieved with $\tilde{\mu}_{EOF} = 1.66$, which would have a corresponding run time of only 10.5 min, compared to the 18 min required without EOF, as found experimentally by [1] (note that this is about the run time

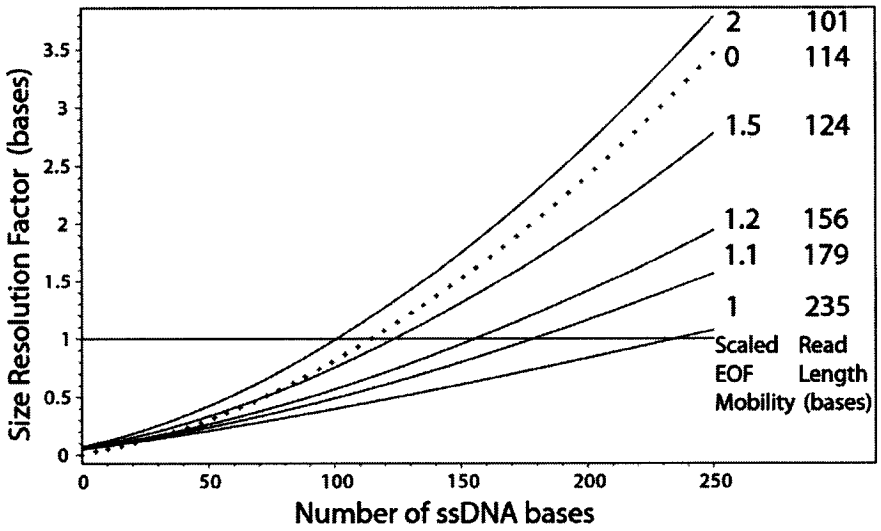


Figure 2. Size resolution factor S_m for the various indicated values of $\tilde{\mu}_{EOF}$, as a function of ssDNA size, for the experimental system of [1]. Dotted line shows the size resolution factor for negligible-EOF conditions; it is clear that $\tilde{\mu}_{EOF} = 1$ provides increasingly better resolution (*i.e.*, smaller and smaller differences in the number of monomers can be resolved) with DNA size beyond about 24 ssDNA bases. Predicted read lengths are also indicated and correspond to the intersection of the size resolution factor curves with the horizontal line at $S_m = 1$ base.

Electrophoresis 2006, 27, 1693–1701

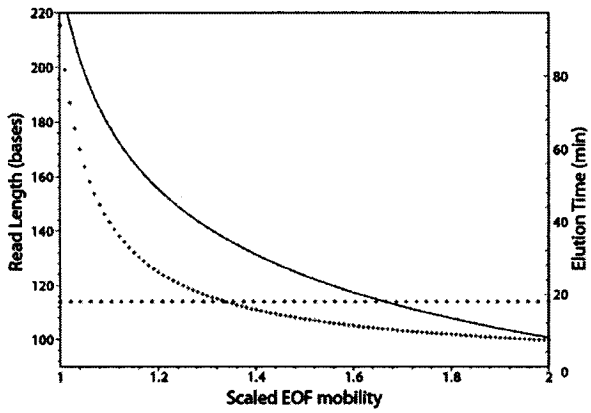


Figure 3. Graph of predicted read length as a function of $\tilde{\mu}_{EOF}$ (solid line), and the corresponding migration time of the largest resolvable conjugate as a function of $\tilde{\mu}_{EOF}$ (dotted line). Horizontal line connects the experimental run time of 18 min of [1] to the predicted optimal read length without EOF for their conditions of 114 ssDNA bases.

we would expect were a snap-shot detection mode available for CE without EOF, see Section 1). The horizontal line connects this experimental run time of 18 min to the optimized negligible-EOF read length prediction of 114 bases. As we can see from Fig. 3, running ELFSE in the presence of EOF not only allows for a substantial increase in the read length but also shortens the total run time for all values of scaled EOF mobility $\tilde{\mu}_{EOF} \geq 1.34$. For values of scaled EOF mobility $\tilde{\mu}_{EOF} \geq 1.34$, more time is required for the resulting increase in read length. For $\tilde{\mu}_{EOF} \geq 1.66$ the separations have a shorter read length with the EOF, but take less time. In the intermediate regime $1.34 \leq \tilde{\mu}_{EOF} \leq 1.66$, increased read length is accompanied by shorter migration times.

3.2 EOF mobility less than the mobility of the fastest conjugate: two migration directions

In this section we look at the situation where the EOF is small enough ($\tilde{\mu}_{EOF} \leq 1$) that some of the faster conjugates can fight it and migrate forward, in the same direction as they would in the absence of EOF. Hence, we have smaller molecules moving backwards and larger molecules that are fast enough to overcome the EOF moving forward. In order to detect both sets of molecules, we would require a different experimental setup, such as injection in the middle of the capillary with detection occurring at both ends, or using multiple runs each geared for a specific size range (and direction). For simplicity we will take the length L from injection to the detector to be the same for both sets of molecules (although different migration lengths might improve the throughput). An EOF mobility μ_{EOF} slightly less than that of unlabeled DNA μ_0 would be even closer to the mobility of very long DNA μ_e (which is slightly less than μ_0 due to the presence of the label) than it would be for $\mu_{EOF} = \mu_0$. Therefore, the longer conjugates would be given even more time to separate from each other; thereby further increasing the read length. Figure 4 shows the size resolution factor S_m as a function of the number of ssDNA bases for a scaled EOF mobility of 0.9, for the conditions of [1] as given in Fig.1, taking the migration length to be 34 cm for both directions of migration. The dotted horizontal line at $S_m = 1$ indicates the cut-off for single-monomer resolution and gives a read length of 328 ssDNA bases for this EOF. The minimum in the curve occurring at 216 ssDNA bases corresponds to the size of DNA which has a mobility approximately equal to the EOF mobility. Hence for a scaled EOF mobility of 0.9, molecules having more than 216 ssDNA bases can fight the EOF and move forward and molecules less than this size are carried backwards by the EOF, while the conjugate with 216 ssDNA bases barely moves. The best resolution occurs near this minimum. Since all of the conjugates with less than 216 bases have a size resolution factor less than 1, they are all resolved, while larger and larger molecules have lesser resolution (due to their ability to fight the EOF and attain speeds that do not allow for adequate separation), eventually reaching that of single-monomer resolution for 328 bases. This read length is excellent for ELFSE;

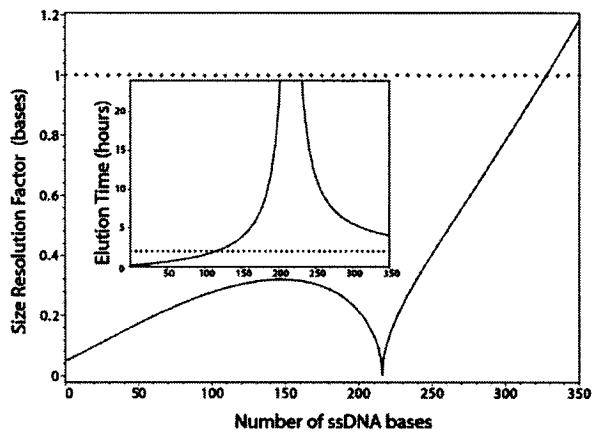


Figure 4. Graph of size resolution factor S_m as a function of number of ssDNA bases, for a scaled EOF mobility of 0.9. Dotted horizontal line at $S_m = 1$ indicates the cut-off for single-monomer resolution, occurring at about 328 ssDNA bases for this EOF. Conditions are as given in Fig. 1 for [1]; the migration length is taken to be 34 cm for both directions of migration. Inset: corresponding migration time in hours; conjugates with more than 115 ssDNA bases require more than 2 h to reach the detector (horizontal dotted line).

Electrophoresis 2006, *27*, 1693–1701

unfortunately, for DNA sizes near the size resolution factor minimum, *i.e.*, those that barely move because their own proper mobility is almost the same as the EOF, the migration time would be very long; in fact the migration time diverges. The inset of Fig. 4 shows the corresponding migration time for the scaled EOF mobility of 0.9. Conjugates with more than 115 ssDNA bases require more than 2 h to reach the detector, while those molecules with mobilities approximately equal to that of the EOF, *i.e.*, those having 216 plus or minus a few ssDNA bases, barely move and hence do not reach the detector in any reasonable amount of time.

Figure 5 shows the predicted read length as a function of $\tilde{\mu}_{EOF}$ for the conditions of [1] as given in Fig. 1. Also shown is the number of ssDNA bases for which the mobility is zero ($\mu_e = \mu_{EOF}$), *i.e.*, the conjugate size for which the migration time diverges (curve a); between curves b and c, separations take longer than 2 h. Unfortunately, although the read lengths are exceptional for $0.9 \lesssim \tilde{\mu}_{EOF} \lesssim 0.96$, separations in this EOF range require too much time and are likely only of interest for special applications. A scaled EOF mobility of 0.99 would still allow for all conjugates (under the conditions of [1]) to move in the same direction while slightly increasing the read length over that of $\tilde{\mu}_{EOF} = 1$ (235 ssDNA bases) to 248 ssDNA bases, with a corresponding increase in migration time from 1.6 to 1.9 h. It should also be noted that even if the curves are resolved, for these long migration times the bands will also be fairly wide and will take some time to pass in front of the detector. For example, with a scaled EOF mobility of 0.99, under the conditions of [1], the slowest band (*i.e.*, the largest ssDNA) would take about 28 s to pass in front of the detector; it is possible that for these spread-out bands, the S/N may be too low to detect.

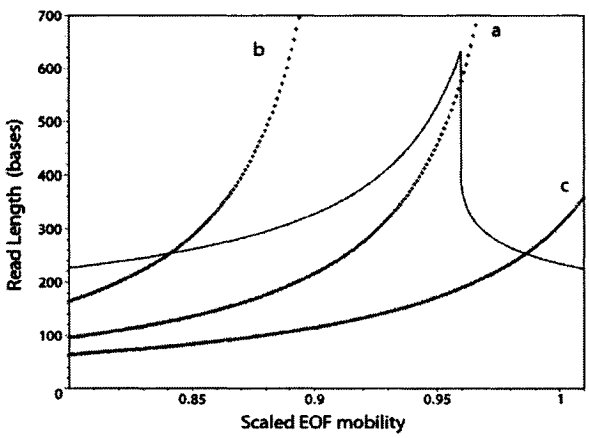


Figure 5. Graph of predicted read length (solid line) as a function of scaled EOF mobility, for the conditions of [1] as given in Fig. 1; the migration length is taken to be 34 cm for both directions of migration. Also shown is the number of ssDNA bases for which the mobility is zero $\mu_e = \mu_{EOF}$, *i.e.*, the conjugate size for which the migration time diverges (curve a); between curves b and c, separations take longer than 2 h.

Since the migration time becomes a limiting factor for the read length, systems which shorten the run time would increase the gains expected through use of EOF for ELFSE. All of our predictions are based on the capillary electrophoretic system of [1]; with the increased speed of microchip electrophoretic systems even better gains due to the EOF could be expected by overcoming the time restraints. The theory presented here could be easily adapted for such systems which may indeed make EOF-based ELFSE a competitive sequencing technique, allowing for rapid, high-read length separations void of the need for gels or entangled polymer solutions.

4 Discussion

We have shown that the EOF can be used to dramatically extend the read length of DNA separations by ELFSE by improving the resolution of larger molecules. For the case of all molecules migrating in the same direction (*i.e.*, $\tilde{\mu}_{EOF} \equiv \frac{\mu_{EOF}}{\mu_0} \geq 1$), the best resolution is expected when the scaled EOF mobility is near unity, and positive effects drop quickly with an increase in $\tilde{\mu}_{EOF}$. For example, a scaled EOF mobility of unity is predicted to more than double the read length for the system of [1] (for which optimal conditions would be expected to yield a read length of 114 ssDNA bases without the EOF), extending it to 235 ssDNA bases. For the case of smaller molecules migrating backwards with the EOF and larger molecules moving forward against the EOF (*i.e.*, $\tilde{\mu}_{EOF} \leq 1$), even more exceptional improvements to the read length are expected; however, the long run time makes this useful for special applications only. For the conditions of [1], a scaled EOF mobility of 0.99 would still allow all the molecules to migrate in the same direction, and the read length is predicted to be 248 ssDNA bases, an exceptional improvement over the predicted optimal read length of 114 bases for ELFSE without EOF.

It is clear that in order to take advantage of the EOF-based resolution increase, the exact value of the scaled EOF mobility must be well controlled. The coating on the capillary wall surface is a key factor determining EOF. Heller *et al.* [10] reduced the EOF from that of an uncoated capillary by 50%, to 1×10^{-3} cm^2/V·s through use of a thin polyacrylamide coating. This corresponds to a scaled EOF mobility in the range $2 < \tilde{\mu}_{EOF} < 10$, given that values of μ_0 typically range from 1×10^{-4} cm^2/V·s to

5×10^{-4} cm^2/V·s. Hence, the EOF would typically need to be reduced by 75% or more in order to achieve a $\tilde{\mu}_{EOF}$ value near unity, for example. Another means of controlling the EOF is by the application of an external electric field which forms a potential gradient with the usual internal electric field thereby creating a radial field; this adjustable gradient is perpendicular to the capillary wall and changes the density of electric charge on the inner capillary wall, thereby allowing for control of the EOF [19–21]. In addition to the EOF, all factors influencing the mobility would also need to be well controlled so as to maintain a constant μ_0 since the desired EOF mobility depends upon this value.

In addition to the clear resolution advantage of performing ELFSE in the presence of EOF, the decrease in run time would also be a big benefit; indeed even nonoptimal EOF values, ($\tilde{\mu}_{EOF} \geq 1.34$) which would not substantially improve resolution, would still shorten the total time required for the electropherogram. For values of scaled EOF mobility $1 \leq \tilde{\mu}_{EOF} \leq 1.34$, more time is required for the resulting increase in read length. The EOF would also change the order of detection as the smaller conjugates reach the detector first, followed by the larger conjugates, restoring the usual order, as with standard (gel/entangled polymer) sequencing, and eliminating the unnecessary wait for small, already resolved molecules to travel to the detector. If the EOF could be maintained at $\tilde{\mu}_{EOF} = 1$, one could also expect evenly spaced peaks which may allow for easier base calling algorithms; somewhat larger values of $\tilde{\mu}_{EOF}$ would give approximately constant peak spacing which would also be beneficial for base calling.

In order to achieve comparable read lengths without EOF, very powerful voltage supplies would be necessary. For example, to obtain a read length comparable to the 235 bases predicted with $\tilde{\mu}_{EOF} = 1$ without the EOF (for the system of [1]), one would need a 3.3 m long capillary which would require a much greater voltage in order to maintain the electric field strength at approximately 333 V/cm. Similarly, comparable read lengths obtained *via* an increase in the electric field would require an electric field strength of about 3300 V/cm, which would also be very demanding indeed in terms of the power supply source. Not only would the field strengths required be extreme but also they might be accompanied by an unfavorable increase in peak widths. Using the EOF is a powerful alternative to these extreme and unrealistic approaches.

We also note that while one could use a method other than the EOF to create the counterflow in an attempt to take advantage of the potential gains, such as a pressure difference, it would lack the characteristic EOF plug-like flow. Typically, non-EOF-based counterflows have a parabolic profile, in contrast to the flat profile across the bulk fluid obtained with EOF. It is only with a flat profile that all molecules across the diameter of the capillary experience the same rate of counterflow; a parabolic profile would mean that molecules near the center would be subject to a greater counterflow than those closer to the outside, leading to an undesirable band broadening.

Our predictions of ELFSE behavior in the presence of EOF are based on negligible band loading width and assume that any EOF-based band broadening effects are negligible. For systems where this assumption is not entirely justified, adjustments would need to be made to the above theory. It is important to note as well that the drag molecule for ELFSE in the presence of EOF would need to be free of problems of sticking to the uncoated (or less coated) capillary wall.

This work was supported, in part, by a Natural Science and Engineering Research Council of Canada Discovery Grant to GWS, and 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 or the NIH.

5 References

[1] Ren, H., Karger, A. E., Oaks, F., Menchen, S. *et al.*, *Electrophoresis* 1999, *20*, 2501–2509.

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

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

[4] Meagher, R. J., Won, J.-I., McCormick, L. C., Nedelcu, S. *et al.*, *Electrophoresis* 2005, *26*, 331–350.

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

[6] Stellwagen, N. C., Stellwagen, E., *Electrophoresis* 2002, *23*, 1935–1941.

[7] Olivera, B. M., Baine, P., Davidson, N., *Biopolymers* 1964, *2*, 245–257.

[8] Völkel, A. R., Noolandi, J., *Macromolecules* 1995, *28*, 8182–8189.

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

[10] Heller, C., Slater, G. W., Mayer, P., Dovichi, N. *et al.*, *J. Chromatogr. A* 1998, *806*, 113–121.

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

[12] Vreeland, W. N., Desruisseaux, C., Karger, A. E., Drouin, G. *et al.*, *Anal. Chem.* 2001, *73*, 1795–1803.

[13] McCormick, L. C., Slater, G. W., Karger, A. E., Vreeland, W. N. *et al.*, *J. Chromatogr. A* 2001, *924*, 43–52.

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

[15] Sinton, D., Escobedo-Canseco, C., Ren, L., Li, D., *J. Colloid Interface Sci.* 2002, *254*, 184–189.

[16] Ghosal, S., *Electrophoresis* 2004, *25*, 214–228.

[17] Potocek, B., Gaš, B., Kenndler, E., Štědrý, M., *J. Chrom. A* 1995, *709*, 51–62.

[18] Gaš, B., Štědrý, M., Kenndler, E., *J. Chromatogr. A* 1995, *709*, 63–68.

[19] Kašička, V., Prusik, Z., Sazelova, P., Chiari, M. *et al.*, *J. Chrom. B* 2000, *741*, 43–54.

[20] Kašička, V., Prusik, Z., Sazelova, P., Brynda, E., Stejskal, J., *Electrophoresis* 1999, *20*, 2484–2492.

[21] Hartley, N. K., Hayes, M. A., *Anal. Chem.* 2002, *74*, 1249–1255.

6 Addendum: a brief derivation of Eq. (9)

In the following we give a brief derivation of Eq. (9), the size resolution factor for the system of [1]. The definition of this factor is given by Eq. (8). First, we start with the numerator, the temporal FWHM (we assume Gaussian peaks):

$$\mathrm{FWHM_t} = 2\sqrt{2\ln(2)}\sigma_t \tag{10}$$

where σ_t is the temporal standard deviation and can be given as follows when the initial peak width is negligible and diffusion is the only significant source of band broadening:

$$\sigma_t = \frac{\sqrt{2Dt}}{v} \tag{11}$$

where v is the velocity, $D = \dfrac{k_B T}{4\pi\eta R_G}$ is the Zimm diffusion coefficient of the hybrid ssDNA molecule, k_B is the Boltzmann constant [13], T is the absolute temperature, η is the viscosity of the free solution, and R_G is the radius of gyration. Hence, the numerator of Eq. (8) can be rewritten as follows, where we have replaced v by L/t

$$\mathrm{FWHM_t} = \frac{2}{L}\sqrt{\frac{\ln(2)k_B T}{\pi\eta R_G}} \times t^{3/2} \tag{12}$$

Following the blob approach presented in [4, 13] which rescales the charged and uncharged segments to account for their different hydrodynamic sizes, the total radius of gyration of the conjugate molecule can be given by that of its charged and uncharged segments

$$R_G = \sqrt{R_{G_c}^2 + R_{G_u}^2} \tag{13}$$

If we assume excluded volume effects to be negligible, the radii of gyration are given by

$$R_{G_i} = \sqrt{\frac{b_{K_i} b_i M_i}{6}} \tag{14}$$

where b_{K_i} is the Kuhn length of polymer i, a measure of its stiffness, and b_i is the monomer size of polymer i. Hence, the total radius of gyration of the conjugate can be written as

$$R_G = \sqrt{\frac{b_{K_c} b_c M_c}{6} + \frac{b_{K_u} b_u M_u}{6}} = \sqrt{\frac{b_{K_c} b_c}{6}} \times (M_c + \alpha_1 M_u) \tag{15}$$

where $\alpha_1 \equiv \dfrac{b_{K_u} b_u}{b_{K_c} b_c}$, as given by [4, 13]. Using Eqs. (5) and (7) for the denominator of the size resolution factor, we find

$$\left|\frac{\partial t}{\partial M_c}\right| = \frac{\alpha_1 M_u t}{(\tilde{\mu}_{EOF} - \tilde{\mu}_e)(M_c + \alpha_1 M_u)^2} \tag{16}$$

Substituting Eq. (15) into the expression for the numerator, Eq. (12), and using Eq. (16) for the denominator, the size resolution factor becomes

$$S_m \approx \frac{2}{L}\sqrt{\frac{\ln(2)k_B T}{\pi\eta\sqrt{\frac{b_{K_c} b_c}{6}}}} \times \frac{(|\tilde{\mu}_{EOF} - \tilde{\mu}_e|)(M_c + \alpha_1 M_u)^{7/4} t^{1/2}}{\alpha_1 M_u} \tag{17}$$

Again making use of Eq. (5) for the migration time, we find

$$S_m \approx 4\sqrt{\frac{\ln(2)D_0}{L\mu_0 E}} \times \frac{(|\tilde{\mu}_{EOF} - \tilde{\mu}_e|)^{1/2}(M_c + \alpha_1 M_u)^{7/4}}{\alpha_1 M_u} \tag{18}$$

where the constant D_0 is defined by

$$D_0 \equiv \frac{k_B T}{4\pi\eta\sqrt{\frac{b_{K_c} b_c}{6}}} = D\sqrt{M_c + \alpha_1 M_u} \tag{19}$$

and can be found from the Ren *et al.* [1] value of the diffusion coefficient $D = 4.8 \times 10^{-7}$ cm^2/s, reported for $M_c = 61$ bases and $\alpha_1 M_u = 24$ bases to be $D_0 = 4.43 \times 10^{-6}$ cm^2/s. Using this value, along with the experimental values from [1] presented above ($L = 34$ cm, $\alpha_1 M_u = 24$, $\mu_0 = 1.95 \times 10^{-4}$ cm^2/V·s and $E = 333$ V/cm) we arrive at Eq. (9) for the size resolution factor for the experimental conditions of [1]

$$S_m(M_c, \tilde{\mu}_{EOF}) \approx \frac{(|\tilde{\mu}_{EOF} - \tilde{\mu}_e|)^{1/2}(M_c + 24)^{7/4}}{5088}$$

The general equation for the ELFSE size resolution factor for charged-uncharged conjugates solely experiencing thermal-based diffusion is

$$S_m \approx 4\sqrt{\frac{\ln(2)D}{L\mu_0 E}} \times \frac{(|\tilde{\mu}_{EOF} - \tilde{\mu}_e|)^{1/2}(M_c + \alpha_1 M_u)^2}{\alpha_1 M_u} \tag{20}$$

Since we expect the mobilities to be independent of electric field, we can see that it is the total voltage drop, *i.e.*, the factor $E \times L$, that determines the resolution rather than either the electric field strength or the migration length independently. Also we see that the viscosity η of the electrophoresis medium does not affect the size resolution of the system (η cancels out in the ratio D/μ_0).

4

Deformation during ELFSE

L.C. McCormick, G.W. Slater, *to be submitted to Electrophoresis* (2006)

In this paper we investigate the behaviour of ELFSE molecules with next-generation labels and/or high field strengths. We make predictions as to the conditions leading to segregation of the DNA and label coils, as well as the expected impact on ELFSE performance. Optimal label architecture design and experimental conditions are suggested. This work also highlighted the potential gains in read length possible with high voltage use.

I was primarily responsible for the research and writing of this article, with the critical support and guidance of my supervisor.

Molecular deformation and performance of free solution electrophoresis of DNA-uncharged polymer conjugates: theoretical predictions

Laurette C. McCormick and Gary W. Slater

University of Ottawa, Canada

Running Title: Deformation of DNA-polymer conjugates during free solution electrophoresis

Correspondence: Gary W. Slater, Département de Physique, Université d'Ottawa, 150 Louis-Pasteur, Ottawa, Ontario, Canada, K1N 6N5

Email: gary.slater@uOttawa.ca

Fax: 613-562-5190

Abbreviations: **ELFSE**, End Labelled Free Solution Electrophoresis; **FWHM**, full width at half maximum.

Keywords: free solution electrophoresis; DNA sequencing; ELFSE; electrophoretic mobility; deformation.

Summary

Recent advancements to DNA sequencing by End Labelled Free Solution Electrophoresis (ELFSE) show the promise of this novel technique, which overcomes the need for a gel by using a label (or *drag-tag*) to render the free solution mobility of the DNA size-dependent. It is the attachment of an uncharged drag-tag molecule of a set size to all the various lengths of DNA in the sample that selectively slows down smaller DNA chains, which have less force to pull the drag-tag than larger DNA. Taking advantage of the modified hydrodynamic properties of labelled DNA, ELFSE has been used to successfully sequence up to about 100 bases of DNA in the absence of a gel or other sieving medium (Ren, H., Karger, A.E., Oaks, F., Menchen, S., Slater, G.W., Drouin, G., *Electrophoresis* 1999, *20*, 2501-2509). So far, only globally random coil conformations have been associated with ELFSE, i.e. the DNA and the label together form a single, undeformed hydrodynamic unit. This paper investigates the conditions under which the DNA and label coils will segregate becoming hydrodynamically distinct, in that the flow of the solvent around one does not affect the flow of the solvent around the other. Optimal experimental conditions as well as label architecture, based on various possible regimes, are predicted.

1. Introduction

End labelled free solution electrophoresis (ELFSE) is a relatively new technique that uses an uncharged (or nearly so) end label called a drag molecule (or *drag-tag*) to achieve separation of various lengths of DNA in free solution [1, 2, 3, 4]. Each DNA fragment is attached to a drag-tag of a set size in order to render the resulting conjugate's electrophoretic mobility length-dependent, and overcome the free-draining phenomenon which normally leads to co-migration of all lengths of DNA in free solution (except very small fragments [5, 6]) [7, 8, 9, 10]. It is this co-migration, or "free draining" property of DNA in free solution that usually necessitates the use of a gel which selectively slows down longer polymers more, by forcing them to collide more frequently with gel fibers [11]). Separation of DNA by the ELFSE technique is made possible by the drag-tag adding a set resistance (friction) to the motion of each DNA fragment, meaning that the more charged monomers a conjugate has (*i.e.*, the longer the DNA component), the more force it has to pull the drag-tag. Hence larger conjugates go faster, leading to size-based separation in free solution. ELFSE has been used to sequence up to about 100 base long ssDNA molecules in about 18 minutes in a 34 cm long capillary, using the globular protein streptavidin as a drag-tag [1]. Current research is directed at finding larger drag-tags, possibly with non-linear architectures [12, 13] to extend the read length of ELFSE [4].

The theory utilized to describe ELFSE and analyze the results was based in part on that developed by Long and co-workers for a polyelectrolyte experiencing both electric and non-electric forces, and more specifically for the electrophoresis of composite objects in free solution [14, 15, 16]. This theory predicted that for a conjugate

molecule composed of two polymer chains that are identical except in their electrical charge properties, the velocity is simply a uniformly weighted average of the individual monomer velocities [15]. For ELFSE, where the DNA chain generally differs in both stiffness and individual monomer size from the drag-tag, a conceptual grouping of both the drag-tag and DNA monomers into blobs having the same hydrodynamic radii allows for utilization of the theory of Long and co-workers for the case where the DNA and the drag-tag together form a random coil. Desruisseaux et al. [2] argued that for labels the size of streptavidin with field strengths achievable in capillary electrophoresis, the label and DNA should together form a single random coil (see Fig 1a). The experimental results have been consistent with the theory for ELFSE conjugates forming a single random coil, and the resulting equation for the velocity of a conjugate with N_D DNA monomers and N_L uncharged label monomers [4]:

$$v = \frac{N_D}{N_D + \alpha} v_D \qquad \text{Eq 1}$$

where $v_D = \mu_0 E$ is the velocity of the unlabelled DNA, and $\alpha = \alpha_1 N_L$ is the number of DNA monomers having the same hydrodynamic radius as the drag-tag; $\alpha_1 = \frac{b_L b_{L_K}}{b_D b_{D_K}}$ (where b_L is the label monomer size, b_D is the ssDNA monomer (base) size, b_{L_K} is the label Kuhn length and b_{D_K} is the ssDNA Kuhn length, a measure of stiffness) is the scaling factor that converts the number of label monomers into the number of hydrodynamically equivalent uncharged DNA monomers [4]. Hence $N = N_D + \alpha_1 N_L$ is

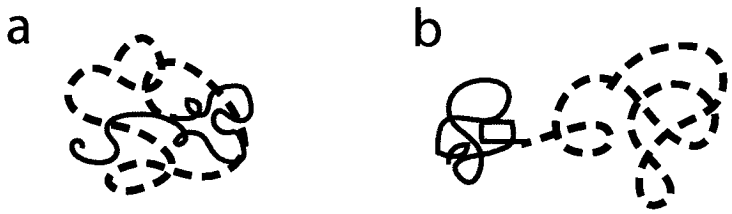


Figure 1: Schematic representation of the ELFSE conjugate comprising a DNA chain, represented by the dashed line, and a polymeric label, represented by the solid line; a) unsegregated conformation, b) segregated conformation. Unlike with the unsegregated conformation, with the segregated conformation the hydrodynamic flow of the solution around the DNA does not affect the label and vice versa.

the effective total number of hydrodynamically equivalent effective monomers in the conjugate. Under certain conditions, such as next generation labels providing more drag and/or higher field strengths, the label and DNA may start to pull apart and segregate (see Fig 1b), becoming hydrodynamically distinct, in that the flow of the solvent around one does not affect the flow of the solvent around the other; this deformation and the resulting impact on ELFSE performance is the focus of this paper.

2. Forces at work in segregated conjugates

We will now briefly outline the relevant forces and develop the electrophoretic mobility for a conjugate where the DNA and label can be considered as hydrodynamically segregated (see Fig 1b), i.e., there is no hydrodynamic coupling between the two sections (the disturbance of the flow of the buffer by the electrophoretic motion of the DNA does not affect the label and vice versa). The procedure is that of Long and co-workers [14, 15, 16]: they linearized the set of coupled electrohydrodynamic equations so that the problem could be solved by a superposition of the situations where either the electric or nonelectric forces act, a procedure that the authors point out is commonly used for rigid particles and is broadly applicable [14]. While this linearization procedure can not be used directly to determine the conformation of the conjugate, it can be used to predict the electrophoretic mobility for various known conformations. The conformation and the mobility can then be determined in a self-consistent manner; the breakdown of a self-consistent state then means a transition to another self-consistent state.

We begin by looking at the forces acting on the DNA. If it were unlabelled, the DNA would have two forces acting on it during electrophoresis, the force due to the

electric field F_{E_D}, and the corresponding force of friction f_{E_D} governed by the motion of the counterions surrounding it, which generates an almost uniform flow that freely penetrates the coil and is proportional to the electric field strength E. These two forces balance each other at the level of each monomer resulting in the unlabelled DNA free solution electrophoretic mobility, which is defined simply as the velocity v_D divided the electric field strength, $\mu_0 \equiv \frac{v_D}{E}$. This mobility is independent from the molecular size of the DNA, hence the need to either use a sieving gel or a drag-tag. When end labelled, it has the additional force acting on it due to the label, $F_{L\text{-}on\text{-}D}$ (subscript D refers to the DNA while subscript L refers to the label), and the corresponding force of friction $f_{L\text{-}on\text{-}D}$. This additional friction, unlike the friction due to the electric field (f_{E_D}), is not governed by the motion of the counterions, rather it is due to a hydrodynamic, Stokes-like flow around the coil that is proportional to $F_{L\text{-}on\text{-}D}$. These additional forces change the velocity of the DNA from v_D to v (note that $v < v_D$ since the label is uncharged) and break the local force balance mentioned before; hence the force of friction acting on the DNA due to its conjugation with the label, $f_{L\text{-}on\text{-}D}$, is simply the product of the coefficient of friction of the DNA ξ_D and the change in velocity $v - v_D$. The net force acting on the DNA is thus:

$$F_{E_D} - f_{E_D} + F_{L\text{-}on\text{-}D} - f_{L\text{-}on\text{-}D} = 0 \qquad \text{Eq 2}$$

where $f_{L-on-D} = \xi_D \times (v - v_D)$. Since the first two forces are balanced, $F_{E_D} = f_{E_D}$, leading to the velocity for the DNA alone v_D, the force on the DNA due to its conjugation with the label is, from Eq 2:

$$F_{L-on-D} = \xi_D \times (v - v_D). \qquad \text{Eq 3}$$

Note that $F_{L-on-D} < 0$, i.e. the positive direction has been chosen to be that of the electrophoretic motion such that the force of the label on the DNA is negative. Similarly, the force on the label due to the DNA is

$$F_{D-on-L} = \xi_L \times (v - v_L) \qquad \text{Eq 4}$$

where $v_L = 0$ because the label is uncharged in the situations of interest in this article. Since these two forces, that of the label on the DNA and that of the DNA on the label must be equal and opposite (this is the tension at the connecting point), setting Eq 3 equal to the negative of Eq 4 gives the velocity for a conjugate comprising hydrodynamically segregated sections [15]:

$$v = \frac{\xi_D v_D}{\xi_D + \xi_L}. \qquad \text{Eq 5}$$

Hence the electrophoretic velocity of a conjugate consisting of two hydrodynamically segregated sections is simply a uniformly weighted average of the individual velocity of

each section, where the weighting is now the hydrodynamic friction coefficient of each section, in contrast to the case for the unsegregated conjugate where the weighting is the effective number of monomers of each section (Eq 1). While one would expect that an increase in the size of the uncharged label (i.e. an increase in the label's friction coefficient) would result in a decrease in velocity, it is not necessarily intuitive that an increase in the friction coefficient of DNA, ξ_D, would increase the conjugate's velocity. This is however clearly predicted by Eq 5, and is a direct and counterintuitive result of the greater weighting a larger DNA would receive in the average determining the velocity. The friction coefficients must be determined carefully as they depend on conformation, which in turn depends on velocity; Eq 5 must therefore be solved self-consistently with a model for both $\xi_L(v)$ and $\xi_D(v_D,v)$.

Equations B and C respectively give the force applied by the label on the DNA, and vice versa, when the two components are hydrodynamically segregated; these are the forces that may act to deform the conjugate under certain conditions. For example, an extremely large label, of sufficient size to prevent the DNA from moving, could result in a situation similar to that of tethered (end anchored) DNA, along with the ensuant deformation [14, 15]. This paper is focused on the segregation that may result from the intramolecular mechanical forces present during ELFSE separations. We note that the approach taken assumes that the linearization procedure of Long and coworkers, where the electric and hydrodynamic forces are considered independently and then superposed, is still a valid approximation for high field strengths (and/or large labels). We also note that, for simplicity, we are neglecting the end effects that were predicted by the theory of Long and coworkers [17] and subsequently examined for the case of ELFSE [18, 19].

3. Critical electric field for segregation

It is expected that deformation of a conjugate during ELFSE would begin with the uncharged label separating and lagging slightly behind the charged DNA, resulting in a hydrodynamically segregated conjugate consisting of two coils (see Fig 1b). A simple estimate of the critical electric field at which this would occur can be obtained by comparing the force acting to deform the label (given by Eq 3) to that acting to resist deformation of the coiled hybrid molecule. We note that there may be some hysteresis because the drag forces depend upon the molecular conformation; i.e. the field at which a segregated conjugate would return to an unsegregated conformation may vary somewhat from the field at which an unsegregated conjugate would segregate. However this calculation is intended to give a rough estimate of the field at which segregation would occur.

The force acting to resist deformation is an elastic force due to the entropic property of polymers. The two ends of a Gaussian chain, separated by the end-to-end distance h, can be modeled as behaving like two points connected by a harmonic spring with a spring constant $k_{spring} = \frac{3k_BT}{\langle h^2 \rangle_0}$, where $\langle h^2 \rangle_0$ is the equilibrium average square end-to-end distance. This model hence yields the following entropic force acting to resist deformation of the coil [20]:

$$F_e = \frac{3k_BT}{\langle h^2 \rangle_0} h \qquad \text{Eq 6}$$

where k_B is the Boltzmann constant, T is the temperature. The average square equilibrium end-to-end distance is directly proportional to the square of the radius of gyration in the random coil R (note that excluded volume interactions are neglected):

$$\left\langle h^2 \right\rangle_0 = 6R^2 . \qquad \text{Eq 7}$$

Polymer conjugates are hydrodynamically coupled at equilibrium and together form a random coil [2]. The radius of gyration for a random coil can be written as [20]

$$R^2 = \tfrac{1}{6} N b b_K , \qquad \text{Eq 8}$$

from which it can be shown that the total radius of gyration of a conjugate comprising N_D monomers (with size b_D and Kuhn length b_{D_K}), and N_L monomers (with size b_L and Kuhn length b_{L_K}), is $R = \sqrt{R_D^{\,2} + R_L^{\,2}}$, where R_D is defined as the equilibrium radius of gyration of the DNA, (i.e. in the random coil conformation), and R_L is the corresponding random coil value for the label. Hence the average equilibrium end-to-end distance is given by $\left\langle h^2 \right\rangle_0 = 6\left(R_D^{\,2} + R_L^{\,2}\right)$. In contrast, the end-to-end distance for a segregated but otherwise undeformed conjugate (i.e. neither the DNA, nor the label, is deforming from their separate random coil conformations) is approximately equal to the sum of the individual component end-to-end distances (i.e. $R \approx R_D + R_L$):

$$h \cong h_D + h_L = \sqrt{6}\left(R_D + R_L\right) \qquad \text{Eq 9}$$

where we have used Eq 7 for the end-to-end distance of a random coil. Hence the entropic force resisting segregation, for a segregated but otherwise undeformed conjugate is:

$$F_e \cong \frac{3k_BT}{\sqrt{6}} \times \frac{\left(R_D + R_L\right)}{\left(R_D{}^2 + R_L{}^2\right)}. \qquad \text{Eq 10}$$

The counter force, that acting to segregate the two components, is the absolute value of the force of the label on the DNA, given by Eq 3, which we combine with Eq 5:

$$\left|F_{L\text{-}on\text{-}D}\right| \cong 4\pi\eta\mu_0 E \times \frac{R_D R_L}{R_D + R_L} \qquad \text{Eq 11}$$

where we used the definition $v_D = \mu_0 E$, and the fact that the friction coefficient $\xi_i = 6\pi\eta R_{H_i}$ for a random coil with hydrodynamic radius $R_{H_i} \cong 2/3\,R_i$, where R_i is the equilibrium radius of gyration of section i, in a buffer of viscosity η [21], is

$$\xi_i^0 \cong 4\pi\eta R_i. \qquad \text{Eq 12}$$

Setting these two opposing forces ($\left|F_{L-on-D}\right|$ and F_e) equal gives an estimate for the critical electric field at which a segregated conjugate would return to an unsegregated state, E^0_{seg},

$$4\pi\eta\mu_0 E^0_{seg} \times \frac{R_D R_L}{R_D + R_L} \cong \frac{3k_B T}{\sqrt{6}} \frac{\left(R_D + R_L\right)}{\left(R_D^2 + R_L^2\right)}, \qquad \text{Eq 13}$$

yielding, as an estimate of the critical field for segregation,

$$E^0_{seg} \cong E_0 \times \frac{\left(1+\sqrt{\frac{\alpha}{N_D}}\right)^2}{\sqrt{N_D \alpha} \times \left(1+\frac{\alpha}{N_D}\right)}, \qquad \text{Eq 14}$$

where we have used the definition $\alpha = \alpha_1 N_L$ and Eq 8. The universal field intensity E_0 is given by:

$$E_0 \equiv \frac{18k_B T}{\sqrt{96}\pi\eta\mu_0 b_D b_{D_K}}. \qquad \text{Eq 15}$$

Note that since F_{L-on-D} and F_{D-on-L} are equal and opposite forces, we could also have set $F_e = F_{D-on-L}$ (i.e. using Eq 4 instead of Eq 3) and obtained Eq 13.

This rough estimate of the critical field for segregation indicates that for an increase in the size of either the label and/or the DNA, segregation would occur for lower

fields. In other words, it is expected that a larger conjugate would be easier to segregate. Under typical experimental conditions, e.g. $T = 300$ K, $\eta = 1$ cP, $\mu_0 = 3\times10^{-4}$ cm^2/Vs, $b_D = 0.43$ nm, and $b_{D_K} = 7$ nm (the last two values are typical for ssDNA), the prefactor is $E_0 \cong 268$ kV/cm; since this far exceeds typical electric field strengths used in electrophoresis, the second factor on the right hand side of Eq 14 would need to be very small for segregation to occur. For the pioneering experiments of Ren et al. with a streptavidin label [1], the α value was about 30 and the read length was about 100 bases; under these conditions the critical field is very high, $E^0_{seg} \approx 9$ kV/cm. However, current goals are for a read length of about 400 bases with a label having an alpha value of at least 150 [4]. If this were achieved, deformation would occur at lower field strengths, i.e. $E^0_{seg} \approx 2$ kV/cm.

4. Electrophoretic mobility of segregated ELFSE conjugates

In this section we address the mobility of segregated but otherwise undeformed conjugates (see Fig 1b). Defined as the ratio of velocity to electric field, the mobility of the ELFSE conjugate in the segregated conformation, where both the label and the ssDNA remain in random coil conformations can easily be determined through use of Eq 5 where the random coil friction coefficients are given by Eq 12:

$$\mu = \mu_0 \frac{R_D}{R_D + R_L} = \mu_0 \frac{1}{1+\sqrt{\alpha/N_D}} \qquad \text{Eq 16}$$

where we again used the definition $v_D = \mu_0 E$, and Eq 8. As with ELFSE where the conjugate is unsegregated, the mobility increases with increasing DNA size (i.e. a larger engine pulls the conjugate faster), and decreases with increasing label size (i.e. increasing frictional drag). We note that the mobility for the unsegregated conjugate (see Fig 1a) can be written as (using Eq 1):

$$\mu = \mu_0 \frac{1}{1 + \alpha / N_D}. \qquad \text{Eq 17}$$

The mobilities for the segregated and unsegregated states are presented in Fig 2, scaled by the length-independent free solution mobility μ_0. The mobility for the segregated conformation becomes less than that of the unsegregated conformation when the number of DNA monomers exceeds that of the effective number of monomers of the label, i.e. when $N \geq \alpha$. This is important to know for future use of larger labels in ELFSE because it may manifest as size inversion. For example, if segregation were to occur abruptly for $N_D = 4\alpha$ bases at a given electric field strength, one can see from Fig 2 that the mobility would drop from about 80% of the free solution value (μ_0) to only 67% of that value, for conjugates of this size. Hence, the order of elution would no longer directly correspond to increasing lengths of DNA. Clearly this would represent a large problem for sequencing and great care would need to be taken in the future when larger labels will make segregation possible. However, if the transition from the random coil conformation to the segregated conformation is smooth rather than abrupt, we may simply see a change in

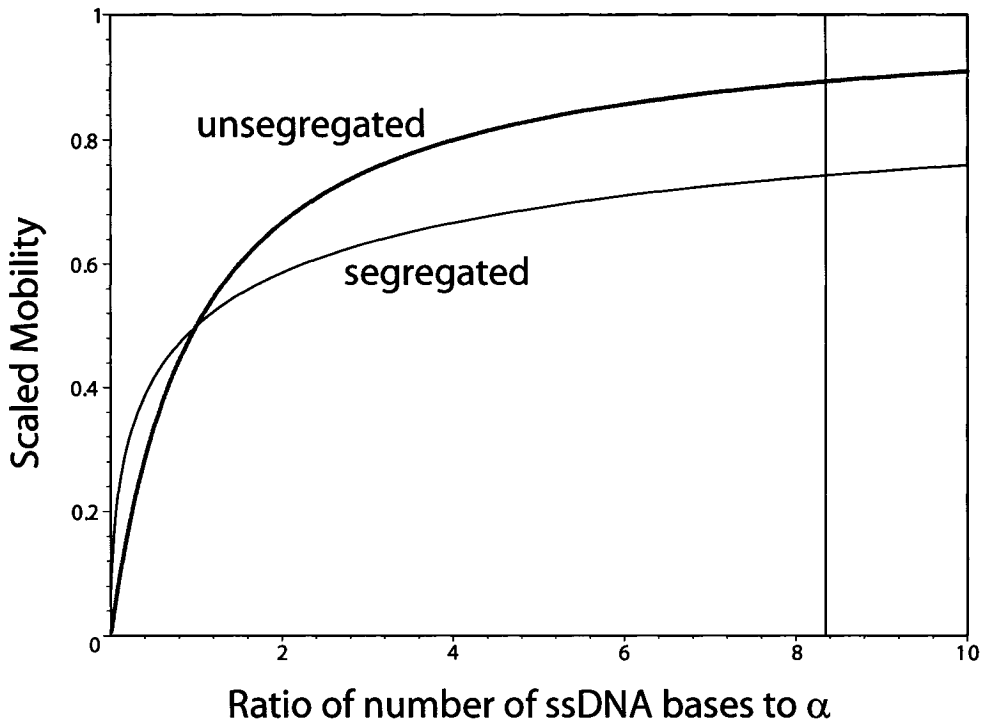


Figure 2: Mobility of the segregated (but otherwise undeformed) and unsegregated conformations as a function of the ratio of the number of ssDNA bases N_D to α, scaled by the length-independent free solution mobility μ_0. The rate of change of the mobility with respect to the number of ssDNA bases (as measured by the slope of these curves) is lesser for the segregated state up until about $N_D / \alpha = 8.3524$; this crossover is indicated by the vertical line.

peak spacing rather than a size inversion, with the transition from the top curve to the lower curve occurring over a large range of sizes.

The only difference between the equations for the mobility in the segregated and unsegregated regimes is that the term α/N_D becomes square rooted once segregation occurs. Since $\alpha/N_D < 1$ for the majority of ELFSE conjugates except really short DNA for which the peaks are needlessly over-resolved [22], the square root increases this term (i.e., $\sqrt{\alpha/N_D} > \alpha/N_D$ when $\alpha/N_D < 1$), thereby decreasing the mobility. The crossover is at $N_D = \alpha$, as can be seen in Fig 2. This decreased mobility is expected to allow more time for greater separation to occur, which should specifically help with the resolution of larger conjugates which experience critically small peak spacing in the unsegregated state. However, there is a competing effect due to the decreased sensitivity of the mobility on the number of DNA monomers N_D in the segregated state, which would act to decrease peak spacing. From equations Y and V, one can find that the rate of change of the mobility with respect to N_D is lesser for the segregated state up until $\frac{N_D}{\alpha} = \frac{1}{4}\left(1+\sqrt{5}+\sqrt{2+2\sqrt{5}}\right)^2 \cong 8.3524$. We examine the effect of segregation on peak spacing next.

5. Peak spacing of segregated ELFSE conjugates

The peak spacing for a segregated (but otherwise undeformed) conjugate can simply be taken as the derivative of the arrival time, $t = l_{cap}/\mu E$ (where l_{cap} is the effective capillary length) with respect to the number of DNA bases, since each consecutive peak

corresponds to one additional ssDNA base. Hence, using Eq 16 for the mobility we obtain the peak spacing:

$$\left|\frac{\partial t_{seg}}{\partial N_D}\right| = \frac{l_{cap}}{\mu_0 E} \times \frac{\sqrt{\alpha}}{2N_D^{3/2}}. \qquad \text{Eq 18}$$

The corresponding peak spacing for the unsegregated conformation is obtained similarly, using Eq 17 instead of Eq 16:

$$\left|\frac{\partial t_{unseg}}{\partial N_D}\right| = \frac{l_{cap}}{\mu_0 E} \times \frac{\alpha}{N_D^{2}}. \qquad \text{Eq 19}$$

The above two equations, X and W, illustrate that the peak spacing is greater for the segregated conformation than the unsegregated conformation whenever $\frac{N_D}{\alpha} > 4$, regardless of the field strength; i.e. when the DNA size gives it at least four times the amount of frictional drag of the label, peak spacing is increased by segregation.

6. Resolution of segregated ELFSE conjugates

In this section we develop the resolution for DNA sequencing purposes. We define the size resolution factor as the ratio of the temporal full width at half maximum ($FWHM$) of the peak, to the temporal peak spacing at the detector:

$$S \equiv \frac{FWHM}{\left|\frac{\partial t}{\partial N_D}\right|}. \qquad \text{Eq 20}$$

The size resolution factor gives the smallest difference in the number of ssDNA bases that can be resolved; hence $S = 1$ base is required for single monomer resolution. If only thermal-based band broadening is present, the numerator can be found by dividing the spatial peak standard deviation $\sigma = \sqrt{2Dt}$ (where the diffusion coefficient is $D = k_B T/\xi$) by the velocity (μE) to convert to temporal spacing, and multiplying by the factor $\sqrt{8\ln 2}$ to convert from standard deviation to full width at half maximum:

$$FWHM = \sqrt{\frac{16\ln(2) k_B T l_{cap}}{(\mu E)^3 \xi}} \qquad \text{Eq 21}$$

Since we have $R \approx R_D + R_L$ for the segregated but otherwise undeformed conjugate, making use of Eqs 12 and 8, along with the definition $\alpha = \alpha_1 N_L$ where $\alpha_1 = \frac{b_L b_{L_K}}{b_D b_{D_K}}$, we find $\xi \cong 4\pi\eta\sqrt{b_D b_{D_K}}/6\left(\sqrt{N_D} + \sqrt{\alpha}\right)$ for the friction coefficient. Along with the mobility for the segregated conformation given by Eq 16, this yields:

$$FWHM_{seg} \cong \sqrt{\frac{\sqrt{96}\ln(2) k_B T l_{cap}}{(\mu_0 E)^3 \pi\eta\sqrt{b_D b_{D_K}}} \times \frac{\left(1 + \sqrt{\alpha/N_D}\right)^3}{\sqrt{N_D} + \sqrt{\alpha}}}. \qquad \text{Eq 22}$$

The size resolution factor for the segregated (but otherwise undeformed regime) can hence be given as:

$$S_{seg} \approx 2\sqrt{\frac{V_s}{V}}\frac{N_D^{5/4}}{\alpha^{1/2}}\left(1+\sqrt{\alpha/N_D}\right) \qquad \text{Eq 23}$$

where we used Eqs X and 22, and $V = E \times l_{cap}$ is the voltage drop along the capillary length, while the characteristic voltage V_s is a constant given by

$$V_s \equiv \frac{\sqrt{96}\ln(2)k_B T}{\pi\eta\mu_0\sqrt{b_D b_{D_K}}}. \qquad \text{Eq 24}$$

Under typical experimental conditions, T = 300 K, η = 1 cP, μ_0= 3×10^{-4} cm^2/Vs, b_D = 0.43 nm, and b_{D_K} = 7 nm, this constant is $V_s \approx 0.17$ V. In a similar way, using Eq 17 instead of Eq 16, we obtain the temporal full width at half maximum for the unsegregated case:

$$FWHM_{unseg} \approx \sqrt{\frac{\sqrt{96}\ln(2)k_B T l_{cap}}{(\mu_0 E)^3 \pi\eta\sqrt{b_D b_{D_K}}} \times \frac{\left(1+\alpha/N_D\right)^3}{\sqrt{N_D+\alpha}}} \qquad \text{Eq 25}$$

which, along with Eq 19 gives the corresponding size resolution factor:

$$S_{unseg} \approx \sqrt{\frac{V_s}{V}} \frac{N_D^{7/4}}{\alpha} \left(1 + \alpha / N_D\right)^{5/4}. \qquad \text{Eq 26}$$

From Eqs 23 and 26 it is clear that higher voltages are desirable in ELFSE since they increase the resolution. It is interesting to note that the resolution in ELFSE does not depend on either the electric field strength or the capillary length individually, but only on their product, the voltage. This is unlike gel electrophoresis where biased reptation, which occurs for higher electric fields, restricts the field strength that can be used [11]. These equations also illustrate that the unsegregated state provides better resolution than the segregated state, up until a crossover which occurs at $N^* = 5.3465\alpha$ bases (this can be obtained by setting the two equations equal and solving numerically); beyond N^* segregation gives better resolution. The crossover depends exclusively on the relative sizes of the label and DNA. The maximum DNA length that can be sequenced, known as the *read length*, (i.e., for which S=1 base) can be found by setting Eq 23, or Eq 26 for the unsegregated case, equal to one base, and solving numerically.

Rough equations for the read length can be obtained in the $N_D >> \alpha$ limit for the segregated and unsegregated cases respectively:

$$\overline{N}_D \approx \frac{\alpha^{2/5}}{2} (V/V_s)^{2/5} \qquad \text{Eq 27}$$

$$\overline{N}_D \approx \alpha^{4/7} (V/V_s)^{2/7} \qquad \text{Eq 28}$$

We note that the read length increases less quickly with α, and more quickly with V, in the segregated state than in the unsegregated state. This is due to the square root in the

mobility equation (Eq 16) predicted by Long and coworkers for segregated conformations.

7. Small α regime

The read length is shown in Fig 3 for the specific example of $\alpha = 30$, which is the value quoted by Ren et al for their work with the label molecule streptavidin [1], and the following typical experimental conditions: $T = 300$ K, $\eta = 1$ cP, $\mu_0 = 3\times10^{-4}$ cm^2/Vs, and $b_D = 0.43\,\text{nm}$, $b_{D_K} = 7\,\text{nm}$, for which $V_s \approx 0.17\,\text{V}$. The crossover to the segregated conformation providing better resolution, as indicated by the horizontal line, occurs at $N^* = 5.3465\alpha = 160\,\text{bases}$. Unfortunately, for the experimental conditions of Ren et al, where the field strength was 15 kV and the capillary lengths ranged from 45 to 110 cm, the critical field for segregation (as given by Eq 14) would only be reached for exceedingly large conjugates, having at least 24 672 ssDNA bases (for the shortest capillary, or more for a longer capillary). Therefore, only the curve corresponding to the unsegregated case is meaningful here. If, however, a higher voltage were used, for example 30 kV instead of 15 kV, then segregation could occur for much smaller conjugates. Since the resolution of ELFSE conjugates, either segregated or unsegregated (Eqs 23 and 26), does not depend on the electric field or capillary length individually, but only on the voltage, changing the capillary length for a set voltage does not affect the resolution for a given conformation, but it may make reaching the critical electric field possible. At 30 kV, the read length could be increased from 199 bases to 218 bases simply by shortening the capillary length so as to take advantage of the increased resolution with the segregated conformation. In order to segregate all conjugates with 160

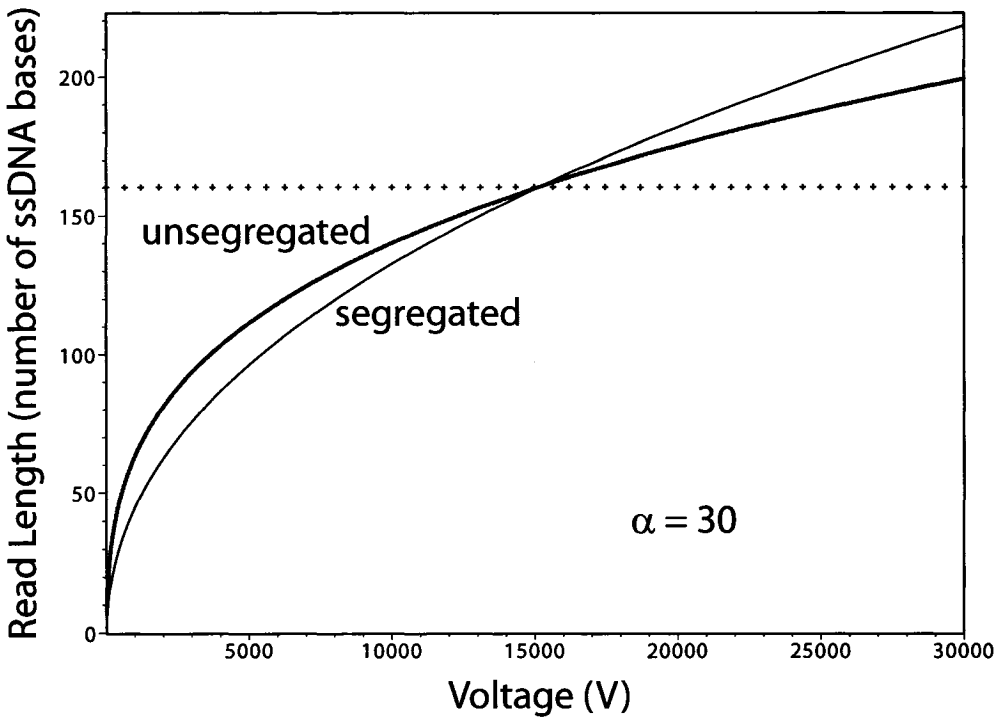


Figure 3: Predicted optimal read length for both the segregated and unsegregated conformations, as a function of voltage (V) for an α value of 30 and typical experimental conditions.

ssDNA bases or more, the required field strength would be 6.7 kV/cm (from Eq 14). (This field strength is significantly higher than those normally used and it may be necessary to, for example, use very narrow channels and low conductivity buffers to prevent the currents from getting too high and leading to problematic Joule heating and buffer depletion.) With a voltage of 30 kV, this means using a capillary length, or separation length on a electrophoresis chip, of only 4.5 cm. While it was previously unexpected that merely using a shorter capillary could increase the read length, this is simply a result of the correspondingly increased field strength causing larger conjugates to segregate under conditions where segregation provides better resolution (i.e. $N \geq N^*$). One caveat however, is the possibility of a size-inversion if the segregation transition takes place over a short range of DNA sizes. This potential phenomenon must be carefully watched for, as discussed above.

We note that we have predicted a read length of about 160 bases for the voltage used in the Ren et al. experiments; their read length was only about 100 bases [1]. However, the typical conditions we have used in our predictions may vary somewhat from their experimental conditions. Also, since we assume only diffusional band broadening, if they had less than optimal conditions, this may also have decreased their read length, contributing to the discrepancy.

8. Large α regime

Since $N^* = 5.3465\alpha$ bases, taking advantage of the segregated regime providing better resolution is only feasible for smaller labels. This is because larger labels (i.e. larger α values) mean that the crossover occurs for larger DNA, and resolving larger DNA

requires very high field strengths. For example, for a label with $\alpha = 150$, the crossover occurs for $N^* = 802$ bases; under the typical conditions we are considering, resolving DNA of this length would require an unrealistic voltage of 169 kV (see Fig 4). Hence DNA lengths for which better resolution comes from segregation ($N > N^*$) are not being resolved and therefore, for the smaller lengths of DNA that are being resolved, the unsegregated conformation provides better resolution ($N < N^*$).

Longer capillaries may provide one means of keeping the conjugates in the unsegregated conformation, by lowering the field for a given voltage below the critical one for segregation. For example, for $\alpha = 150$ and a voltage of 30 kV, the read length for the unsegregated conformation is predicted to be about 450 bases, as compared to 359 bases for the segregated conformation. (The critical read lengths are obtained by setting either Eq 26, for the unsegregated case, or Eq 23 for the segregated case, equal to one base, and solving for N.) The critical field for segregation for a conjugate of this size is 1.9 kV/cm (from Eq 14); at 30 kV, a capillary about 15.5 cm or longer in length would be required to keep the field strength below this critical one.

9. Selecting the optimal conjugate conformation

The optimal conjugate conformation is determined by N^*; if the read length is greater than this critical value then the segregated state gives better resolution, whereas if the read length is smaller than this critical value then the unsegregated state is preferable. Depending on whether conjugates with $N > N^*$ bases are resolvable or not (i.e the read length is greater than N^* or not), either a shorter or longer capillary length can be chosen to select the optimal conformation for the given conditions.

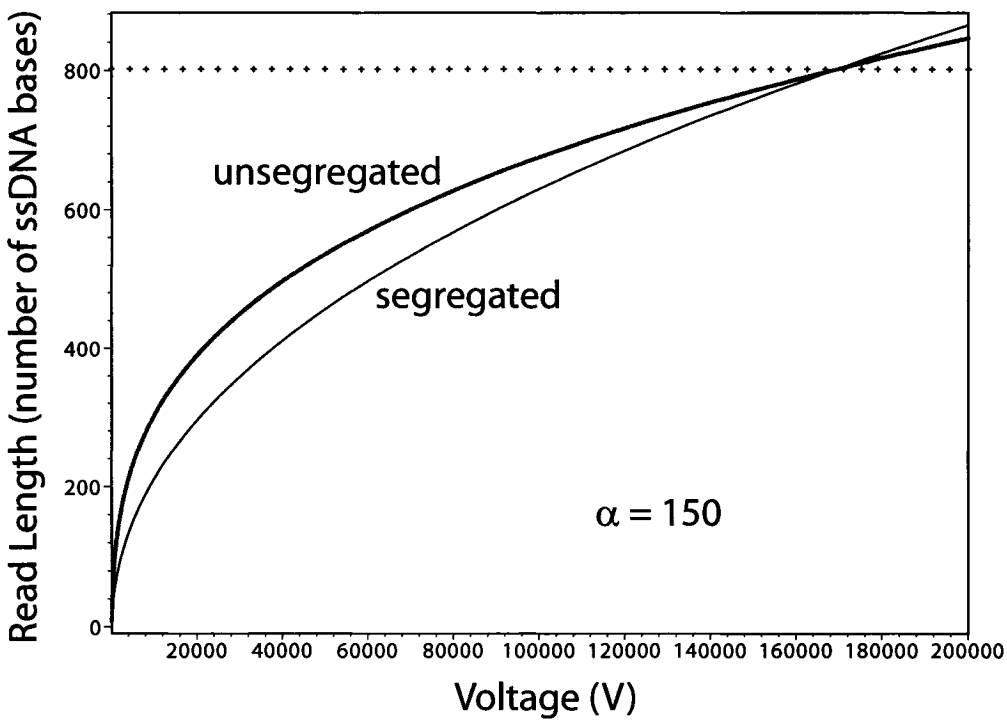


Figure 4: Predicted optimal read length for both the segregated and unsegregated conformations, as a function of voltage (V) for an α value of 150 and typical experimental conditions.

A shorter capillary is desirable when the α value is small enough that obtainable voltages provide resolution for $N \geq N^* = 5.3465\alpha$ bases, i.e. when the read length (which is smaller for smaller α values) is greater than N^*. This is because for conjugates with $N \geq N^*$ bases the segregated conformation provides better resolution, and if a shorter capillary length is chosen properly it may increase the field strength for the given maximum voltage to that required for segregation, thereby enabling better resolution via segregation. The voltage required to resolve $\overline{N} > N^*$ bases for a given α can be determined by setting the resolution factor from Eq 26, $S(\alpha, \overline{N}, V)$, equal to one base and solving for the voltage.

Conversely, a longer capillary is desirable when the α value is too large for obtainable voltages to provide resolution for $N \geq N^* = 5.3465\alpha$ bases. From the equations for resolution (Eqs 23 and 26) we see that a higher voltage gives better resolution and hence is preferable, however when the read length is less than N^*, better resolution is obtained with the unsegregated conformation and the field strength should be kept below the critical one for segregation. With a properly chosen longer capillary length, the field strength would decrease (for the given maximum voltage) to below the critical field for segregation, thereby maintaining the unsegregated conformation for which the lengths that are being resolved, $N < N^*$, are afforded better resolution.

The prediction that segregation decreases resolution for conjugates comprising DNA with less than $N^* = 5.3465\alpha$ bases can help guide label design. For labels that are too large to allow resolution of N^* bases with obtainable voltages, the label architecture should be one that discourages segregation of the label and the DNA chain due to steric repulsion (i.e. where the label architecture inhibits the mingling of DNA and label chains

in a single random coil) so as to avoid decreasing the read length (see Fig 4). For a label with a given α value, Eq 26 can be used to predict whether or not N^* bases would be resolved with the maximum voltage (i.e. $S(\alpha, N^*, V) \leq 1$ if resolved); if not, then the label should be designed so as to avoid steric repulsion.

10. Discussion

We have developed in this paper the theoretical predictions for ELFSE conjugate behaviour under conditions which could lead to segregation, based on the theory of Long et al. [15]. We illustrated how the capillary length could be adjusted so as to induce the optimal conformation for a given label size and voltage.

Higher α values and voltages are always preferable in terms of ELFSE resolution. However, a higher α value means that the crossover to better resolution with segregation occurs for a larger conjugate, $N^* = 5.3465\alpha$. This larger conjugate may not be resolvable even with high voltages, such that the read length is less than the critical value for the crossover $\overline{N} < N^*$. Under these conditions the unsegregated conformation gives better resolution and hence if possible, it should be selected through use of a capillary long enough to keep the field strength (for the given voltage) lower than the critical one for segregation (see Eq 14). In summary, although segregation is expected to provide better resolution for conjugates with $N > N^*$, with next-generation labels this crossover occurs for DNA sizes that are not resolvable with currently available voltages and hence the regime for these labels is one in which segregation should be avoided by keeping the electric field strength below the critical one for segregation.

The prediction that segregation decreases resolution for conjugates comprising DNA with less than $N^* = 5.3465\alpha$ bases can also help guide label design. For labels that are too large to allow resolution of N^* bases with obtainable voltages, the label architecture should be one that discourages segregation of the label and the DNA chain due to steric repulsion. Under these conditions steric segregation would decrease the read length (see Fig 4). Developing large labels that induce steric segregation from the DNA were, until now, considered an experimental goal.

It is important to note that once the force acting to deform the ELFSE conjugate reaches that required for segregation, increasing this force is expected to induce further deformation by stretching the DNA and/or a deformable label. Stretching would further impact the conjugate mobility because a stretched conformation is expected to have a lower mobility due to its increased friction coefficient. (The equation determining the mobility of a segregated conjugate, the first expression of Eq 16, is expected to also apply to segregated and stretched conjugates, with the hydrodynamic radii of the two components increasing with an increase in the extent of deformation.) The larger friction coefficient would also be expected to decrease diffusion, and therefore decrease peak broadening, thereby increasing the resolution. Hence the predictions given in this paper, applying exclusively to either unsegregated conjugates or conjugates where the components are segregated but otherwise undeformed, would need to be adjusted to take into account any further deformation. Depending on the conditions, it may be possible to have a range of conformations, from unsegregated to segregated but otherwise undeformed, to segregated with various extents of deformation. Since larger conjugates are easier to deform, we expect that this progression would be seen over increasing DNA

sizes, perhaps contained within a single DNA sequencing sample. Further theoretical analysis (and/or computer simulations) is necessary to predict the effect of stretching from the segregated state.

We note that the read lengths presented in this paper are based on ideal conditions where the band width is limited only by diffusion, all other sources of band broadening being negligible. We have assumed an ideal, negligible band loading width; a non-negligible loading width would decrease the resolution and affect our predictions. Clearly care must be taken to avoid Joule heating of the solution in the capillary, which may arise with use of higher voltages causing increased current as an additional source of band broadening.

We have neglected the end effects predicted by Long et al. [17], for which we amended our theory specifically for ELFSE in a previous paper [18]. These effects, confirmed experimentally in [19], result from monomers near the end of the ELFSE conjugate receiving a higher weighting in the average determining the mobility. We use a uniformly weighted average here for the sake of simplicity; however the end effects could be taken into account by properly weighting the mobility average. Clearly such second order effects are not expected to affect our main conclusions.

Acknowledgements

The authors would like to thank Robert J. Meagher of Sandia National Laboratories (USA) for his experimental insight. The work was supported, in part, by a Natural Science and Engineering Research Council of Canada Discovery Grant to GWS, and a University of Ottawa Admission Scholarship to LM, as well as the National Institutes of

Health (NIH) of the U.S.A. (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 or the NIH.

References

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

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

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

[4] Meagher, R.J., Won, J.-I., McCormick, L.C., Nedelcu, S., Bertrand, M.M., Bertram, J.L., Drouin, G., Barron, A.E., Slater, G.W., *Electrophoresis* 2005, *26*, 331–350.

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

[6] Stellwagen, N.C., Stellwagen, E., *Electrophoresis* 2002, *23*, 1935-1941.

[7] Olivera, B.M., Baine, P., Davidson, N., *Biopolymers* 1964, *2*, 245-257.

[8] Völkel, A.R., Noolandi, J., *Macromolecules* 1995, *28*, 8182-8189.

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

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

[11] Viovy, J.L., *Rev. Mod. Phys.* 2000, *72*, 813-872.

[12] Haynes, R.D., Meagher, R.J., Won JI, Bogdan F.M., Barron AE, *Bioconjugate Chemistry* 2005, *16*, 929-938.

[13] Nedelcu, S., Slater, G. W., *Electrophoresis* 2005, *26*, 4003-4015.

[14] Long, D., Viovy, J.-L., Ajdari, A., *Physical Review Letters* 1996, *76*, 3858-3861.

[15] Long, D., Viovy, J.-L., Ajdari, A., *J. Phys.: Condens. Matter* 1996, *8*, 9471-9475.

[16] Long, D., Ajdari, A., *Electrophoresis* 1996, *17*, 1161-1166.

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

[18] McCormick, L.C., Slater, G.W., Karger, A.E., Vreeland, W.N., Barron, A.E., Desruisseaux, C., and Drouin, G., *J. Chrom. A* 2001, *924*, 43-52.

[19] Meagher, R.J., McCormick, L.C., Haynes, R.D., Won, J.-I., Lin, J.S., Slater, G.W., Barron, A.E., *Electrophoresis* (in press).

[20] Teraoka, I., "Polymer Solutions, An Introduction to Physical Properties", *Wiley Interscience*, New York, 2002.

[21] Munk, P., "Introduction to Macromolecular Science", *Wiley*, New York, 1989.

[22] McCormick, L.C., Slater, G.W., *Electrophoresis* (in press).

5

Free Solution Conjugate Electrophoresis (FSCE)

L.C. McCormick, G.W. Slater, A.E. Karger, W.N. Vreeland, A.E. Barron, C. Desruisseaux, G. Drouin, *Journal of Chromatography A* **924**, 43–52 (2001)

In this article we present the general theory for Free Solution Conjugate Electrophoresis (FSCE), a technique analogous to End Labelled Free Solution Electrophoresis (ELFSE) where instead of using a uniform set of uncharged drag-tag molecules to separate various lengths of DNA, a uniform set of DNA molecules are used as "'engines"' to separate various lengths of uncharged polymers in a sample. We show that for the most favourable, diffusion-limited electrophoresis conditions, there is actually an optimal DNA size to achieve the separation of a given polymer sample.

Please note that at the time this paper was written, the end effect had not yet been carefully investigated; see Chapter 2 for an in-depth analysis of the end effect.

This article was researched and written together with my supervisor, and borrowed from an analogous theory written by Claude Desruisseaux for ELFSE. Experimentalists in the Annelise Barron group (Northwestern University) as well as Guy Drouin (University of Ottawa) and Achim Karger (Applied Biosystems) assisted with the comparison to experimental data.

ELSEVIER

Journal of Chromatography A, 924 (2001) 43–52

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Capillary electrophoretic separation of uncharged polymers using polyelectrolyte engines
Theoretical model

L.C. McCormick[a], G.W. Slater[a,*], A.E. Karger[b], W.N. Vreeland[c], A.E. Barron[c], C. Desruisseaux[d], G. Drouin[d]

[a]*Department of Physics, University of Ottawa, 150 Louis-Pasteur, Ottawa, Ontario K1N 6N5, Canada*
[b]*Applied Biosystems, Foster City, CA, USA*
[c]*Department of Chemical Engineering, Northwestern University, Evanston, IL, USA*
[d]*Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*

Abstract

We recently demonstrated that the molecular mass distribution of an uncharged polymer sample can be analyzed using free-solution capillary electrophoresis of DNA–polymer conjugates. In these conjugates, the DNA is providing the electromotive force while the uncharged polydisperse polymer chains of the sample retard the DNA engine with different amounts of hydrodynamic drag. Here we present a theoretical model of this new analytical method. We show that for the most favourable, diffusion-limited electrophoresis conditions, there is actually an optimal DNA size to achieve the separation of a given polymer sample. Moreover, we demonstrate that the effective friction coefficient of the polymer chains is related to the stiffness of the two polymers of the conjugate, thus offering a method to estimate the persistence length of the uncharged polymer through mobility measurements. Finally, we compare some of our predictions with available experimental results.

Keywords: Mathematical modelling; Capillary electrophoresis; Polymers; DNA

1. Introduction

The polydispersity of a polymer solution is not easily determined using conventional methods such as gel permeation chromatography [1] and mass spectrometry [2]. Recently, we proposed and tested a new method to characterize the distribution of the varying degrees of polymerization of a water-soluble, uncharged polymer species [3]. The method is based on the idea that if a set of "engines" of uniform size and charge distribution is conjugated in an unbiased way to all the polymer chains of the sample, one can in fact use free-solution capillary electrophoresis (CE) to study the distribution of polymer sizes. Indeed, the uncharged polymer then acts like a parachute, or "drag", and the retardation due to the drag molecule is directly proportional to its contour length. We called this method free solution conjugate electrophoresis (FSCE).

In our original paper [3], we showed the separation of each of 3400, 5000 and 20 000 nominal molecular mass poly(ethylene glycol) (PEG) samples using conjugated oligomeric DNA engines 20 and 35 bases long. Excellent quantitative results were obtained and compared with matrix-assisted laser de-

*Corresponding author.
E-mail address: gslater@science.uottawa.ca (G.W. Slater).

sorption ionization time-of-flight (MALDI-TOF) MS data. Two results were particularly remarkable: (1) the 20-base DNA engine appeared to always give better results for a given set of experimental conditions, and (2) the effective friction coefficient of a PEG monomer was only about 14% of that of a DNA monomer. In this article, we derive the theory of FSCE and we examine these two results very carefully, i.e. we make predictions regarding the optimal engine size and the effective friction coefficient of the drag polymer.

Early experiments with FSCE-like separation techniques include the work of Stefansson and Novotny [4] who achieved separation of uncharged oligosaccharides through complexation with charged tags, and Bullock [5] who has also successfully separated uncharged PEG polymers through derivatization with phthalic anhydride. Although both of these early experiments were promising, they were not as precise and as easily controllable as the recent work of Vreeland et al., for example, in specifically varying the engine size. Nevertheless, the theory developed here could be adjusted for the experimental situations of the early studies.

Note that FSCE is essentially the complementary separation method to ELFSE, or end-labelled free-solution electrophoresis [6–8]. In ELFSE, a uniform (i.e., monodisperse) drag molecule is used to achieve the CE separation of a polydisperse polyelectrolyte sample. For instance, using streptavidin (a globular, uncharged protein) as the drag-label, we were able to sequence up to about 110 base long DNA samples in less than 20 min [8]. DNA sequencing using CE normally requires the use of a sieving polymer matrix (e.g., a gel or an entangled polymer solution) [9], but the extra drag added by the label allows us to achieve separation based only on the free-solution hydrodynamic properties of the conjugates. We have presented a theory of ELFSE in a recent paper [10], but this theory does not directly apply to FSCE since the latter is a method to resolve the drag label itself.

We will thus study the free-solution capillary electrophoresis of an uncharged sample composed of a polydisperse polymer solution conjugated to a monodisperse polyelectrolyte chain (Fig. 1). Hence, both components of the resulting complex are assumed to be flexible, water-soluble polymers. It would be quite easy to rewrite the theory for the situation where the "engine" is a globular object such as a colloidal particle or a protein using the theoretical elements described in this article.

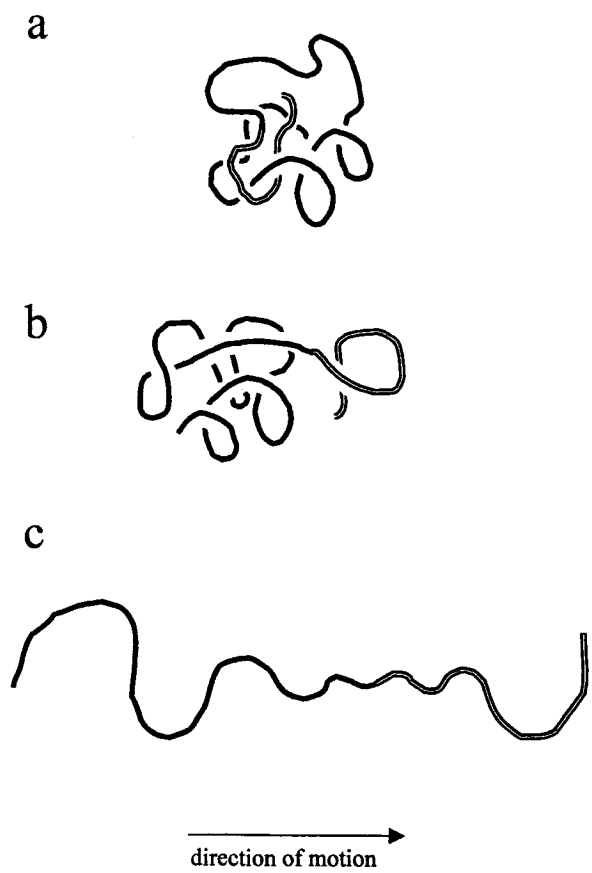


Fig. 1. Schematic representation of a block polyampholyte. The solid line shows the uncharged block while the double line shows the (shorter) charged block. (a) Equilibrium conformation; (b) segregation of the two blocks; (c) segregation and stretching of the blocks.

2. Theory for homogeneous charged–uncharged block copolymers

In this section, we examine the electrophoresis of a block copolymer consisting of a linear chain of charged monomers joined to a linear chain of uncharged monomers. The charged and uncharged blocks considered here are assumed to be homoge-

neous in the sense that they share the same flexibility (i.e. the same Kuhn length, b_K) and monomer size b.

2.1. The mobility

The mobility of polyampholytes was investigated by Long et al. [11]. Their theoretical results were fairly general, allowing for various charge distributions, including the specific case of our charged–uncharged complex. The theory takes into account the following five factors affecting the mobility of a polyampholyte in an approximately Gaussian conformation. (1) The force experienced by the charged monomers due directly to the electric field. (2) Hydrodynamic interactions that arise from the effect of the electric field on both the charged monomers and the counter ions surrounding them in the solution. (3) The tension along the polymer chain. (4) Thermal agitation. (5) The additional flow resulting from the previous two non-electrical factors. Factors 2 and 5 together comprise the "flow field" at the location of the nth monomer, due to the other monomers of the chain. Average equilibrium values of this flow field are utilized, since the focus is on the linear response of the chain. The mobility of a Gaussian and linear polyampholyte in a solution of high salt concentration was found to be a weighted average of the individual monomer mobilities, with the weight being essentially uniform and equal to $1/N$, where N is the total number of monomers. The weighting factor is actually somewhat higher for the last 10% of the monomers; e.g., it is about $1.5/N$ for the last 1% of the chain. This "end-effect" is a consequence of the greater hydrodynamic friction experienced, on average, by monomers at the end of the chain compared to those located inside the coil [11]. For sufficiently long charged sections, this effect can be neglected; however, as the number of charged monomers decreases, the small fraction of monomers that contribute with a greater weighting to the overall mobility may become important. Here we will analyze the situation where the end-effects can be neglected; a short discussion of the end-effects shall be presented in the Discussion section.

Hence, for a linear chain consisting of M_c charged monomers having a free electrophoretic mobility μ_0, and M_u uncharged monomers having no electrophoretic mobility, the overall mobility is simply an average wherein the weighting factor is approximately uniform:

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

This equation is valid so long as the two blocks retain significantly their equilibrium Gaussian conformational statistics [11] and remain hydrodynamically coupled, i.e., the two chains form a single random coil (see Fig. 1a). This condition of a single hydrodynamic unit is no longer met if the force with which the uncharged segment resists the movement via hydrodynamic drag is greater than the elastic, spring-like force which acts so as to maintain a single coiled conformation (see Fig. 1b). The drag force, i.e. the increased force of friction due to the uncharged block, depends on the electric field strength as this affects the speed which in turn determines the drag force. Hydrodynamic segregation can occur if the electric field strength and/or the hydrodynamic drag are too large. The conditions under which such decoupling may take place are typically extreme for electrophoresis; a calculation for the specific case of a DNA–streptavidin complex is presented in Ref. [10], wherein the critical electric field was indeed found to be very large ($E \approx 10$ kV/cm). Beyond the hydrodynamic segregation regime, an even greater difference between the electric force on the charged segment and the hydrodynamic drag force on the uncharged segment would result in a stretching of one or both blocks of the polyampholyte (Fig. 1c). We will not discuss these effects in this article since they are not relevant for currently available results and experimental setups.

2.2. The diffusion coefficient

We first note the classical result that the diffusion coefficient for a sphere of radius R, moving in a fluid, is given by the Stokes–Einstein equation [12]:

$$D = \frac{k_B T}{\xi} = \frac{k_B T}{6\pi\eta R} \tag{2}$$

where k_B is the Boltzmann factor, T is the absolute temperature, and η is the viscosity of the fluid. The radius which determines the coefficient of friction $\xi = 6\pi\eta R$ is not always as simple as for a hard

impermeable sphere; in the general case, the effective radius to be used is termed the hydrodynamic radius, R_H. For a Gaussian coil, the hydrodynamic radius is approximately two-thirds of the radius of gyration: $R_H \approx \frac{2}{3}R_G$ [13], where R_G can be obtained from the Kratky–Porod equation [14]. In the limit where the contour length l is much larger than the Kuhn length b_K, we obtain:

$$R_G^2 \simeq \frac{b_K l}{6} \tag{3}$$

In order to simplify the calculations, we neglect excluded volume effects (such effects become important only for rather long polymer chains [14] and would only affect some quantitative aspects of our theoretical predictions). The contour length l is equal to the product of the total number of monomers and the distance b between two monomers: $l = (M_c + M_u)b$. The diffusion coefficient of our block polyampholyte is thus given by:

$$D = \frac{D_0}{\sqrt{M_c + M_u}} \tag{4}$$

where D_0 is defined as:

$$D_0 \equiv \frac{kT}{4\pi\eta\sqrt{bb_K/6}} \tag{5}$$

Note that this result is not affected by the presence of the electric field in the case of free-flow electrophoresis.

2.3. Optimal resolution for homogeneous complexes

In a recent paper by Vreeland et al. [3], two different charged chains (single-stranded, ss, DNA) were investigated for the optimal resolution of a DNA–polymer conjugate. The smaller of the two lengths (20 and 35 bases) was found to yield better resolution, for the same experimental parameters, leading to the apparent (but premature) conclusion that the smaller the charged chain utilized, the better the resolution. This can be anticipated to some degree, since the relative difference in mobility between consecutive uncharged chain lengths (M_u and $M_u + 1$) increases for decreasing numbers of charged monomer "engines" (M_c), as can be seen from utilizing Eq. (1):

$$\frac{1}{\mu}\cdot\frac{\partial\mu}{\partial M_u} = \frac{1}{M_c + M_u} \tag{6}$$

Hence the peak separation would benefit from shorter charged chain segments. However, there are competing effects which would indicate that a longer charged block could also benefit resolution; for example, a longer charged segment would lead to higher electrophoretic mobilities, allowing for less diffusion. It would also be a larger molecule, such that the diffusion coefficient would be smaller. Its increased speed, however, would also allow it less time to separate over the length of the capillary tube. Consequently, the capillary length and electric field strength, among other experimental parameters, could be expected to have an effect on the optimum length of the charged "engine" to be used.

We shall define the resolution factor R as the ratio of the final time width of two consecutive peaks, $\sigma_t(M) \simeq \sigma_t(M+1)$, at the detector (expressed as the standard deviation of their assumed Gaussian distribution), to the difference in their elution times, $t(M)$:

$$R(M) = \frac{\sigma_t(M)}{t(M) - t(M-1)} \simeq \frac{\sigma_t(M)}{\partial t/\partial M} \tag{7}$$

where $M = M_c + M_u = l/b$ is the total number of monomers. Note that by defining the resolution factor in this manner, the lower the value the better the resolution [8] since R has the units of monomer units (i.e., $R = r$ means that we can resolve molecules that differ by r monomers). To resolve a distribution of molecular sizes, we thus need $R \leq 1$.

The final spatial width includes a component due to the initial width, $\sigma_0(M) \simeq \sigma_0(M+1)$, and a component due to the diffusion, which depends on the elution time, $t(M)$:

$$\sigma^2(M) = 2D(M)t(M) + \sigma_0^2(M) \tag{8}$$

where the diffusion coefficient $D(M)$ is given by Eq. (4). For a peak velocity of $v(M) = \mu(M)E$, where E is the electric field, the time-width is related to the spatial width via:

$$\sigma_t(M) = \frac{\sigma(M)}{v(M)} = \frac{\sqrt{2D(M)t(M) + \sigma_0^2(M)}}{v(M)} \quad (9)$$

The elution time $t(M)$ is simply the ratio of the effective length of the capillary (i.e. the distance to the detection point) L, to the velocity $v(M) = \mu(M)E$. Hence, the denominator of Eq. (7) can be written as:

$$\frac{\partial t(M)}{\partial M} = \frac{\partial}{\partial M}\left[\frac{L}{\mu(M)E}\right] \quad (10)$$

and we can rewrite the resolution factor as:

$$R = \sqrt{\frac{2D(M)}{\mu^3(M)EL[(\partial/\partial M)(1/\mu)]^2} + \left[\frac{\sigma_0(M)}{\mu(M)L(\partial/\partial M)(1/\mu)}\right]^2} \quad (11)$$

Since we are interested in the optimal resolution for chains with a set number of charged monomers, M_c, and differing uncharged chain lengths, M_u, we take the partial derivative of the mobility (Eq. (1)) with respect to M_u (i.e., $\partial/\partial M = \partial/\partial M_u$). Thus, using the equations for the diffusion coefficient and mobility (Eqs. (1) and (4)), the resolution becomes, for $M = M_c + M_u$:

$$R = \sqrt{\frac{2D_0}{\mu_0 EL} \cdot \frac{(M_c + M_u)^{5/2}}{M_c} + \frac{\sigma_0^2}{L^2}(M_c + M_u)^2} \quad (12)$$

The optimal length of the charged section can now be obtained by solving the optimization condition $(\partial R/\partial M_c) = 0$. This yields the following fourth-order polynomial in M_c:

$$\Sigma_0^2 M_c^4 = (\tfrac{5}{2})^2 \cdot (M_c + M_u)M_c^2 - 5(M_c + M_u)^2 M_c + (M_c + M_u)^3 \quad (13)$$

where the only term representing the relevant experimental parameters is:

$$\Sigma_0 = \frac{\sigma_0^2 \mu_0 E}{LD_0} \quad (14)$$

The exact solutions of this fourth-order polynomial are very long and hence are not given here; instead we present series solutions for the two limiting cases. Σ_0 is clearly a term representing the two factors which affect the final peak width: the initial loading width ($\sim\sigma_0$), and the diffusion ($\sim D_0$). In fact, it is proportional to the critical length, $L_0 \equiv \sigma_0^2 \mu E/2D$, at which these two terms are equal [15]:

$$\Sigma_0 = \frac{L_0}{L} \cdot \frac{\sqrt{M_c + M_u}}{M_c} \quad (15)$$

For typical experimental parameters (see Appendix A), Σ_0 may range from about 10^{-3} to 10^{+4}. In order to find a manageable solution for the optimal length of the charged segment, we will use series solutions for these two limits.

2.4. Optimal number of charged monomers for the diffusion-limited regime

First we note that in the limit of $\Sigma_0 = 0$ (i.e., negligible loading widths), Eq. (13) yields:

$$M_c = \tfrac{2}{3} \cdot M_u \quad (16)$$

Hence, in this limit, the optimal number of charged monomers for the "engine" is not $M_c \to 1$, unlike the condition suggested by the data presented in Ref. [3]. Also of interest is that the optimal value of $\frac{2}{3}M_u$ does not depend on any system parameters such as the length of the capillary. A series solution around this value for M_c yields:

$$\frac{M_c}{M_u} \simeq \frac{2}{3} - \frac{8\sqrt{15}}{135} \cdot (\Sigma_0^2 M_u)^{1/2} + \frac{32}{225} \cdot (\Sigma_0^2 M_u) + \cdots \quad (17)$$

This series agrees quite well with the full solution of Eq. (13), for the range of $0 \lesssim \Sigma_0^2 M_u \lesssim 1$, as can be seen in Fig. 2. The optimal value of M_c decreases smoothly from $\frac{2}{3}M_u$ when Σ_0 increases.

2.5. Optimal number of charged monomers for the injection-limited regime

A series solution for the region of large Σ_0 yields:

$$\frac{M_c}{M_u} \simeq \frac{1}{(\Sigma_0^2 M_u)^{1/4}} - \frac{1}{2(\Sigma_0^2 M_u)^{1/2}} + \cdots \quad (18)$$

Fig. 2 also displays this series solution. In contrast with the diffusion-limited regime, the optimal num-

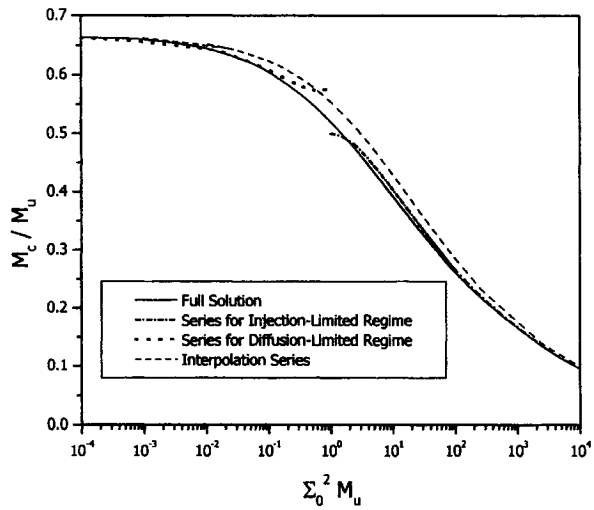


Fig. 2. Ratio of the optimal engine size M_c to the size of the uncharged block M_u, vs. the expansion parameter $\Sigma_0^2 M_u$ which measures the impact of the injection width.

ber of charged monomers can now be quite small compared to M_u (of course, values smaller than unity are meaningless).

2.6. Discussion of FSCE for homogeneous block polyampholytes

Of course, one normally has a distribution of molecular sizes M_u to be analyzed. Therefore, what our analysis really suggests is that one should choose the engine size M_c to maximize the resolution near the peak of this distribution. However, since the resolution factor R increases with size M_u, it might be preferable to choose the engine size M_c to maximize the resolution for the largest polymers to be resolved (since the latter are the most difficult to resolve anyway). Let us call M_u^* this molecular size. The behaviour of the optimal M_c over the full range of Σ_0 is then approximated very well (see Fig. 2) by the following interpolating function:

$$M_c \approx \frac{\frac{2}{3}M_u^*}{\sqrt{1+\frac{4}{9}(\Sigma_0^2 M_u^*)^{1/2}}} \tag{19}$$

Perhaps the most interesting result is that the optimal engine size M_c for the separation of homogeneous molecules can be as large as $\frac{2}{3}\cdot M_u^*$. This nontrivial result comes from the competition between diffusion and peak spacing. When the impact of the injection width is larger, as described by the parameter $\Sigma_0 \sim \sigma_0$, smaller engines are needed because more time (slower separations) is necessary in order to overcome the initial peak widths.

The optimal resolution (Eq. (12)) for the largest molecule to be resolved, in the limit of negligible loading width, is given by:

$$R(M_u^*)|_{M_c=2/3M_u^*} = \sqrt{\frac{D_0}{\mu_0 EL}\cdot\frac{5^{5/2}}{3^{3/2}}\cdot M_u^{*3/2}} \tag{20}$$

Using the numerical values given in Appendix A, this can be rewritten as $R(M_u^*) \simeq (M_u^*/1300)^{3/4}$. This suggests that molecules up to $\simeq 1300$ monomers long could be resolved under these ideal conditions. Higher electric fields and longer capillaries can be used to increase this number. The corresponding elution time would then be (using Eq. (1)):

$$t(M_u^*) = \frac{L}{\mu(M_u^*)E} = \frac{5}{2}\cdot\frac{L}{\mu_0 E} \tag{21}$$

Remarkably, the elution time of the M_u molecule is a universal multiple of the time $L/\mu_0 E$, the elution time of an unlabeled charged molecule. The latter, being less than 20 min for most experimental conditions, indicates that FSCE is a fairly fast separation process.

For the injection-limited regime where the optimal M_c approaches 1, however, the optimal resolution would be reduced to:

$$R(M_u^*)|_{M_c=1} = \frac{\sigma_0}{L}\cdot(M_u^*+1) \tag{22}$$

The corresponding elution time would be (using Eq. (1)):

$$t(M_u^*)|_{M_c=1} = (M_u^*+1)\cdot\frac{L}{\mu_0 E} \tag{23}$$

Although one can achieve fractionation of the sample using this limit, it is clearly not as good as the diffusion-limited situation discussed above. In particular, the time duration of the separation can be enormous.

In the next section, we generalize our analysis to treat inhomogeneous block copolymers made of blocks with different hydrodynamic properties.

3. Non-homogeneous charged–uncharged complexes

The equation for the mobility of a block polyampholyte, developed by Long et al. [11], assumes a homogeneous backbone, that is, charged and uncharged blocks having the same zero-field hydrodynamic properties; hence it will need to be adjusted for the case of non-homogeneous backbones. For example, Vreeland et al. [3] studied ssDNA–PEG conjugates. In order for their equation to be utilized, we visualize a regrouping of the monomers of both charged and uncharged chain segments into "supermonomer" units (called blobs) of uniform hydrodynamic radii (see Fig. 3), so as to "homogenize" the complex [10]. For example, three uncharged monomers could comprise a hydrodynamic blob equivalent to a blob made of five charged monomers. Hence the number of blobs, M_{B_i}, of type i (i = c or u, for charged or uncharged, respectively) can be expressed as the total number of Kuhn lengths of type i, M_{K_i}, divided by the number of Kuhn lengths in each homogenized unit of type i, m_{K_i}:

$$M_{B_i} = \frac{M_{K_i}}{m_{K_i}} \tag{24}$$

with

$$M_{K_i} = \frac{l_i}{b_{K_i}} = \frac{M_i b_i}{b_{K_i}} \tag{25}$$

After redefining the blob-monomers, we can use Eq. (1) for the mobility of the complex made of M_{B_c} charged blobs (of mobility μ_{0_B}) and M_{B_u} uncharged blobs since all blobs are hydrodynamically equivalent. Using Eqs. (24) and (25), we then obtain:

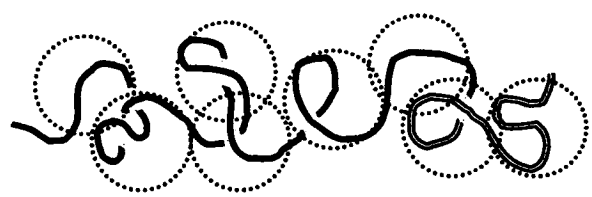


Fig. 3. Schematic representation of the blob construction for inhomogeneous backbones. We chose an extended molecular conformation for clarity.

$$\mu = \mu_{0_B} \cdot \frac{M_{B_c}}{M_{B_c} + M_{B_u}} = \mu_{0_B} \cdot \frac{M_c}{M_c + M_u \cdot (m_{K_c} b_u b_{K_c} / m_{K_u} b_c b_{K_u})} \tag{26}$$

This equation can be further simplified by using the necessary equivalence of the hydrodynamic radii r_H of the charged and uncharged blobs:

$$r_{H_c} = r_{H_u} \tag{27}$$

From Section 2.2, we know that the hydrodynamic radius of a polymer chain of contour length l_i can be expressed in terms of its Kuhn length b_{K_i} as $r_{H_i} \simeq \frac{2}{3}(b_{K_i} l_i / 6)^{1/2}$. The total contour length within a blob of hydrodynamic radius r_H being given by $l_i = m_{K_i} b_{K_i}$, we obtain $r_{H_i} \simeq \frac{2}{3}(b_{K_i}^2 m_{K_i}/6)^{1/2}$. Therefore, Eq. (27) leads to the important result:

$$\frac{m_{K_c}}{m_{K_u}} = \frac{b_{K_u}^2}{b_{K_c}^2} \tag{28}$$

and Eq. (26) can be expressed as:

$$\mu = \mu_{0_B} \cdot \frac{M_c}{M_c + M_u \alpha} \tag{29}$$

where we now introduce the microscopic parameter:

$$\alpha \equiv \frac{b_u b_{K_u}}{b_c b_{K_c}} \tag{30}$$

It is interesting to note that this is the only parameter that contains information about the chemistry of the two polymers. Moreover, this information is only relative since we find a ratio between various microscopic length scales (monomer sizes and Kuhn lengths).

Noting that the hydrodynamic radius of the block polyampholyte is $R_H = (R_{H_u}^2 + R_{H_c}^2)^{1/2}$, the diffusion coefficient from (Eq. (2)) becomes:

$$D = \frac{k_B T}{6\pi\eta\sqrt{R_{H_c}^2 + R_{H_u}^2}} = \frac{D_{0_c}}{\sqrt{M_c + \alpha M_u}} \tag{31}$$

where we used the fact that:

$$R_{B_i} \simeq \tfrac{2}{3} R_{G_i} = \tfrac{2}{3}\sqrt{b_i b_{K_i} M_i / 6}$$

and the definition:

$$D_{0_c} \equiv \frac{k_B T}{(4/\sqrt{6})\pi\eta\sqrt{b_c b_{K_c}}}$$

Utilizing the mobility and diffusion coefficients from Eqs. (29) and (31), Eq. (11) becomes, for non-homogeneous backbones:

$$R = \sqrt{\frac{2D_{0_c}}{\mu_{0_B} EL} \cdot \frac{(M_c + M_u\alpha)^{5/2}}{M_c\alpha^2} + \frac{\sigma_0^2}{L^2\alpha^2}(M_c + M_u\alpha)^2} \tag{32}$$

Taking the derivative of R with respect to M_c, and again setting it equal to zero, yields a fourth-order polynomial similar to that for homogeneous backbones (Eq. (13)):

$$\Sigma_{0_c}^2 M_c^4 = (\tfrac{5}{2})^2 \cdot (M_c + M_u\alpha)M_c^2 - 5(M_c + M_u\alpha)^2 M_c + (M_c + M_u\alpha)^3 \tag{33}$$

where $\Sigma_{0_c} \equiv \sigma_0^2 \mu_{0_B} E/LD_{0_c}$. Again the solution is similar to the homogeneous case, being closely approximated by:

$$\frac{M_c}{\alpha M_u} \approx \frac{2/3}{\sqrt{1 + (4/9)(\alpha M_u \Sigma_{0_c}^2)^{1/2}}} \tag{34}$$

Note that this solution is indeed equal to the corresponding equation (Eq. (19)) for the homogeneous condition $\alpha = 1$. As we can see, the microscopic parameter α simply rescales M_u: each charged monomer corresponds hydrodynamically to α uncharged monomers. This is the result of the fact that the monomers are no longer hydrodynamically equivalent.

Using Eq. (32), we can calculate the optimal resolution for the largest non-homogeneous block copolymer, which we again take to be of size M_u^*. In the limit of negligible loading widths, we obtain:

$$R(M_u^*)|_{M_c=(2\alpha/3)M_u^*} = \frac{1}{\alpha^{1/4}} \cdot \sqrt{\frac{D_{0_c}}{\mu_{0_B} EL} \frac{5^{5/2}}{3^{3/2}} \cdot M_u^{*3/2}} \tag{35}$$

and the corresponding elution time would be:

$$t(M_u^*) = \frac{L}{\mu(M_u^*)E} = \frac{5}{2} \cdot \frac{L}{\mu_{0_B} E} \tag{36}$$

which, remarkably, is the same as for the homogeneous case (Eq. (21)), and does not depend on the number of monomers or the nature of the polymers. Fig. 1 of Vreeland et al. [3] gives a factor of about 2.1 instead of 5/2, but their engine was not of optimal size. Using the numerical values described in Appendix A, we obtain $R(M_u^*) \simeq (M_u^*/680)^{3/4}$. For PEG, as used by Vreeland et al. [3], this predicts that the maximum molecular mass that we could resolve using FSCE, an optimal engine and these experimental conditions would be about 30 000 Daltons, close to what is suggested by these authors' results.

For the injection-limited regime where the optimal value of M_c approaches 1, the optimal resolution would be:

$$R(M_u^*)|_{M_c=1} = \frac{\sigma_0}{L} \cdot \left(M_u^* + \frac{1}{\alpha}\right) \tag{37}$$

while the corresponding elution time would be (using Eq. (29)):

$$t(M_u^*)|_{M_c=1} = (M_u^*\alpha + 1) \cdot \frac{L}{\mu_{0_B} E} \tag{38}$$

From Eqs. (35) and (37), it is evident that values of $\alpha < 1$ increase the resolution factor R, while $\alpha > 1$ results in smaller R's. Of course, the latter is preferable. Since $\alpha = b_u b_{K_u}/b_c b_{K_c}$, this indicates that one should choose the charged block such that it has a smaller monomer size b_c and/or a smaller Kuhn length b_{K_c} than the uncharged polymer sample that one wishes to separate. This is not trivial since most charged polymers tend to be rigid because of the extra electrostatic contribution to their Kuhn length.

4. Discussion

Free-solution conjugate electrophoresis has been shown to be a potential alternative to gel permeation chromatography and mass spectrometry for the analysis of the mass distribution of polydisperse polymer samples [3]. In this article, we have derived the basic theoretical framework for the development and optimization of FSCE. Three main elements were

needed: (1) a model for the electrophoretic mobility $\mu(M_u)$ of a linear polymer of degree of polymerization M_u end-conjugated to a polymeric charged engine of size M_c; (2) a model for the thermal diffusion coefficient $D(M_u,M_c)$ of these molecules; (3) a model to represent the effective friction coefficient of the two polymer blocks during the electrophoresis process. We demonstrated that the recent polyampholyte theory of Long et al. [11] could be generalized to treat this problem. One of the main results is that the effective frictional contribution of the polymer chains (described by the constant α) is proportional to the microscopic parameters (monomer size and chain stiffness) of both polymers. This readily suggests a new method to estimate the Kuhn length of uncharged polymers as a function of the Kuhn length of, for example, ssDNA under the same conditions. Our blob theory also explains why the FSCE electropherogram shows equally-spaced peaks (see Eqs. (1) and (29)), a most useful feature of this separation method [3].

It is interesting to use the recent results of ssDNA–PEG FSCE separations to test our theoretical model for the parameter α. Vreeland et al. reported that $\alpha = 0.138$ for their experimental conditions [3]. Using Eq. (30) and the fact that $b_{K_{ssDNA}} \approx 7$ nm and $b_{ssDNA} = 0.43$ nm [10], we find that $b_{K_{PEG}} \cdot b_{PEG} \approx 0.42$ nm^2. Since $b_{PEG} = 0.36$ nm, we get $b_{K_{PEG}} \approx 1.15$ nm (or ≈ 3 monomers), and a characteristic ratio $C_\infty = 3.2$, in fair agreement with the experimental value of $C_\infty = 3.8$. Incidentally, their value of $\alpha = 0.138$, together with Eq. (34), suggests that optimal results would be obtained, for the molecular mass 5000 PEG samples ($M_u = 114$), using a ssDNA label of size $M_c \leq (2\alpha/3)M_u = 10$ bases (the exact value depends on the importance of the injection width), while they used 20 and 35 base ssDNA engines.

Moreover, we determined how the resolution provided by FSCE depends on the actual size of the engine. For a diffusion-limited situation (always the best case scenario in separation science), there is actually an optimal engine size for a given polymer mass distribution. A general theory was also derived for a more general case where both the diffusion and the injection width limit the final resolution. Our analysis of these theoretical predictions indicates that much could indeed be gained by optimizing the engine size M_c, increasing the microscopic parameter α (e.g., by changing the ionic strength of the buffer) and decreasing the experimental parameter Σ_0 (e.g., by increasing the field strength). Our theory thus introduces a systematic way to optimize FSCE.

Our model has neglected a number of effects that can modify its quantitative predictions to some extent, although none would modify its qualitative predictions. First, it must be noted that the mobility of the last charged blob, μ_{0_B}, can be slightly size dependent if the blob size is not large enough (e.g., the free-solution mobility of DNA is slightly size dependent below about 20 bases). This may introduce a small correction to our equations when the predicted optimal engine size is small. As we mentioned before, the theory of Long et al. [11] also introduces a small correction factor for both ends of the polymer. We also neglected excluded volume effects because they are often negligible for short polymers [14]; however, such effects can easily be added to the model. Finally, we must stress the fact that the version of the theory presented here assumed that the charged block was long enough to be represented by a Gaussian blob (i.e., we assumed that $M_c \gg b_{K_c}/b_c$); it is rather straightforward to modify the theory if the charged block is very rigid (e.g., one then has $r_{H_c} \sim M_c$).

Since FSCE, like ELFSE, is based on the hydrodynamic properties of macromolecules, anything that could alter the hydrodynamics of the analyte could potentially affect (negatively or positively) the performance of this new method. For example, very high field intensities could potentially lead to polymer deformation: this would greatly modify the effective frictional properties of both the engine and the polymer drag [10]. As discussed previously, this is not expected to happen under most circumstances. Of great interest would be the effect of branching and other nontrivial polymer structures. Indeed, the hydrodynamic properties of a non-linear macromolecule depend not only on the mass of the latter, but also on its precise configuration. We thus predict that FSCE could potentially be used to separate and characterize, for example, polymers of identical masses but different degrees of branching. The theory presented in this paper can be generalized to treat such cases.

Acknowledgements

The authors would like to thank Drs. Frank Oaks, Wendy Petka and Didier Long for useful discussions. This work was supported, in part, by a Natural Science and Engineering Research Council (NSERC) of Canada scholarship to L.C.M., and a NSERC Research Grant to G.W.S. W.N.V. is supported by Northwestern University's NIH Predoctoral Biotechnology Training Grant (5-T32 GM 08449-06). A.E.B and W.N.V. acknowledge support from the US Department of Energy, Office of Biological and Environmental Research Grant DE-FG02-99ER62789. Further financial support came from Applied Biosystems.

Appendix A

In order to get some rough numerical predictions, we used the following "typical" ranges for the CE experimental parameters: $T \approx 300$–340 K, $\eta \approx 400$–1000 μPa s, $b_{K_{ssDNA}} \approx 5$–10 nm, $\sigma_0 = 0.1$–10 mm, $E \approx 100$–1000 V/cm, $\mu_0 \approx 1 \cdot 10^{-4}$–$5 \cdot 10^{-4}$ cm^2/V s, $L \approx 10$–100 cm. Finally, the ssDNA monomer size is $b_{ssDNA} \approx 0.43$ nm. Using these ranges, we find that Σ_0 may range from about 10^{-3} to 10^{+4}.

In addition, we used the following parameters to make predictions using Eqs. (20), (35) and (36): $D_0 = 7.0 \cdot 10^{-6}$ cm^2/s, $\mu_0 = 3 \cdot 10^{-4}$ cm^2/V s, $E =$ 300 V/cm, $L = 40$ cm. We also used $\alpha \approx 0.138$, as determined by Vreeland et al. [3] for PEG.

References

[1] H.G. Barth, B.E. Boyes, C. Jackson, Anal. Chem. 70 (1998) 251R.
[2] A. Marie, F. Fournier, J.C. Tabet, Anal. Chem. 72 (2000) 5106.
[3] W.N. Vreeland, C. Desruisseaux, A.E. Karger, G. Drouin, G.W. Slater, A.E. Barron, Anal. Chem. 73 (2001) 1795.
[4] M. Stefansson, M. Novotny, Anal. Chem. 66 (1994) 1134.
[5] J. Bullock, J. Chromatogr. 645 (1993) 169.
[6] P. Mayer, G.W. Slater, G. Drouin, Anal. Chem. 66 (1994) 1777.
[7] C. Heller, G.W. Slater, P. Mayer, N. Dovichi, D. Pinto, J.-L. Viovy, G. Drouin, J. Chromatogr. A 806 (1998) 113.
[8] H. Ren, A.E. Karger, F. Oaks, S. Menchen, G.W. Slater, G. Drouin, Electrophoresis 20 (1999) 2501.
[9] P.G. Righetti (Ed.), Capillary Electrophoresis in Analytical Biotechnology, CR Press, New York, 1996.
[10] C. Desruisseaux, D. Long, G. Drouin, G.W. Slater, Macromolecules 34 (2001) 44.
[11] D. Long, A.V. Dobrynin, M. Rubinstein, A.J. Ajdari, J. Chem. Phys. 108 (1998) 1234.
[12] P. Munk, Introduction to Macromolecular Science, Wiley, New York, 1989.
[13] J. des Cloizeaux, G. Jannink, Les Polymères en Solution: leur Modélisation et leur Structure, Les Éditions de Physique, Paris, 1987.
[14] A.Y. Grosberg, A.R. Khokhlov, Statistical Physics of Macromolecules, AIP Press, New York, 1994.
[15] L. Bousse, C. Cohen, T. Nikiforov, A. Chow, A.R. Kopf-Sill, R. Dubrow, J.W. Parce, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 155.

6

Diffusion Coefficient of DNA Molecules During Free Solution Electrophoresis

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

This seminal paper showed that the Nernst-Einstein equation relating the mobility and the diffusion coefficient during electrophoresis, while appropriate for most gel-based systems, is invalid for free solution electrophoresis. Many authors simply assumed that this equation still holds under free solution conditions, however our results unequivocally show that a simplistic use of the Nernst-Einstein relation fails and that the electric field actually has no effect on the thermal diffusion process. Since the diffusion coefficient is not affected by the electric field in free solution, the correct equation for the diffusion coefficient is given by Zimm's equation for non-electrophoretic conditions. This is a crucial result, impacting on predictions of ELFSE performance.

Claude Desruisseaux was responsible for the ssDNA experimental data in this paper, while I assisted my supervisor with the related analysis and writing. The dsDNA sections were the responsibility of Bernard Tinland and coworkers.

2424 *Electrophoresis* 2001, *22*, 2424–2432

Axel E. Nkodo[1]
Jean M. Garnier[2]
Bernard Tinland[1]
Hongji Ren[3]
Claude Desruisseaux[3]
Laurette C. McCormick[4]
Guy Drouin[3]
Gary W. Slater[4]

[1]Institut Charles Sadron, Strasbourg, France
[2]IGBMC, Strasbourg, France
[3]Department of Biology
[4]Department of Physics, University of Ottawa, Ottawa, Canada

Diffusion coefficient of DNA molecules during free solution electrophoresis

The free-draining properties of DNA normally make it impossible to separate nucleic acids by free-flow electrophoresis. However, little is known, either theoretically or experimentally, about the diffusion coefficient of DNA molecules during free-flow electrophoresis. In fact, many authors simply assume that the Nernst-Einstein relation between the mobility and the diffusion coefficient still holds under such conditions. In this paper, we present an experimental study of the diffusion coefficient of both ssDNA and dsDNA molecules during free-flow electrophoresis. Our results unequivocally show that a simplistic use of Nernst-Einstein's relation fails, and that the electric field actually has no effect on the thermal diffusion process. Finally, we compare the dependence of the diffusion coefficient upon DNA molecular size to results obtained previously by other groups and to Zimm's theory.

Keywords: Capillary electrophoresis / DNA sequencing / Band broadening / Diffusion coefficient / Free-draining DNA EL 4546

1 Introduction

Although electrophoretic methods have been used extensively to separate nucleic acids over the last several decades, there are still numerous unresolved theoretical issues that have important consequences for the optimization of new and current technologies [1]. A remarkable example of the recent evolution of our theoretical understanding is provided by the work of Long *et al.* [2–5] on the electrophoretic properties of composite macromolecules (*e.g.*, DNA molecules end-labeled with streptavidin) and polyelectrolytes with inhomogeneous charge distributions (*e.g.*, block copolymers or polyampholytes). These new developments have redefined the concept of free-draining polymers and have made clear and rather surprising predictions regarding situations where both electrophoretic and mechanical forces (*e.g.*, extra frictional forces coming from an uncharged component attached to the polyelectrolyte) are applied to a flexible polyelectrolyte. We have recently demonstrated the validity of some of these nontrivial predictions [6]. Another example is the recent experimental work on diffusional band broadening during gel electrophoresis: Tinland *et al.*'s results [7–11], for instance, fully support the theoretical prediction that diffusion is increased by the field-driven motion of the DNA chain in its tube [12–14], a nontrivial phenomenon that actually reduces the contribution of Joule heating during gel electrophoresis [15]. Neglecting this field-induced increased diffusion may thus lead to completely erroneous conclusions about the mechanisms that limit the performance of DNA separation systems (*e.g.*, capillary sequencing systems) based on gel sieving.

Capillary electrophoresis (CE) is now the standard method to sequence DNA. The CE methods currently rely mostly on the sieving properties of highly entangled polymer solutions for the separation of ssDNA molecules of approximately 30–1000 bases [16–17]. The maximum read length is limited by the combination of reduced peak spacing and broader peaks found for the largest and slowest molecules. The reduced peak spacing is well understood in terms of the field-driven molecular orientation predicted by the biased reptation model [18] and observed by videomicroscopy [19] and other techniques. Detailed experimental studies have provided us with an excellent empirical body of quantitative information about the relation between mobility (μ, in cm^2/Vs), polymer concentration, field intensity (E, in V/cm) and DNA size (M, in bases or base pairs) [18]. However, the band broadening processes are not yet fully understood because the dynamics of the uncross-linked polymers clearly affect DNA diffusion, a contribution that the reptation model does not take into account. Given the current importance of DNA CE sequencing in polymer solutions for the different genome projects, it is odd that so little is known about these coupled diffusion processes.

Surprisingly enough, the diffusion coefficient $D(M)$ to be used to analyze the data from experiments involving the free-flow electrophoresis of DNA molecules also remains ill-understood and the relevant literature is often somewhat confusing. For example, it is often not realized that

Correspondence: Professor Gary W. Slater, Department of Physics, University of Ottawa, 150 Louis-Pasteur, Ottawa, Ontario K1N 6N5, Canada
E-mail: gslater@science.uottawa.ca
Fax: +613-562-5190

Abbreviations: FRAP, fluorescence recovery after photobleaching; **pDMA**, polydimethylacrylamide; **TBE**, Tris-borate-EDTA

the Nernst-Einstein relation [20] between the mobility $\mu(M)$ and the diffusion coefficient $D(M)$ of a DNA molecule of size M (in bases),

$$\frac{D}{\mu} = \frac{k_B T}{Q} \quad (1)$$

where $Q \sim M$ is the charge of the polyelectrolyte, should not be valid during free flow electrophoresis. However, it is known to be valid during gel electrophoresis, in the limit of low field intensities [12–14, 18], because the gel fibers then essentially screen the hydrodynamic interactions over distances larger than the mean pore size (note that it is not valid in gels at high fields because of the nonlinear effects discussed previously [12–14, 18]).

In fact, the diffusion coefficient $D(M)$ of DNA fragments should be independent of field intensity E in (capillary) free-flow electrophoresis. As far as we can tell, this simple fact has never been tested before. In other words, we expect the diffusion coefficient to be given by Zimm's expression for non-electrophoretic conditions [21]:

$$D(M) = \frac{k_B T}{6\pi\eta R_H(M)} \quad (2)$$

where $R_H(M)$ is the hydrodynamic radius of the DNA coil, k_B is Boltzmann's constant, T is the temperature (in K), and η is the buffer viscosity. For a long flexible polymer, we should thus have $D(M) \sim 1/M^{3/5}$ if excluded volume interactions are important and $D(M) \sim 1/M^{1/2}$ otherwise [21]. On the other hand, the mobility μ is not a function of molecular size M for so-called free-draining polymers such as DNA (*i.e.*, $\mu \sim M^0$), although we sometimes have a very weak dependence for small molecular sizes (for example, for sizes less than about 400 bp for dsDNA) [17, 18]. Taken together, these scaling laws clearly contradict Eq. (1). Therefore, either the Nernst-Einstein relation is not valid for DNA free-solution CE, or the Zimm diffusion coefficient is not valid during such experiments. As we will see, our results clearly indicate that the former is the correct conclusion.

The main goal of this paper is thus to test the validity of Eq. (1) by measuring the diffusion coefficient $D(M)$ of both ssDNA and dsDNA fragments with and without an electric field being applied. Furthermore, we will study the size-dependence of the diffusion coefficient $D(M)$ during free-flow electrophoresis. The experimental and theoretical details of our study are described in Section 2 while the results are presented in Section 3. We conclude (Section 4) by looking at the consequences of these findings for the development of various DNA separation methods.

2 Materials and methods

2.1 ssDNA experiments

2.1.1 The electrophoretic conditions

All experiments were performed with the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) utilizing user-modified modules for free-solution electrophoresis. The buffer filling time between each run was 1.5 min, and the data were collected at frequencies of 1.6 or 3.1 Hz. Uncoated capillaries with 20 μm ID and 60 cm total length (49 cm effective length) were used. The buffer contained 1 × (TAPS) (*i.e.*, 0.1 M TAPS; the pH was adjusted to 8.5 with NaOH), 0.1% POP4 polymer (Applied Biosystems) and 7 M N-tris (hydroxy methyl) methyl-3-aminopropanesulfonic acid urea. The small amount of POP polymer was used to reduce the EOF of the uncoated capillaries to negligible values. All experiments were done at 30°C.

2.1.2 DNA samples

DNA fragments 50, 75, 100, 160, 200 and 250 bp long were obtained by separating GeneScan Rox labeled DNA standards (Applied Biosystems) on an agarose gel, cutting the bands out, and extracting the DNA fragments from the gel using a GeneClean kit. These samples were further purified using CentriSep spin columns (Princeton Separations, Adelphia, NJ, USA). The identity of each of the DNA fragments was confirmed by capillary electrophoresis using a denaturing POP4 buffer (sieving matrix) and the GeneScan 500 Tamra DNA standards (Applied Biosystems). The 18-mer DNA fragment was from one of the Applied Biosystems dye primer sequencing kit. A mixture of 2 μL of purified DNA fragments, 6 μL of water, and 8 μL of 8M urea was heated at 65°C for 2 min, quenched on ice, and spun briefly (to pellet the impurities) before loading on the CE instrument.

2.1.3 Measuring the diffusion coefficient D

Two different methods were employed to measure $D(M)$. The first one is the stop-flow method [22], where the voltage was turned off after the analyte had migrated nearly halfway along the capillary. The molecule was then left in the capillary for different time durations; this allowed zero-field diffusional band broadening to take place. Finally, the voltage was turned on again to drive the analyte towards the detector window. The diffusion coefficient at zero-field, $D_0(M)$, is then obtained from a fit of the final band variance $\sigma(t)$ *vs.* the stop-time t:

$$\sigma^2(t) = \sigma_e^2 + 2D_0(M)t \quad (3)$$

where σ_e is the peak standard-deviation due to the electrophoretic migration itself (including the contribution σ_0 coming from the initial band width at injection). The chosen stop times were long enough (0.5–20 h) for the last term to clearly dominate the final band width. Such a fit thus gives both the diffusion coefficient D_0 and an approximate value for σ_e. In all cases, we also loaded a second sample prior to starting the second phase of the electrophoresis experiment in order to obtain a separate measure of the electrophoretic band width σ_e. Fits of the residual variance σ^2-$(\sigma_e)^2$ *vs.* stop-time t gave zero-field diffusion coefficients D_0 that were essentially identical to those obtained without measuring σ_e. However, because of the variations of the EOF with time, the diffusion coefficients D_0 obtained using σ^2–$(\sigma_e)^2$ usually had smaller error bars (the double loading process removes these fluctuations from the final result). All stop-flow experiments were done with a field E = 15kV/60cm = 250 V/cm.

Uninterrupted experiments were carried out to obtain the diffusion coefficient $D(M)$ in the presence of the field E. A simple way to measure $D(M)$ is then to calculate the plate height $H = (\sigma_e)^2/L$, where L is the elution length. If Joule heating is negligible, H is given by :

$$H = \frac{\sigma_e^2}{L} = \frac{\sigma_0^2}{L} + \frac{2D}{u} + Wu \qquad (4)$$

where W is a constant and $u = \mu E$ is the electrophoretic velocity. The three terms are due, respectively, to (i) the initial injection width σ_0, (ii) thermal (Brownian) diffusion, and (iii) the analyte-wall interactions. If D is not a function of the field intensity E (our main hypothesis here), a plot of H *vs.* $2/u$ should show a linear regime at low velocity u with a slope $D = D_0$. Note that if D does depend on E, Eq. (4) will not be adequate to fit the data; therefore, this approach is also a direct test for the hypothesis that $D \neq D(E)$, as we shall see. Note that in practice it is easier to measure the full width at half maximum (FWHM) of a peak than its variance σ^2. For a Gaussian peak (an excellent approximation for the situation studied here), the relation between the two is simply $(\text{FWHM})^2 = 8\ln(2) \times \sigma^2 \cong 5.545 \times \sigma^2$.

2.1.4 The Zimm and Kratky-Porod theories

In the framework of the Rouse model [21], the diffusion coefficient of a random coil polymer chain scales like $D \sim 1/M$. Incidentally, Eq. (1) does predict this scaling law for DNA in the presence of an electric field (since $Q \sim M^1$ and $\mu \sim M^0$). However, the hydrodynamic interactions between the monomers make the coil impermeable to the solvent; in this case, Zimm's theory predicts that the diffusion coefficient is given by Eq. (2) instead. If excluded volume interactions are negligible, we have $R_H \sim M^{1/2}$ (if the contour length $L_D \sim M$ of the DNA molecule is much longer than its persistence length p) and Eq. 2 then predicts $D \sim 1/M^{1/2}$; this was the result found by Tinland for ssDNA in Tris-borate-EDTA (TBE) buffer with 8M urea [11]. If excluded volume interactions cannot be neglected, however, the scaling becomes $R_H \sim M^{3/5}$ for long enough polymers and Eq. (2) then predicts $D \sim 1/M^{3/5}$. For short, rod-like polymers, on the other hand, we expect roughly $R_H \sim M$ and $D \sim 1/M$, as we will discuss later.

We will also try to analyze our ssDNA data using the Kratky-Porod equation [23] for the radius-of-gyration $R_g(L_D)$ of a worm-like chain of contour length $L_D = Mb$ (where b is the contour length of one monomer; note that b = 0.43 nm for ssDNA and b = 0.34 nm for dsDNA) [23]:

$$R_g^2(M) = \frac{pL_D}{3} \times \left[1 - 3\left(\frac{p}{L_D}\right) + 6\left(\frac{p}{L_D}\right)^2 - 6\left(\frac{p}{L_D}\right)^3 (1 - e^{-L_D/p})\right] \qquad (5)$$

In order to do so, we will also assume that the relation $R_H \cong (2/3)R_g$, which holds for a Gaussian coil [24, 25], can also be used to relate the hydrodynamic radius R_H and the radius-of-gyration R_g of our Kratky-Porod polymer chains. The Kratky-Porod equation allows us to study the intermediate regime where the chain is between the rigid rod ($R_H \sim M$) and random coil ($R_H \sim M^{1/2}$) regimes and excluded volume can be neglected.

2.2 dsDNA experiments

2.2.1 Materials

For all experiments the buffer was 0.01 M TBE at pH 8.3 at 23°C. The λ phage DNA was purchased from Biolabs (Beverly, MA, USA) and the T2 DNA from Sigma (St. Louis, MO, USA). All the other fragments (ranging from 2.1 to 19.1 kbp) were restricted and linearized from the corresponding vectors. The samples were checked by performing a gel electrophoresis and preheated at 55°C before injection in the capillary in order to open the circular phage. All sample solutions were diluted well below their overlap concentration and mixed with a solution of YOYO (Molecular Probes, Eugene, OR, USA) in the ratio of 1 dye per 40 bp. The pDMA (polydimethyl-acrylamide) used for coating the capillary walls had a molecular weight of 156 000 g/mol. It was a gift from P. Galin (ICS, Strasbourg). After the coating procedure, fluoresceinated dextran ($M_w = 2.10^6$ g/mol) from Sigma was used to estimate the electroosmotic flow (EOF).

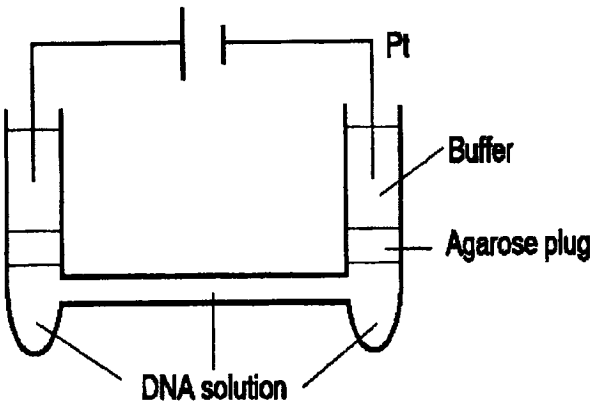


Figure 1. Schematic diagram of the electrophoretic cell. The agarose plugs separate the DNA solution from the electrode regions and thus minimize convective effects.

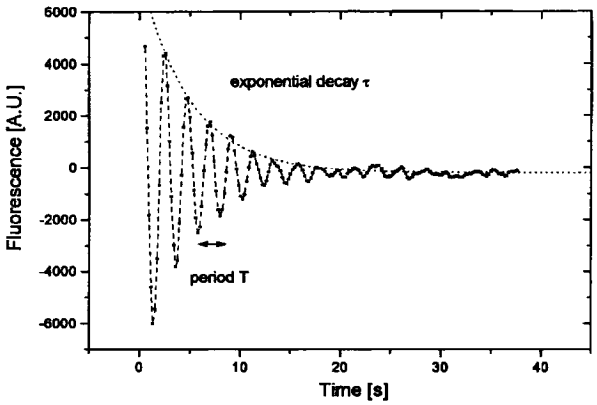


Figure 2. A typical FRAP signal for DNA electrophoresis. The time of the oscillations corresponds to the drift of DNA chains, so to the mobility. The time of the exponential decay corresponds to the diffusion.

2.2.2 Capillary electrophoresis

Measurements were carried out in fused silica, square-section capillaries (Wales Apparatus Co., Hellertown, PN, USA) with 0.5 mm ID and 100 mm length. Both ends of the capillary were connected to plastic tubes filled with the DNA solution, agarose gels plugs and buffer from the bottom up (Fig. 1). The electrodes were plunged into the top part of the tube containing the buffer. The EOF was significantly reduced by the pDMA coating. The capillary walls were "conditioned" with a 1% w/w pDMA solution overnight. The EOF was measured for electric fields ranging from 1–20 V/cm using fluorescent dextran, and was found to be less than 0.2×10^{-4} cm^2/Vs, a value that agrees with the observation made by Madabhushi [26]. The residual EOF is thus only about 5% of the dsDNA free solution mobility.

2.2.3 The fluorescence recovery after photobleaching (FRAP) setup

The setup is similar to the one described by Davoust *et al.* [27]. Two coherent laser beams (λ = 488 nm) are crossed in the sample to create a fringe pattern. The fringe spacing $i = 2\pi/q$ is set by the crossing angle Θ and ranges from 3 to 60 μm while $q = (4\pi/\lambda)\sin(\Theta/2)$ is the scattering vector. Fluorescent bleaching of the YOYO-labeled DNAs is obtained by producing a 1 s full intensity pulse. This generates a fringe pattern of concentration of fluorescent molecules. The amplitude of this pattern is detected by modulation of the illuminating fringe position using a piezoelectrically driven mirror and lock-in detection of the emerging fluorescence collected at the photomultiplier by an optical fiber. In the Guinier regime, where the size of the particles is much smaller than the fringe spacing, the amplitude of the experimental signal decays exponentially because of thermal diffusion and, when an electric field is applied, the migration of the fluorescent particles provides an additional sinusoidal signal (Fig. 2).

The experimental signal is thus fitted with the expression:

$$y = Ae^{-\frac{(t-t_0)}{\tau}} \sin\left(\frac{2\pi\ (t-t_0)}{T}\right) \quad (6)$$

When an electric field E is applied perpendicularly to the fringes, the molecules migrate over the distance x with a velocity $u = \mu E$ where μ, the electrophoretic mobility, is given by $\mu = i/ET$. In the absence of an electric field (a single exponential decrease), the translational diffusion coefficient is given by $D = 1/q^2\tau$. When the field is applied, one can still use this relation under the condition that the drift of the sample, which is also a cause of the decrease of the signal, is small compared with diffusion. In the diffusive regime, $\tau \sim 1/q^2$ whereas in the drift regime $\tau \sim 1/q$ [10]. Depending on the values of the characteristic times T and τ, we must choose q large enough so that the drift is negligible and small enough to stay in the Guinier regime while keeping τ above 1 s. For the dsDNAs we used, these conditions imply that we could not apply fields higher than about E = 3 V/cm (see Fig. 3).

3 Results

3.1 ssDNA

3.1.1 The free-flow mobility μ_0 and the absence of Joule heating

Figure 4 shows the velocity of our 75 base ssDNA fragment as a function of the electric field intensity E. The straight line fit indicates that Joule effects are negligible

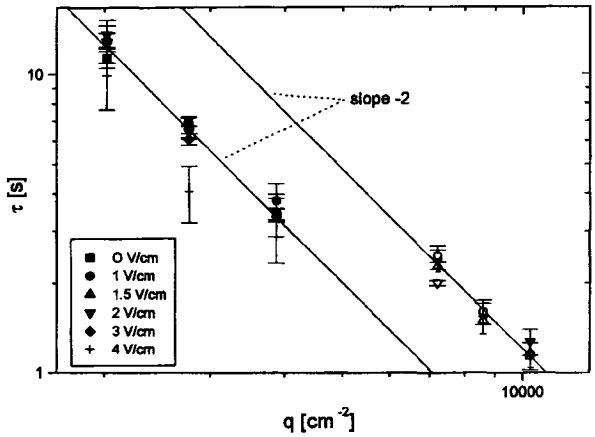


Figure 3. Log-log plot of τ *vs.* q. For E ≤ 3 V/cm, the data are compatible with a –2 slope which indicates a diffusive process.

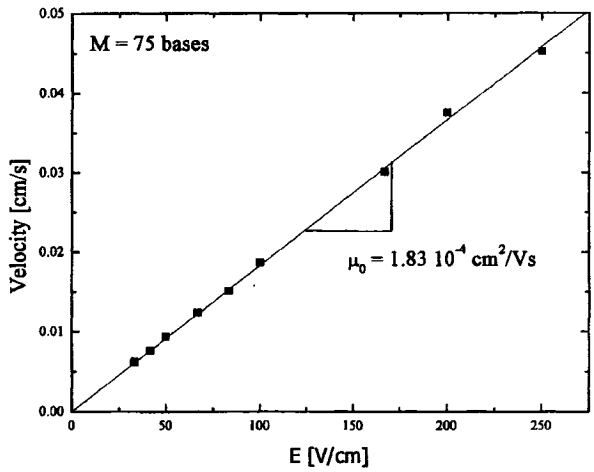


Figure 4. Electrophoretic velocity *u(E) vs.* field intensity E for an M = 75 base ssDNA fragment. The slope gives the (molecular size-independent) free-solution mobility μ_0, as shown. No nonlinear effect (*e.g.*, due to Joule heating) is present.

over this range of electric field intensities, an important factor for the rest of the analysis. The slope of the fit gives the ssDNA free-flow mobility $\mu_0 = 1.83 \times 10^{-4}$ cm^2/Vs.

3.1.2 The zero-field diffusion coefficients

Figure 5 shows the final band variance $\sigma^2(t)$ as a function of twice the stop-time $(2t)$ for a M = 75 base ssDNA molecule. According to Eq. (3), the diffusion coefficient is thus given by the slope, which yields the zero-field diffusion coefficient D_0 = $(4.6 \pm 0.3) \times 10^{-7}$ cm^2/s in this case. The fit also indicates that the other contributions to the final band width (*i.e.*, the

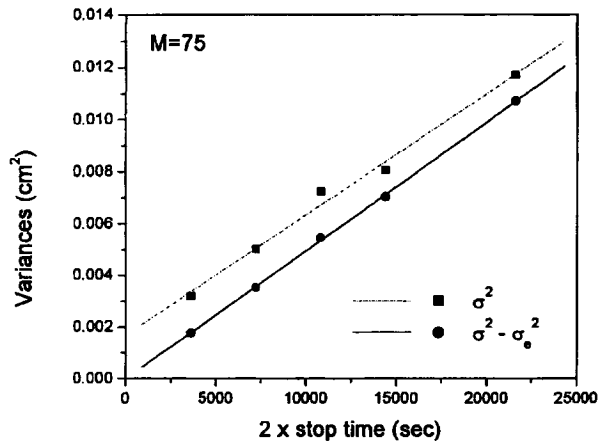


Figure 5. Final peak variance *vs.* twice the stop-time $(2t)$ for a M = 75 base ssDNA fragment. The slope gives the zero-field diffusion coefficient D_0. Upper curve: raw data, the straight line fit is variance = $[(1.70 \pm 0.036)\ 10^{-3}\ \text{cm}^2] + [(4.65 \pm 0.27) \times 10^{-7}\ \text{cm}^2/\text{s}] \times [2t]$. Lower curve: we subtracted the variance of the second loading; the straight line fit is variance = $[(-0.04 \pm 0.08) \times 10^{-3}\ \text{cm}^2] + [(4.95 \pm 0.06) \times 10^{-7}\ \text{cm}^2/\text{s}] \times [2t]$.

band broadening that comes from injection and electrophoretic migration) are $(\sigma_e)^2 \cong (17 \pm 4) \times 10^{-4}$ cm^2. The mean electrophoretic elution time was t_e = 1063 ± 11 s under these conditions; assuming that the diffusion coefficient is also $D_e = D_0 = (4.6 \pm 0.3) \times 10^{-7}$ cm^2/s during the electrophoresis process (a result that we will demonstrate later), we would expect a residual band variance of $(\sigma_D)^2 = 2D_e t_e \cong (9.8 \pm 0.6) \times 10^{-4}$ cm^2 from this process alone. We thus conclude that the contribution of the injection process, given roughly by $(\sigma_e)^2 - (\sigma_D)^2 \cong (0.27 \pm 0.17\ \text{mm})^2$, is too small to be reliably estimated in these experiments. Therefore, we will not study the injection process any further. Table 1 gives the values of the zero-field diffusion coefficients $D_0(M)$ for the various ssDNA molecules that we studied. In each case, two values are given: the first one was obtained as described in Fig. 4 for the M = 75 base fragment (*i.e.*, without an independent measurement of the electrophoretic band width σ_e), while the other was obtained after subtracting the value of σ_e obtained from the second sample loading (as described in Section 2.3). The consistency of our values suggests that our data is reliable to within about ± 10% in all cases. Also, there is no systematic trend.

3.1.3 Thermal diffusion during free-flow electrophoresis

In order to estimate the diffusion coefficient during the electrophoresis itself, we measured the final band width at elution and calculated the plate height, as described in

Table 1. Diffusion coefficient D for different ssDNA molecular sizes M

M bases	D_{0+} 10^{-7} cm^2/s	D_{0-} 10^{-7} cm^2/s	D_e 10^{-7} cm^2/s
18	9.94 (75)	9.81 (38)	9.7 (11)
50	6.15 (41)	6.16 (58)	6.61 (48)
75	4.65 (27)	4.95 (06)	5.04 (38)
100	4.03 (34)	4.04 (17)	4.09 (18)
160	2.72 (27)	2.36 (23)	2.76 (32)
200	1.74 (20)	1.80 (04)	2.17 (07)
250	1.97 (18)	1.88 (21)	1.66 (11)

D_0 and D_e stand for the diffusion coefficients at zero electric field and with the electric being applied, respectively. In the first case, D_{0+} means that we did not subtract the variance of the second loading, while D_{0-} means that we did (see Fig. 5). The error on the last two significant digits is given in parenthesis.

Section 2.3. Figure 6 shows the plate height H *vs.* the ratio $2/u$ for the 75 base fragment. The slope of the straight-line fit gives $D_e = (5.0 \pm 0.4) \times 10^{-7}$ cm^2/s, in good agreement with the value found in the previous section using the stop-flow method. Since H does not seem to increase for small values of $2/u$, we conclude that the wall-analyte interactions (the Wu term in Eq. (4)) play a very minor role. The intercept at the origin is often negative and always very small; therefore, we again conclude that we cannot obtain reliable values of the injection width σ_0 using our CE approach. Table 1 also gives the values of the diffusion coefficients obtained using this method. As we can see, the values are in excellent agreement with those obtained in Section 3.1 using the stop-time method. We thus conclude that the electric field has no effect on the diffusion coefficient of ssDNA fragments during electrophoresis.

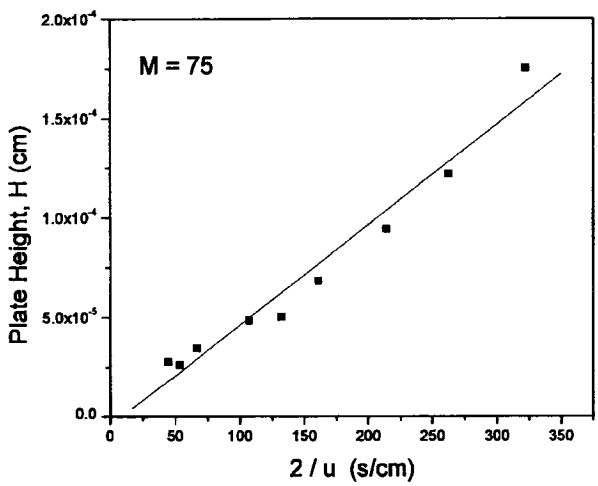


Figure 6. Plate height H *vs.* twice the inverse velocity, $2/u$, for the M = 75 base ssDNA fragment. The slope of the linear fit, $H = [(-4.3 \pm 6.8) \times 10^{-6}$ cm$] + [(5.04 \pm 0.38) \times 10^{-7}$ cm^2/s$] \times [2/u]$, gives directly the diffusion coefficient D.

3.1.4 Size-dependence of the diffusion coefficient

The three sets of diffusion coefficients given in Table 1 are plotted in Fig. 7 as a function of the ssDNA size M. The slope of the linear fit (solid line) shows an approximate scaling law $D \sim 1/M^{(0.68 \pm 0.03)}$. It is interesting to note that both Stellwagen *et al.* [22] and Sorlie and Pecora [25] have reported the same scaling law for short dsDNA fragments. Since the exponent ($\cong$0.68) is significantly larger than even Flory's excluded volume exponent of 3/5, our results strongly suggest that our ssDNA fragments were too short to satisfy the asymptotic behaviour $D \sim 1/M^{3/5}$ which is expected only for very long flexible polymers. Since the persistence length of ssDNA is roughly $p \cong 7.2$ nm or about 17 bases for this buffer composition [6], our ssDNA fragments ranged from about 1 to about 15 persistence lengths only. As suggested by Sorlie and Pecora [25], one can instead analyze the data using Eqs. (2) and (5) since it is then possible to obtain a $D(M)$ *vs.* M curve with an effective exponent $\cong$0.68 over a fair range of molecular sizes. On the other hand, omitting the first point gives an excellent fit with a scaling $D \sim 1/M^{(0.83 \pm 0.04)}$ (not shown). The first point corresponds to the M = 18 base fragment, a rigid rod molecule for which the approximation $R_H = (2/3)R_G$ is certainly not appropriate.

In fact, the diffusion coefficient of a rigid rod DNA molecule should be given by the expression [25]:

$$D \cong \frac{k_B T}{3\pi\eta L_D/(\ln(L_D/d_D) + \gamma)} \qquad (7)$$

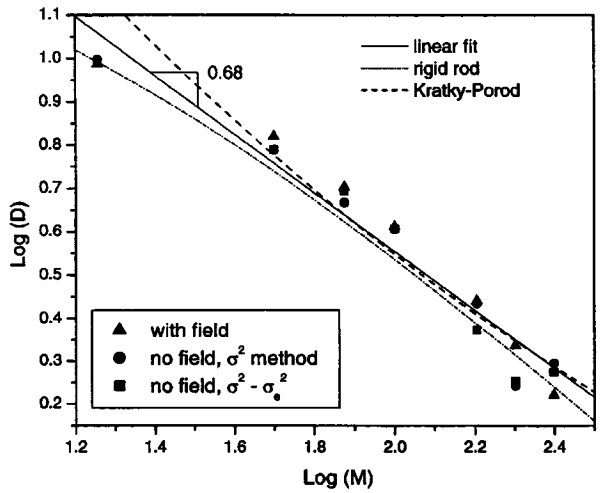


Figure 7. Diffusion coefficient D *vs.* ssDNA molecular size M (in bases). The three fits are described in Section 3.1.4.

where $d_D \cong 1.1$ nm is the diameter of the ssDNA backbone, $L_D = Mb$ is the contour length of the fragment, and $\gamma_{(M=18)} \cong -0.73$ is a correction for end effects. If we use the Table 1 average value of $D = (9.82 \pm 0.46) \times 10^{-7}$ cm^2/s for the $M = 18$ base fragment, Eq. (7) gives $k_BT/3\pi\eta b = (9.21 \pm 0.43) \times 10^{-5}$ cm^2/s. This means that the viscosity of the 7M urea solvent is predicted to be $\eta = (1.12 \pm 0.05) \times 10^{-3}$ Pas, about 41% higher than the viscosity of water at this temperature, a reasonable value. We can then use our estimated value of $k_BT/3\pi\eta b$ together with Eqs. (2) and (5) to fit the data for the $M > 18$ bases points. The resulting fit (upper line in Fig. 7) is quite good, and we obtain a persistence length $p = 8.1 \pm 0.8$ nm, in decent agreement with the value of $p = 7.2$ nm recently obtained by Desruisseaux *et al.* for similar buffer conditions. We thus conclude that the diffusion properties of these small ssDNA fragments can be well explained by Zimm's theory and the proper expression for the hydrodynamic radius of the molecule (*i.e.*, either Eq. 5 or Eq. 7). The lower line in Fig. 7 also shows the prediction given by Eq. (7) using the numerical values described above (which, again, were obtained solely from the $M = 18$ bases point). The prediction is actually rather good too (although it is expected that it would diverge more and more for larger DNA sizes). This again indicates that these ssDNA fragments are rather small compared to their persistence length.

3.2 dsDNA

3.2.1 Free-flow mobility measurements

Figure 8 shows the velocity obtained from the value of T as a function of the electric field for dsDNA sizes ranging from 2.1 kbp to 164 kbp. In Fig. 9, we show the results

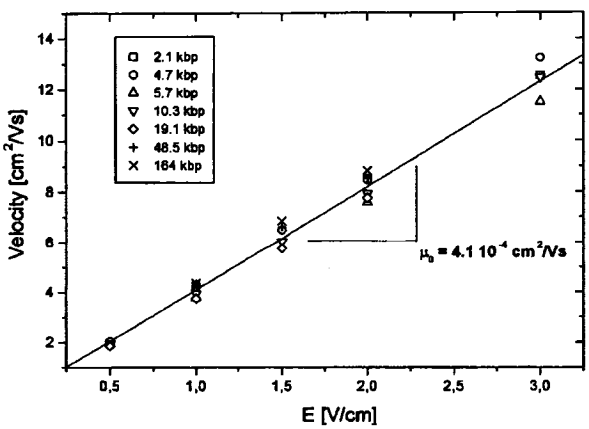


Figure 8. Velocity *vs.* the electric field intensity ($E < 3$ V/cm) for various dsDNA sizes.

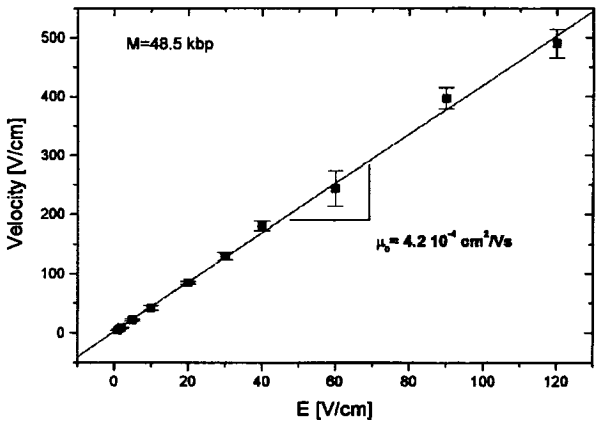


Figure 9. Velocity of the λ- dsDNA fragment *vs.* the field intensity E (< 120 V/cm).

obtained for larger electric fields (up to 120 V/cm) for the λ-DNA. Each point is an average over 6 consecutive measurements. A clear result is that the velocity is linear with the field , so μ does not depend on E. Also, there is no size effect for the six DNA samples used. These types of results have already been reported in previous experiments using indirect [28] or less accurate methods [29]. Stellwagen *et al.* [22] recently performed direct measurements on double stranded DNA with field intensities above 50 V/cm. Here we provide experimental values for a wide range of DNA sizes and several electric fields. Using the FRAP setup allows quasi instantaneous measurements preventing the effects of possible DNA/capillary wall interactions (the molecules move over microns instead of several centimeters). The mean value obtained for all the samples at different fields is $\mu_0 = (4.1 \pm 0.4) \times 10^{-4}$ cm^2/Vs.

3.2.2 Diffusion measurements

Figure 10 shows the diffusion coefficient obtained from the value of τ as a function of the electric field intensity E. There is no systematic dependence with E. Two factors explain the slight variations observed : the first is the effect of the drift that increases with the field ; the second is the uncertainty that arises when we change the q vector in order to minimize the contribution of the drift to the experimental signal. These results clearly indicate that the diffusion coefficient of dsDNA is not affected by the electric field.

3.2.3 Size-dependence of the diffusion coefficient

In Fig. 11, we show a log-log plot of D *vs.* M for different field intensities. The scaling law $D \sim 1/M^{(0.57)}$ can fit all of our data ($E = 0$ and $E > 0$). Thus, the Nernst-Ein-

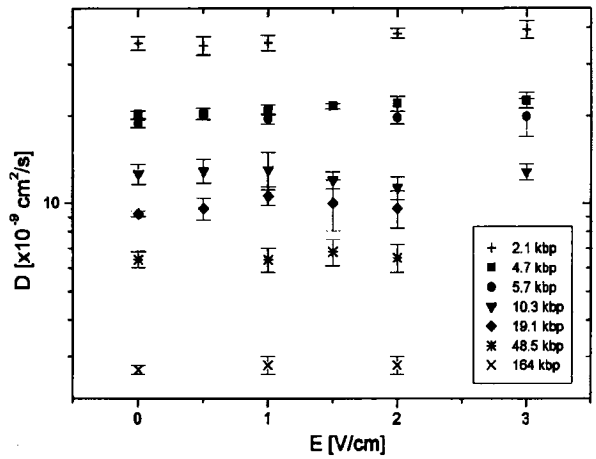


Figure 10. Translational diffusion coefficient *D* for various DNA sizes *vs.* field intensity *E*.

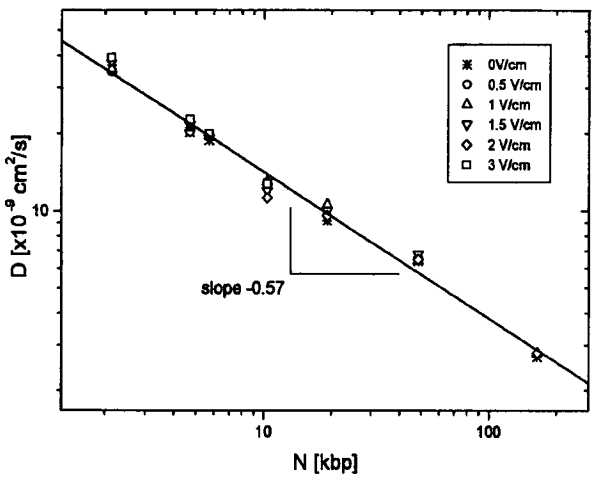


Figure 11. Diffusion coefficient *vs.* dsDNA lengths (in base pairs). The slope (–0.57) indicates a Zimm's scaling for all molecular lengths.

stein relation is not valid, even for high field intensities. The analysis in Section 3.1.4 can also be made here. One can either see this result as an effective scaling law coming from the use of the Kratky-Porod equation (neglecting excluded volume effects), or as evidence that excluded volume effects are important (in which case we indeed expect ~$1/M^{0.60}$). However, our dsDNAs range from the 2.1 kbp to the T2 (164 kbp). This corresponds to a length variation from 7 to 164 Kuhn segments. Therefore our data are fully compatible with a behavior starting with weakly bending rods and going to Gaussian chains with excluded volume interactions.

4 Discussion

Our results, for both ssDNA and dsDNA, demonstrate unequivocally that the diffusion coefficient of these polyelectrolytes is not a function of the electric field during free-flow electrophoresis. In other words, the Nernst-Einstein equation is not valid under such simple conditions. One might say that the DNA coil is free-draining for the mobility but not for the diffusion; this distinction is actually due to the fact that while the mobility is directly related to the net flow of solvent around and through the coil, the diffusion is simply related to fluctuations in the flow and the latter, unlike the former, are not affected by the presence of the electric field. In other words, while the mobility is related to electrostatic factors, diffusion is purely mechanical. In such cases, the Nernst-Einstein relation cannot be expected to be valid since mobility and diffusion are related to different physical mechanisms.

We note in passing that the failure of the Nernst-Einstein relation, *i.e.* the fact that the diffusion coefficient is not field-dependent, assures the validity of Galilean relativity. In a frame of reference that moves with the DNA molecule, it is not possible to estimate the field intensity from the diffusion coefficient. In gels, the situation is different because the gel fibers affect the hydrodynamic interactions through a screening process [30]. This restores the validity of the Nernst-Einstein equation at low field. However, the reptation process easily brings the system out-of-equilibrium and Eq. (1) is again violated.

Our analysis of the diffusion data agrees well with the Zimm theory for a nonfree draining polymer. This means that with a good model for the hydrodynamic radius $R_H(M)$, one can use Eq. (2) to predict fairly accurately the diffusion coefficient during free-flow electrophoresis. Equation (7) applies to short rod-like fragments, the Kratky-Porod equation provides an excellent model for $R_H(M)$ when the polymer chains are not very long compared to their persistence length, and finally Flory's scaling form is valid for longer chains that are affected by excluded volume effects. Perhaps the largest impact of these results is for the prediction of the optimal performance of both end-labeled free-solution electrophoresis (ELFSE) and free-solution conjugate electrophoresis (FSCE). ELFSE [31] uses a molecular drag-parachute to achieve the separation of DNA molecules in free-solution CE, while FSCE uses a DNA engine to separate a polydisperse polymer solution [32]. In both cases, the diffusion coefficient must be given by the appropriate Zimm expression, and not by the Nernst-Einstein equation. Since using the latter tends to grossly underestimate the diffusion coefficient, predictions based on the Nernst-Einstein relation will predict unrealistically good performance. Our recent theoretical work [31] has indeed

assumed the validity of Zimm's equation, but we were not aware of any experimental test of this subtle point. The current study answers this fundamental question and establishes the validity of these calculations.

The authors would like to thank Didier Long for useful discussions. This work was supported in part by Applied Biosystems and by Research Grants from the Natural Science and Engineering Research Council (NSERC) of Canada to GWS and GD. LCM would like to thank NSERC and the University of Ottawa for scholarship awards and grants. We thank also Dr Galin for the gift of pDMA.

Received March 2, 2001

5 References

[1] Slater, G. W., Desruisseaux, C., Hubert, S. J., Mercier, J.-F., Labrie, J., Boileau, J., Tessier, F., Pépin, M. P., *Electrophoresis* 2000, *21*, 3873–3887.

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

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

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

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

[6] Desruisseaux, C., Long, D., Drouin, G., Slater, G. W., *Macromolecules* 2000, *34*, 44–52.

[7] Pernodet, N., Tinland, B., *Biopolymers*, 1997, *42*, 471–478.

[8] Tinland, B., Pernodet, N., Pluen, A., *Biopolymers*, 1998, *46*, 201–214.

[9] Pluen, A., Tinland, B., Sturm, J., Weill, G., *Electrophoresis* 1998, *19*, 1548–1559.

[10] Meistermann, L., Tinland, B., *Phys. Rev. E*, 1998, *58*, 4801–4806.

[11] Tinland, B., Pluen, A., Sturm, J., Weill, G., *Macromolecules* 1997, *30*, 5763–5765.

[12] Slater, G. W., *Electrophoresis* 1993, *14*, 1–7.

[13] Duke, T. A. J., Viovy, J. L., Semenov, A. N., *Biopolymers* 1994, *34*, 239–247.

[14] Semenov, A. N., Joanny, J. F., *Phys. Rev. E*, 1997, *55*, 789–799.

[15] Slater, G. W., Mayer, P., Grossman, P. D., *Electrophoresis* 1995, *16*, 75–83.

[16] Cottet, H., Gareil, P., Viovy, J. L., *Electrophoresis* 1998, *19*, 2151–2162.

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

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

[19] Gurrieri, S., Rizzarelli, E., Beach, D., Bustamante, C., *Biochemistry* 1990, *29*, 3396–3401.

[20] Pathria, R. K., *Statistical Mechanics*, Butterworth Heinemann, Oxford 1996.

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

[22] Stellwagen, N. C., Magnusdottir, S., Gelfi, C., Righetti, P. G., *Biopolymers* 2001, *58*, 390–397.

[23] Kratky, O., Porod, G., *Rec. Trav. Chim.* 1949, *68*, 1106–1152.

[24] des Cloizeaux, J., Jannink, G., *Polymers in Solution: Their Modelling and Structure*, Oxford Science Publishers, New York 1990.

[25] Sorlie, S. S., Pecora, R., *Macromolecules* 1990, *23*, 487–497.

[26] Madabhushi, R.S., *Electrophoresis* 1998, *19*, 224–230.

[27] Davoust, J., Devaux, P. F., Leger, L., *EMBO J.*, 1982, *1*, 1233–1238.

[28] Hervet, H., Bean, P. C., *Biopolymers* 1987, *26*, 727–742.

[29] Ross, P. C., Scruggs, R. L., *Biopolymers* 1964, *2*, 231–236.

[30] Long, D., Ajdari, A., *Eur. Phys. J. E.* 2001, *4*, 29–32.

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

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

7

Electrophoresis in the Presence of Viscosity Gradients

S. Guillouzic, L.C. McCormick, G.W. Slater, *Electrophoresis* **23**, 1822–1832 (2002)

The effect of viscosity gradients within the capillary during electrophoresis are analysed in this paper. It is shown that such gradients only serve to decrease resolution under optimal conditions where the band width is limited by diffusion. When conditions are less than optimal and the initial band loading width is non-negligible, a gradient can be used to increase resolution, but is ideally located at the beginning of the capillary thereby serving only to increase sample stacking, a process commonly used to sharpen bands upon loading.

This paper is a result of a project that I had started; it is primarily the result of work undertaken by Steve Guillouzic. This third paper was written by Steve; I derived the results presented in section 2.

1822 *Electrophoresis* 2002, *23*, 1822–1832

Steve Guillouzic
Laurette C. McCormick
Gary W. Slater

Department of Physics,
University of Ottawa,
Ottawa, Canada

Electrophoresis in the presence of gradients: I. Viscosity gradients

In many cases, the resolution provided by capillary electrophoresis systems approaches that predicted for diffusion-limited separations. Once all device-related sources of band broadening have been eliminated or minimized, only thermal diffusion remains. In principle, peaks can be sharpened using gradients of various system characteristics such as gel concentration, buffer viscosity and electric field. However, it is not clear whether this can actually increase the resolution of the system. In this article, we focus our attention on viscosity gradients and we examine both continuous and step-like variations. Our results indicate that the performance of electrophoretic systems cannot be improved by viscosity gradients. They may provide extra stacking, and thus improve the resolution, when the injection width is non-negligible. However, for the systems considered here, the best resolution is obtained when the viscosity is uniform and the stacking is entirely performed at injection. We conclude by discussing the link between these results, the fundamental laws of thermodynamics, the nature of the detection process and the importance of having nonlinear effects in nonuniform systems.

Keywords: Band broadening / Capillary electrophoresis / Resolution / Self-focusing bands / Viscosity gradients EL 4909

1 Introduction

Although electrophoretic methods are being used extensively to separate biological macromolecules such as DNA and proteins, there are still numerous unresolved theoretical issues related to the optimization of these technologies. The performance of a separation system is characterized by three factors: (i) the elution times of the analytes; (ii) the peak spacing; (iii) the peak widths. Optimization of the experimental conditions usually focuses on these parameters, or more frequently on combinations of these parameters such as the number of theoretical plates or the resolution factor. In the ideal situation, the peak widths are controlled solely by diffusion processes and extremely high resolutions can be achieved. Therefore, optimization requires that we possess a precise knowledge of the dependence of the analytes' electrophoretic mobility μ and diffusion coefficient D upon the various experimental variables (field, temperature, ionic strength, *etc.*).

In general, the experimental parameters of an electrophoresis run are spatially uniform and constant in time. Even such straightforward cases are often extremely hard to fully understand from a theoretical point of view, but much progress has been made in several cases, especially for DNA separation methods [1, 2]. Temporal gradients, where one or more parameters are changed during the course of the experiment, are easy to handle when they use low frequencies. The final result is then a linear combination of the results expected for the different values of the experimental parameters, weighted according to the duration of their use. However, one can observe truly remarkable and highly nontrivial phenomena when the temporal changes of the experimental parameters are in resonance with some natural time scale of the separation process. A famous example of this is provided by the various pulsed-field gel electrophoresis techniques that exploit the reorientation times of large DNA molecules in order to separate them [1–3]. Because fast temporal switching of the experimental parameters keep the system far from any equilibrium or steady-state situation, pulses can indeed lead to much better separations than predicted for constant experimental conditions.

Spatial gradients, however, are quite different. First, they are usually applied over very long length scales (compared to the size of the molecules). Indeed, while biological macromolecules may have relaxation times in the seconds or minutes range, their sizes are frequently in the μm or sub-μm range. Unless one uses microlithographic (or perhaps nanolithographic) systems with molecular-size features, a spatial gradient will look like a uniform experimental condition over the length scale of the molecule. Second, spatial gradients have quite a different effect on

Correspondence: Professor Gary W. Slater, Department of Physics, University of Ottawa, 150 Louis-Pasteur, Ottawa, Ontario KIN 6N5, Canada
E-mail: gslater@science.uottawa.ca
Fax: +613-562-5190

Abbreviation: SDE, stochastic differential equation

Electrophoresis 2002, *23*, 1822–1832

finishline (elution) and snapshot (standard slab gel) systems. In the first case, all molecules visit the same gradients, while in the second the slower molecules visit only a fraction of the system before the power is turned off and the result of the separation is visualized. One would thus expect that, based on simple thermodynamic principles, spatial gradients can improve the resolution of snapshot systems. Indeed, this has been studied and used in the past [4, 5]. A special and extremely important example is that of isoelectric focusing of proteins in pH gradients. A snapshot focusing method has also been recently developed for DNA based on subtle nonlinear dynamics effects [6, 7]. The situation with finishline/elution methods has however not been studied very carefully until now.

In this paper, we present the first in a series of investigations that will examine the effect of various long-wavelength (macroscopic) gradients on the performance of elution separation systems. Such studies must treat a number of issues very carefully. For instance, the cases of linear and nonlinear coupling of the analyte dynamics to the spatial gradient(s) are fundamentally different. Continuous and step gradients may also play different roles. Furthermore, we must consider the effect of the gradient on elution times, peak width and peak asymmetry. Finally, each type of gradient (or combination of gradients) has a different impact on the electrophoretic process. In this article, we study the simplest possible gradient system: a viscosity gradient in an otherwise uniform and linear system. This will allow us to develop a theoretical framework for studying gradients and to reach general conclusions that apply to all gradients. Such a gradient can be created, *e.g.*, using the CE loading procedure recently introduced by the group of Chu [8].

The paper is organized as follows. In Section 2, we examine the effect of a simple viscosity step in the middle of the capillary. We treat it as a second stacking problem and show that such gradients are not useful unless injection was poor in the first place. Section 3 provides an in-depth study of viscosity gradients using a more microscopic approach based on the theory of first passage times. This analysis is completed by an appendix in which the diffusion of analytes in the absence of an electric field is considered.

2 The simple viscosity step

Before we investigate the effect of continuous viscosity gradients on electrophoretic separation, let us examine the effect of a step gradient placed between the injection point (at $x = 0$) and the detection region (at $x = L$) (Fig. 1). This will give us useful information about gradients in general, as well as about the difference between continuous and step gradients. The viscosity gradient can in general be defined as

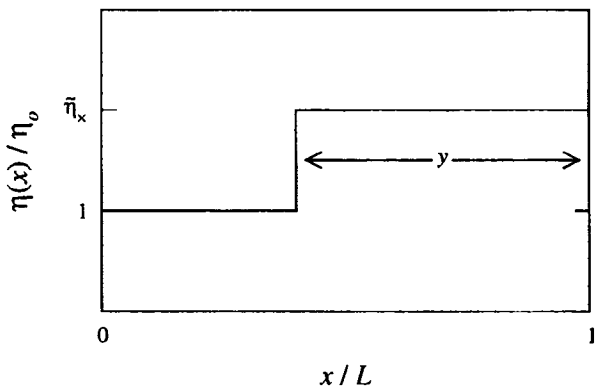


Figure 1. The step viscosity gradient. A second step could exist at $x = 0$ *for* sample stacking at injection, but this is not part of the calculation.

$$\eta(x) = \eta_o \tilde{\eta}(x) \tag{1}$$

where η_o is the viscosity of the buffer solution at a given reference point and $\tilde{\eta}(x)$ is a dimensionless function. For the step gradient, we define η_o as being the viscosity at (or, more precisely, immediately after) the injection point ($x = 0$) and

$$\tilde{\eta}(x) \equiv \begin{cases} 1 & \text{for } 0 < x < (1-y)\,L \\ \tilde{\eta}_x & \text{for } (1-y)\,L < x < L \end{cases} \tag{2}$$

where $\tilde{\eta}_x$ is the multiplicative increment of the step ($\tilde{\eta}_x \geq 1$), y determines its position ($0 \leq y \leq 1$) and L is the elution length. In the linear case studied in this article, the velocity v and the diffusion coefficient D of the analytes are related to the local viscosity of the medium through the equations

$$v(x) = \frac{v_o}{\tilde{\eta}(x)} \tag{3}$$

and

$$D(x) = \frac{D_o}{\tilde{\eta}(x)} \tag{4}$$

where v_o and D_o are the velocity and the diffusion coefficient in regions of space where the viscosity is equal to η_o. (Both v_o and D_o are positive.) Clearly, $v(x) = v_o$ and $D(x) = D_o$ in the low-viscosity region [$0 < x < (1 - y)L$], and $v(x) = v_o/\tilde{\eta}_x \equiv v_o'$ and $D(x) = D_o/\tilde{\eta}_x \equiv D_o'$, in the high-viscosity region [$(1 - y)L < x < L$]. In the next subsections, we will examine, in order, the diffusion-limited case, the case with zero-diffusion, and finally the general case with both diffusion and a finite initial injection band (peak) width.

2.1 The diffusion-limited case

The mean elution time through the low viscosity region is given by

$$t = \frac{(1-y)L}{v_0} \tag{5}$$

and through the high viscosity region, by

$$t' = \frac{yL}{v_0'} \tag{6}$$

The mean elution time from the injection point to the detector is thus equal to

$$t_e(\tilde{\eta}_x) = \frac{L}{v_0}[1 + (\tilde{\eta}_x - 1)y] \tag{7}$$

Using this equation, it is found that increasing v_0 by an increment Δv_0 decreases the mean elution time by

$$\Delta t_e(\tilde{\eta}_x) = \Delta t_e(1)[1 + (\tilde{\eta}_x - 1)y] \tag{8}$$

where

$$\Delta t_e(1) = \left(\frac{1}{v_0} - \frac{1}{v_0 + \Delta v_0}\right)L \tag{9}$$

is the net difference in elution time in the absence of a viscosity step. Equation (8) thus gives us the temporal peak spacing between two analytes whose velocities at the injection point are respectively v_0 and $v_0 + \Delta v_0$. We observe in this equation an increase of temporal peak spacing due to the reduced velocity in the high viscosity region. This effect disappears if $y = 0$, *i.e.*, if the viscosity step is near the detector.

The final spatial variance of a peak whose injection width is negligible is given by

$$\sigma^2(\tilde{\eta}_x) = \frac{2D_0 t}{\tilde{\eta}_x^2} + 2D_0' t' \tag{10}$$

The denominator of the first term is due to the fact that the step change in viscosity stacks the band variance by a factor proportional to the ratio of the velocities across the interface. Equation (10) can also be written as

$$\sigma^2(\tilde{\eta}_x) = \sigma^2(1)\left(\frac{1-y}{\tilde{\eta}_x} + y\right) \tag{11}$$

where, as before, the parameter

$$\sigma^2(1) = \frac{2D_0 L}{v_0} \tag{12}$$

refers to the homogeneous case (no viscosity step). The corresponding temporal width of the band is

$$\tau(\tilde{\eta}_x) = \frac{\sigma(\tilde{\eta}_x)}{v_0'} = \tau(1)\sqrt{1 + (\tilde{\eta}_x^2 - 1)y} \tag{13}$$

where, once again,

$$\tau(1) = \sqrt{\frac{2D_0 L}{v_0^3}} \tag{14}$$

refers to the homogeneous case.

We now proceed with the determination of the resolution factor for two analytes with similar elution times and temporal peak widths. In other words, we are looking at situations close to the limit of resolution of the system. We assume that the difference between the velocities of the analytes, as well as between their diffusion coefficients, is negligible except for the calculation of the peak spacing $\Delta t_e(\tilde{\eta}_x)$. This allows us to define the resolution factor as

$$R(\tilde{\eta}_x) \equiv \frac{\Delta t_e(\tilde{\eta}_x)}{2\tau(\tilde{\eta}_x)} \tag{15}$$

where $\Delta t_e(\eta_x)$ and $\tau(\eta_x)$ are respectively given by Eqs. (8) and (13). For the step gradient and a zero injection width, we find that

$$\frac{R(\tilde{\eta}_x)}{R(1)} = \frac{1 + (\tilde{\eta}_x - 1)y}{\sqrt{1 + (\tilde{\eta}_x^2 - 1)y}} \tag{16}$$

where $R(1)$ is the value of the resolution factor in the homogeneous case. This ratio is plotted in Fig. 2 against y and $\tilde{\eta}_x$, respectively the position and the height of the viscosity step. The resolution is always lowered by the presence of the step. In other words, we cannot gain anything in this limit by using step viscosity gradients. The best position for the step is at $y = 0$ or $y = 1$, in which cases the resolution is left unchanged and the viscosity is uniform. The worst possible location is at $y_{min} = 1/(\tilde{\eta}_x + 1)$, since this minimizes the resolution factor.

2.2 The injection-limited case

Here, to simplify the analysis, we assume that the injection width σ_0 of the sample is the same for the two molecular species and that it is not a function of the position y of the viscosity step. This is assumed to be the case even when $y = 1$, in which case the step is located at the origin and simply adds to the stacking. As far as elution times are concerned, Eq. (8) is still valid. The final spatial width of the peak, however, is now equal to

$$\sigma(\tilde{\eta}_x) = \frac{\sigma_0}{\tilde{\eta}_x} \tag{17}$$

since there is no diffusion. Using this equation, the final temporal width of the peak is found to be

$$\tau(\tilde{\eta}_x) = \frac{\sigma(\tilde{\eta}_x)}{v_0'} = \frac{\sigma_0}{v_0} \tag{18}$$

Electrophoresis 2002, *23*, 1822–1832

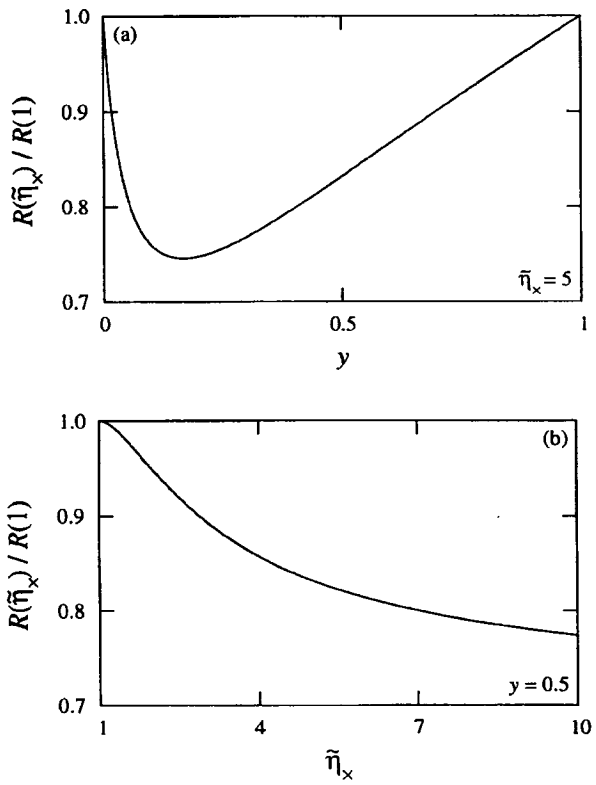


Figure 2. Resolution factor $R(\tilde{\eta}_x)$ divided by its uniform viscosity limit $R(1)$, as a function of (a) the position y and (b) the amplitude $\tilde{\eta}_x$ of the viscosity step for a negligibly small injection width (see Eq. 16). The ratio $R(\tilde{\eta}_x)/R(1)$ is always smaller than unity, indicating a loss of resolution. The worst resolution is found at $y = 1/(\tilde{\eta}_x + 1)$.

and is thus unaffected by the presence of the step gradient. Finally, the resolution factor is in this case given by

$$\frac{R(\tilde{\eta}_x)}{R(1)} = 1 + (\tilde{\eta}_x - 1)y \tag{19}$$

and the viscosity gradient is seen to improve the resolution. This is expected, since the step gradient contributes to the stacking of the initial peak. Indeed, the maximum effect is found for $y = 1$, *i.e.*, when the step gradient serves to stack the injection peak right at the beginning.

2.3 The general case

If both diffusion and injection contribute to the peak variance, we have

$$\sigma^2 = 2D_o't' + \sigma_o'^2 \tag{20}$$

where

$$\sigma_o'^2 = \frac{1}{\tilde{\eta}_x^2}\left(\frac{2(1-y)D_oL}{v_o} + \sigma_o^2\right) \tag{21}$$

is the variance of the peak immediately after it has passed the viscosity step. The corresponding temporal width of the peak at the detector is thus given by

$$\tau = \frac{\sigma}{v_o'} = \sqrt{\frac{2D_oL}{v_o^3}[1 + (\tilde{\eta}_x^2 - 1)y] + \frac{\sigma_o^2}{v_o^2}} \tag{22}$$

and the resolution factor is now found to be

$$R(\tilde{\eta}_x) = \frac{\Delta t_e(1)[1 + (\tilde{\eta}_x - 1)y]}{2\sqrt{\frac{2D_oL}{v_o^3}[1 + (\tilde{\eta}_x^2 - 1)y] + \frac{\sigma_o^2}{v_o^2}}} \tag{23}$$

This equation can also be written as

$$\frac{R(\tilde{\eta}_x)}{R(1)} = \frac{1 + (\tilde{\eta}_x - 1)y}{\sqrt{1 + \frac{(\tilde{\eta}_x^2 - 1)y}{1 + \Sigma_o^2}}} \tag{24}$$

where the new scaled injection width Σ_o is defined as

$$\Sigma_o^2 \equiv \frac{v_o\sigma_o^2}{2D_oL} \tag{25}$$

We note that Eq. (24) reduces to Eq. (16) if $\Sigma_o^2 \ll 1$, *i.e.*, if the injection width is negligible compared to the contribution ($\sim 2DL/v$) of diffusion to the final peak variance. Furthermore, it reduces to Eq. (19) when $1 + \Sigma_o^2 \gg (\tilde{\eta}_x^2 - 1)y$, *i.e.*, when diffusion is negligible.

As expected, a step gradient is only useful if the injection width is the leading contribution to the final peak width. In such a case, the gradient's role is simply to stack the peak further. The ratio given by Eq. (24) is shown in Fig. 3 for two sets of values. In the first case, we have $\tilde{\eta}_x < 1 + 2\Sigma_o^2$, and the resolution is improved no matter where the step is located; however, a step at the injection point $y = 1$ gives optimal performance. In the second case, we have $\tilde{\eta}_x > 1 + 2\Sigma_o^2$, and we then gain only if the viscosity step is located at a position $y > y_1 \equiv [\tilde{\eta}_x - (1 + 2\Sigma_o^2)]/[(\tilde{\eta}_x - 1)(1 + \Sigma_o^2)]$. When $0 < y < y_1$, the situation is actually similar to Section 2.1, where a step gradient degrades the performance. The worst performance happens for $y = y_{min} = [\tilde{\eta}_x - (1 + 2\Sigma_o^2)]/(\tilde{\eta}_x^2 - 1)$.

Perhaps the most remarkable result of this section is the fact that although a step gradient can increase the resolution by stacking the injection width, the location of the gradient can be critical if both the diffusion and injection play non-negligible roles. In this case, placing the step gradient too close to the detector can actually hurt the separation! In all cases, as one would expect, the best performance is obtained when the step gradient is simply used as an additional injection stacking process.

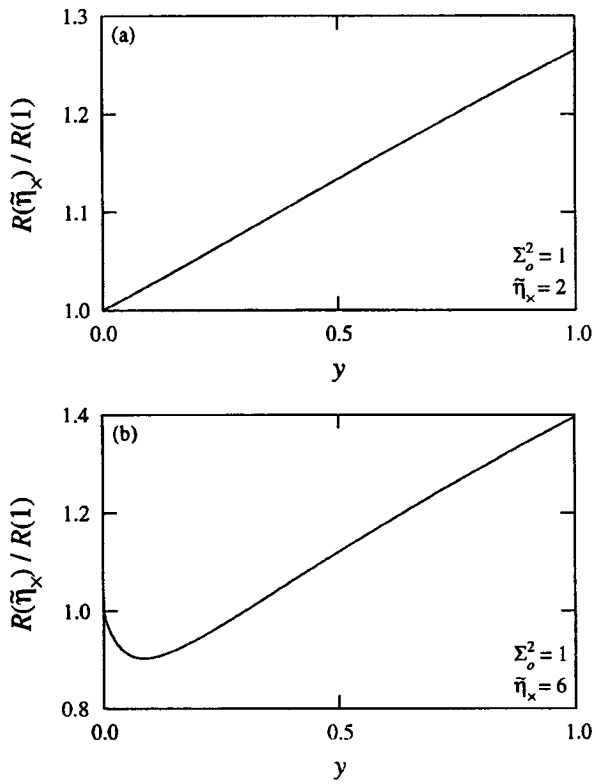


Figure 3. Resolution factor $R(\tilde{\eta}_x)$, divided by its uniform viscosity limit $R(1)$, as a function of the position y *of* the viscosity step for a nonnegligible injection width (see Eq. 24). In graph (a), where $\tilde{\eta}_x < 1 + 2\Sigma_0^2$, the presence of the viscosity step always improves the resolution. In graph (b), on the other hand, $\tilde{\eta}_x > 1 + 2\Sigma_0^2$ and the resolution is decreased when the step is located too close to the detector, *i.e.*, near $y = 0$.

3 Continuous viscosity gradients

In this section, we study the evolution of ensembles of molecules using analytical tools developed in the field of stochastic processes. The position of each molecule is first modeled as a diffusion process, and the moments of the first passage time in front of the finishline detector are then calculated. The resolution factor is obtained from these moments. In this approach, molecules are removed as soon as they reach their destination. This differs from real-world CE experiments, where molecules may pass several times in front of the finishline detector because of diffusion. (We also neglect the fact that the detection zone itself has a finite width.) For a molecule with an average velocity v and a diffusion coefficient D, the displacements due to the drift and the diffusion in a period of time Δt are respectively given by $\Delta x_v = v\Delta t$ and $\Delta x_D = \sqrt{2D\Delta t}$. Multiple readings of individual molecules are expected to happen over intervals of time for which $\Delta x_D \gtrsim \Delta x_v$, or equivalently, $\Delta t \lesssim 2D/v^2$. The theory presented in this section is thus expected to adequately describe all experiments for which times on the order of $2D/v^2$ or smaller do not need to be resolved. The key time for us here is the final peak width. The lower bound for the final temporal width w_t of a peak is given by the diffusion-limited value $w_t = \sqrt{2Dt/v}$, where $t = L/v$ is the elution time of the analyte and L is the length of the capillary. The condition $w_t \gg 2D/v^2$ is thus equivalent to $Lv/2D \gg 1$. This is easily satisfied for essentially all realistic cases. For instance, CE separation would typically correspond to $v \sim 10^{-2}$ cm/s, $D \sim 10^{-7}$ cm^2/s, and $L \sim 50$ cm, giving $Lv/2D \sim 10^6$. The formalism leading to the moments of the first passage time is summarized in Section 3.1, and two examples are discussed in Sections 3.2–3.4. The viscosity step introduced in Section 2 is first revisited in Section 3.2, and an exponential viscosity gradient is then analyzed in Sections 3.3 and 3.4.

3.1 Moments of the first passage time

In this section, the motion of an individual molecule evolving between two reflecting walls located at $x = a$ and $x = b$ (see Fig. 4) is modeled using the Stratonovich stochastic differential equation (SDE)

$$dx = \frac{v_0}{\tilde{\eta}(x)}dt + \sqrt{\frac{2D_0}{\tilde{\eta}(x)}} \circ dW \tag{26}$$

where v_0, D_0 and $\tilde{\eta}(x)$ have the same meaning as in Section 2, W is a Wiener (diffusion) process and "o" denotes the use of Stratonovich calculus. This type of calculus is preferred over Ito calculus for this particular system, because the differential of the Wiener process (dW) appearing in Eq. (26) models a stochastic process with a

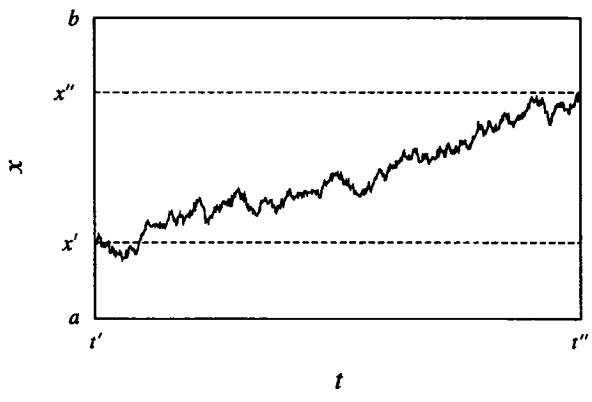


Figure 4. Schematic setup of a first passage time experiment. The analyte is confined to a given region of space by reflecting boundaries located at $x = a$ and $x = b$. It is first injected at $x = x'$ and then moves through both drift and diffusion until it reaches the detector placed at $x = x''$. The first passage time is defined as the transit time from $x = x'$ to $x = x''$, *i.e.*, $t'' - t'$.

Electrophoresis 2002, *23*, 1822–1832

non-zero, though vanishingly small, correlation time. Mathematical manipulations are however more readily performed on Ito SDEs than on Stratonovich SDEs, and Eq. (26) must therefore be converted into the equivalent Ito SDE [9]

$$dx = \frac{v_o}{\tilde{\eta}(x)}\left(1 - \frac{D_o}{2v_o}\frac{d}{dx}\ln\tilde{\eta}(x)\right)dt + \sqrt{\frac{2D_o}{\tilde{\eta}(x)}}\,dW \tag{27}$$

before proceeding with further calculations. When written in this form, the first term models the drift, and the second one the diffusion. This shows that the stochastic term of Eq. (26), which is the sole originator of diffusion in that equation, also contributes to the drift in the presence of a viscosity gradient, *i.e.*, when $d\tilde{\eta}(x)/dx \neq 0$.

The first time that a particle reaches a given point from another one is known as the "first passage", and the duration of such a transit between two points is thus called the "first passage time". This quantity is herein denoted as $T(x', x'')$, where x' and x'' are respectively the starting and ending points (see Fig. 4). In CE, these points correspond to the injection point and the finishline detector. Although $T(x', x'')$ varies from one realization to another in the case of a stochastic process, its moments can be determined using the standard procedure summarized in the following paragraphs [9].

The first step in the derivation of the moments is to convert the SDE (27) into its equivalent backward Fokker-Planck equation

$$\frac{\partial}{\partial t'}p(y,t|x',t') = -\frac{v_o}{\tilde{\eta}(x')}\left(1 - \frac{D_o}{2v_o}\frac{d}{dx'}\ln\tilde{\eta}(x')\right)\frac{\partial}{\partial x'}p(y,t|x',t') - \frac{D_o}{\tilde{\eta}(x')}\frac{\partial^2}{\partial x'^2}p(y,t|x',t') \tag{28}$$

where $p(y, t|x', t')$ is the conditional probability density associated with the state variable x in Eqs. (26) and (27). To quadratic order in dy and with $t > t'$, $p(y,t|x',t')dy$ is the probability that $x(t) \in [y - dy, y]$ given that $x(t') = x'$.

The probability $P(x'', t''|x', t')$ that the particle reaches $x = x''$ at a time $t \leq t''$ given that it started from $x = x'$ at time $t = t'$ is the probability distribution function of $T(x', x'')$. This function can be related to $p(y, t|x', t')$ if an absorbing boundary is set at the final destination $x = x''$. If $x' < x''$, then

$$P(x'', t''|x', t') = 1 - \int_a^{x''} dy\, p(y, t''|x', t') \tag{29}$$

which simply states that the particle remains in the interval $[a, x'')$ until it reaches $x = x''$. A similar expression applies when $x' > x''$. Once $P(x''t''|x't')$ is known, the moments of $T(x', x'')$ are obtained using

$$\langle T^n(x', x'')\rangle = \int_{t'}^{\infty} dt''(t'' - t')^n \frac{\partial}{\partial t''}P(x'', t''|x', t') \tag{30}$$

where $\langle\cdots\rangle$ denotes an ensemble average. Because of normalization, we have

$$\langle T^0(x', x'')\rangle = 1 \tag{31}$$

Under the assumption that $T(x', x'')$ is time-homogeneous (time-shift invariant) and that

$$\lim_{(t''-t')\to\infty}(t'' - t')^n[1 - P(x'', t''|x', t')] = 0 \tag{32}$$

combining Eqs. (28), (29) and (30) leads to

$$\frac{v_o}{\tilde{\eta}(x')}\left(1 - \frac{D_o}{2v_o}\frac{d}{dx'}\ln\tilde{\eta}(x')\right)\frac{\partial}{\partial x'}\langle T^n(x', x'')\rangle + \frac{D_o}{\tilde{\eta}(x')}\frac{\partial^2}{\partial x'^2}\langle T^n(x', x'')\rangle = -n\langle T^{n-1}(x', x'')\rangle \tag{33}$$

which relates each moment $\langle T^n(x', x'')\rangle$ with $n \geq 1$ to the lower order moment $\langle T^{n-1}(x', x'')\rangle$. In order to solve this differential equation, boundary conditions at $x' = x''$ and $x' = a$ must be specified. The absorbing boundary at $x = x''$ implies that

$$\langle T^n(x', x'')\rangle = 0. \tag{34}$$

On the other hand, the reflecting boundary at $x = a$ is implemented by setting

$$\frac{\partial}{\partial x'}p(y, t|x', t')|_{x'=a} = 0 \tag{35}$$

which leads to

$$\frac{\partial}{\partial x'}\langle T^n(x', x'')\rangle_{x'=a} = 0 \tag{36}$$

Applying these boundary conditions to Eq. (33) results in

$$\langle T^n(x', x'')\rangle = \frac{n}{D_o}\int_{x'}^{x''} dy \frac{1}{\alpha(y)}\int_a^y dx\alpha(x)\tilde{\eta}(x)\langle T^{n-1}(x, x'')\rangle \tag{37}$$

where

$$\alpha(y) \equiv \exp\left[\frac{v_o}{D_o}\int^y dx\left(1 - \frac{D_o}{2v_o}\frac{d}{dx}\ln\tilde{\eta}(x)\right)\right] \tag{38}$$

The lower integration bound in Eq. (38) is arbitrary, but it must lie within $[a, x'']$.

When combined with Eq. (31), Eq. (37) allows all the moments of $T(x', x'')$ to be determined iteratively, and from them, all the central moments of interest. The three most often encountered of these central moments are the mean

$$\text{mean}\,(T(x', x'')) \equiv \langle T(x', x'')\rangle \tag{39}$$

the variance

$$\text{var}(T(x', x'')) \equiv \left\langle (T(x', x'') - \langle T(x', x'')\rangle)^2\right\rangle = \langle T^2(x', x'')\rangle - \langle T(x', x'')\rangle^2 \tag{40}$$

and the skewness

$$\mathrm{skew}(T(x',x'')) \equiv \frac{\left\langle (T(x',x'') - \langle T(x',x'')\rangle)^3 \right\rangle}{\left\langle (T(x',x'') - \langle T(x',x'')\rangle)^2 \right\rangle^{3/2}}$$

$$= \frac{\langle T^3(x',x'')\rangle - 3\langle T^2(x',x'')\rangle\langle T(x',x'')\rangle + 2\langle T(x',x'')\rangle^3}{\left(\langle T^2(x',x'')\rangle - \langle T(x',x'')\rangle^2\right)^{3/2}} \quad (41)$$

which respectively characterize the center, the width and the asymmetry of the distribution. More precisely, in the context of CE, mean $(T(x', x''))$ corresponds to the mean elution time of a given molecule species and $\sqrt{\mathrm{var}(T(x',x''))}$ represents its temporal peak width. For the skewness, positive values indicate that the tail of the distribution extending toward $+\infty$ (longer elution times) is larger than the one pointing in the other direction; negative values indicate exactly the opposite. In CE, skew $(T(x', x''))$ is in general expected to be positive, since elution times cannot be negative and, in addition, are not subjected to any upper bound limit when diffusion is present. (Molecular degradation is one of the processes that may induce a negative skewness.)

The moments of the first passage time obtained up to now are applicable only in cases where the injection width is negligible, since all the trajectories considered in the averaging process start from the same point, *i.e.*, $x = x'$. When the injection width is non-zero, we must average the results of Eq. (37) over all the possible starting points in order to obtain the appropriate values of the moments. If $p_0(x')$ is the probability density describing the distribution of analytes immediately after all the molecules have been injected, the moments of the first passage time at $x = x''$ are given by

$$\langle T^n(x'')\rangle = \int dx' p_0(x') \langle T^n(x',x'')\rangle \quad (42)$$

The mean elution time and the temporal peak width can then be determined from these moments in the same way as when the injection width is negligible.

3.2 Case 1: The viscosity step revisited

The theory presented in Section 3.1 can be easily applied to the viscosity step introduced in Section 2. The case of zero injection width, for instance, is obtained by substituting Eq. (2) in Eq. (37) with $x' = 0$ (injection point), $x'' = L$ (finishline detector) and $a \to -\infty$, which leads to

$$\mathrm{mean}(T(0,L)) = \frac{L}{v_0}\left[1 + (\tilde{\eta}_x - 1)y - \frac{D_0}{Lv_0}\left(\tilde{\eta}_x - e^{(\tilde{\eta}_x-1)/2}\right)\left(1 - e^{-yLv_0/D_0}\right)\right] \quad (43)$$

and

$$\mathrm{var}(T(0,L)) = \frac{2D_0L}{v_0^3}\left\{1 + (\tilde{\eta}_x^2 - 1)y + 2\tilde{\eta}_x\left(\tilde{\eta}_x - e^{(\tilde{\eta}_x-1)/2}\right)ye^{-yLv_0/D_0} + \frac{D_0}{2Lv_0}\left[e^{\tilde{\eta}_x-1} + 2(\tilde{\eta}_x+1)e^{(\tilde{\eta}_x-1)/2} - 5\tilde{\eta}_x^2 + 2\left(2\tilde{\eta}_x^2 - e^{(\tilde{\eta}_x-1)/2} - e^{\tilde{\eta}_x-1}\right)e^{-yLv_0/D_0} + \left(e^{(\tilde{\eta}_x-1)/2} - \tilde{\eta}_x\right)^2 e^{-2yLv_0/D_0}\right]\right\} \quad (44)$$

The $a \to -\infty$ limit needs to be considered in order to avoid spurious terms arising from reflections on the wall located at $x = a$.

The extra terms appearing in Eqs. (43) and (44), when compared with the results of Section 2 (see Eqs. 7 and 13), are due to molecules that hop back and forth across the viscosity step. Whereas this phenomenon is neglected by the method discussed in that section, it is taken into account by the one presented here. It must however be noted that these terms are negligible in typical experimental setups, where $D_0/Lv_0 \sim 10^{-6}$ and $\tilde{\eta}_x < 10$. Indeed, in the limit where the diffusion coefficient and the step size are relatively small, Eqs. (43) and (44) respectively yield

$$\mathrm{mean}(T(0,L)) \simeq \frac{L}{v_0}[1 + (\tilde{\eta}_x - 1)y] \quad (45)$$

and

$$\mathrm{var}(T(0,L)) \simeq \frac{2D_0L}{v_0^3}[1 + (\tilde{\eta}_x^2 - 1)y] \quad (46)$$

which are identical to the results obtained in Section 2 for the mean elution time t_e and the temporal peak variance τ^2. However, the presence of extra terms in Eqs. (43) and (44) indicates that one should not represent a continuous viscosity gradient as a series of small steps using the results of Section 2. A full treatment of the problem is required.

3.3 Case 2: The exponential gradient with a perfect injection

The exponential viscosity gradient

$$\eta(x) = \eta_0 \exp\left(\frac{2kx}{L}\right) \quad (47)$$

where $L \equiv x'' - x'$ is the elution length and k is a dimensionless parameter, is another example for which exact expressions can be obtained for the moments of the first passage time. In the limit where $a \to -\infty$, reflections on the left wall are eliminated and the formalism presented in Section 3.1 leads to

$$\mathrm{mean}(T(0,L)) = \begin{cases} \dfrac{L}{2kv_0} \times \dfrac{e^{2k} - 1}{1 + kD_0/Lv_0} & \text{for } k > -\dfrac{Lv_0}{D_0} \\ \infty & \text{for } k < -\dfrac{Lv_0}{D_0} \end{cases} \quad (48)$$

Electrophoresis 2002, *23*, 1822–1832

and

$$\mathrm{var}(T(0,L)) = \begin{cases} \dfrac{LD_0}{2kv_0^3} \times \dfrac{e^{4k}-1}{(1+kD_0/Lv_0)^2(1+3kD_0/Lv_0)} & \text{for } k > -\dfrac{Lv_0}{3D_0} \\ \infty & \text{for } k < -\dfrac{Lv_0}{3D_0} \end{cases} \quad (49)$$

where v_0 and D_0 are both positive. It is interesting to note that, when $k > 0$, mean$(T(0, L))$ remains finite as v_0 approaches zero from positive values, in contrast with the uniform ($k = 0$) viscosity case where it becomes infinite. This is due to the fact that the molecule is repelled by the boundary located at $x = a = -\infty$ when $k > 0$ (see Appendix). Although this is an artifact, its impact on mean $(T(0, L))$ is negligible when $kD_0/Lv_0 \ll 1$, as can be seen by looking at the denominator of Eq. (48).

As in Section 2, we are interested in calculating the resolution factor for two molecular species with similar physical properties and, thus, closeby elution peaks. From Eq. (48), it is found that increasing v_0 by Δv_0, and D_0 by ΔD_0 decreases mean$(T(0,L))$ by

$$\Delta\,\mathrm{mean}(T(0,L)) = \frac{\Delta v_0 + k\Delta D_0/L}{v_0 + kD_0/L}\mathrm{mean}(T(0,L)) \quad (50)$$

to quadratic order in Δv_0 and ΔD_0. The resolution factor is thus given by

$$R = \frac{\Delta\,\mathrm{mean}(T(0,L))}{2\mathrm{var}(T(0,L))} = \frac{1}{2}\frac{\Delta v_0 + k\Delta D_0/L}{v_0 + kD_0/L}\sqrt{\frac{Lv_0(1+3kD_0/Lv_0)\tanh(k)}{2kD_0}} \quad (51)$$

to order $O(\Delta v_0^2, \Delta v_0\Delta D_0, \Delta D_0)$. Dividing this equation by the uniform viscosity limit

$$R_0 \equiv \lim_{k\to 0} R = \frac{\Delta v_0}{2}\sqrt{\frac{L}{2v_0D_0}} \quad (52)$$

and taking the $\Delta v_0, \Delta D_0 \to 0$ limit results in

$$\frac{R}{R_0} = \frac{1+(k/L)(dD_0/dv_0)}{1+kD_0/Lv_0}\sqrt{\frac{(1+3kD_0/Lv_0)\tanh(k)}{k}} \quad (53)$$

where $dD_0/dv_0 \equiv \lim_{\Delta v_0, \Delta D_0 \to 0} \Delta D_0/\Delta v_0$. Under normal circumstances, $kD_0/Lv_0 \ll 1$ and $dD_0/dv_0 \sim D_0/v_0$, thus leading to

$$\frac{R}{R_0} \simeq \sqrt{\frac{\tanh(k)}{k}} \quad (54)$$

As illustrated in Fig. 5a, $\sqrt{\tanh(k)/k}$ is maximum at $k = 0$. The best resolution for this system therefore occurs when the viscosity is uniform. Interestingly, the cases $\pm k$ give similar resolutions. In one case ($k > 0$), the peaks are sharper and closer to one another, while they are broader but further apart in the other case ($k < 0$). The end result is, in both cases, a poorer resolution.

As shown by Eqs. (48) and (49), both the mean and the variance of the first passage time are finite when $v_0 > 0$ and $k > -Lv_0/3D_0$. In this parameter range, the skewness is given by

$$\mathrm{skew}(T(0,L)) = 4\frac{\sinh(3k)}{[\sinh(2k)]^{3/2}}\frac{\sqrt{(kD_0/Lv_0)(1+3kD_0/Lv_0)}}{1+5kD_0/Lv_0} \quad (55)$$

except when $-Lv_0/3D_0 < k < -Lv_0/5D_0$, in which case it is infinite. The uniform viscosity limit of Eq. (55) is

$$\mathrm{skew}_0(T(0,L)) \equiv \lim_{k\to 0}\mathrm{skew}_0(T(0,L)) = 3\sqrt{\frac{2D_0}{Lv_0}} \quad (56)$$

Note that $\mathrm{skew}_0(T(0, L)) > 0$ in the finishline mode, in contrast with the spatial skewness of the snapshot mode, because the distribution of analytes continues to widen as it moves in front of the detector [10]. This fact is rarely mentioned in the literature, because $\mathrm{skew}_0(T(0, L))$ is usually too small to be measured. Assuming that $kD_0/Lv_0 \ll 1$, the ratio of Eqs. (55) and (56) yields

$$\frac{\mathrm{skew}\left(T(0,L)\right)}{\mathrm{skew}_0\left(T(0,L)\right)} = \frac{2}{3}\frac{\sinh(3k)}{[\sinh(2k)]^{3/2}}\frac{\sqrt{2k(1+3kD_0/Lv_0)}}{1+5kD_0/Lv_0} \simeq \frac{2}{3}\frac{\sqrt{2k}\sinh(3k)}{[\sinh(2k)]^{3/2}} \quad (57)$$

which is illustrated in Fig. 5b. Experimentally, the viscosity of the medium in which the separation takes place may vary at most by about an order of magnitude. Thus, $k \lesssim 1.15$ and $\mathrm{skew}(T(0, L))/\mathrm{skew}_0(T(0, L)) \lesssim 1.45$. Although the skewness of the first passage time is moderately increased by the presence of the viscosity gradient compared to the uniform viscosity case, it remains of the same order of magnitude and thus remains quite small. For instance, if $v \sim 10^{-2}$ cm/s, $D \sim 10^{-7}$ cm^2/s and $L \sim 10$ cm, then $\mathrm{skew}_0(T(0, L)) \sim 4 \times 10^{-3}$ and $\mathrm{skew}(T(0, L)) \lesssim 6 \times 10^{-3}$. This is in stark contrast with the spatial distribution of molecules, which can be severely skewed by the presence of a viscosity gradient. Indeed, as the molecules get closer to the detector in a monotically increasing viscosity gradient, their drift and diffusion slow down because of the increased viscosity. This leads to a skewed spatial distribution of molecules, with the longer tail pointing away from the detector. However, since the molecules that are located in this tail are those that drift the most rapidly, they tend to catch-up with those that are closer to the detector. In the end, this significantly reduces the skewness of the first passage times as the molecules pass in front of the detector.

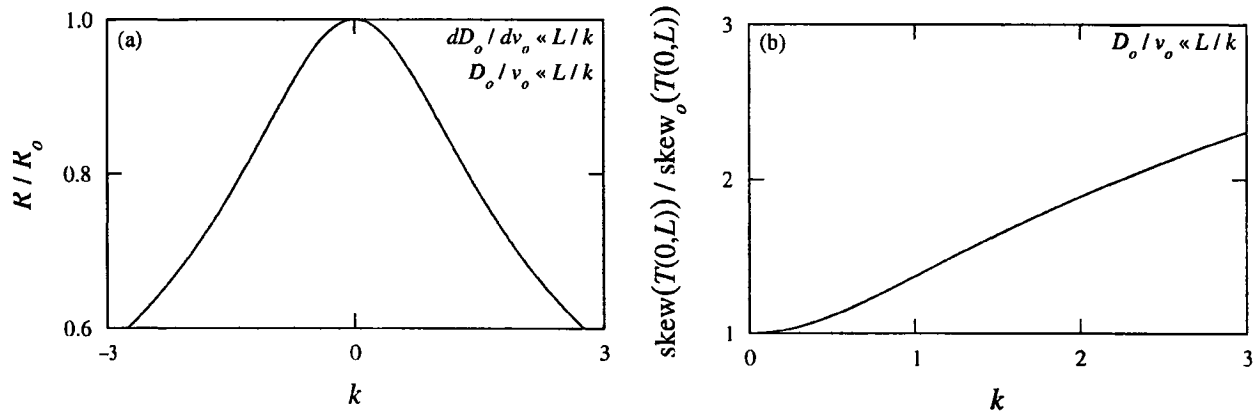


Figure 5. (a) Resolution factor R and (b) skewness of the mean first passage time skew$(T(x', x''))$ for the exponential viscosity gradient and a negligibly small injection width. The ratios of these quantities with their uniform viscosity limits R_0 and skew$_0(T(x', x''))$, as given by Eqs. (54) and (57), are plotted here as a function of the parameter k, which determines the steepness of the viscosity gradient (see Eq. 47). The best resolution occurs when the viscosity is uniform and gradually degrades as $|k|$ increases. On the other hand, the skewness is significantly increased by the presence of an exponential viscosity gradient, but remains of the same order of magnitude as in the uniform viscosity case.

3.4 Case 3: The exponential gradient with a nonzero injection width

In this case, we assume the analytes to be uniformly distributed by the injection. This "plug" initial distribution of analytes is described by the probability density

$$p_0(x) = \begin{cases} 1/l \text{ for } -l/2 \le x \le l/2 \\ 0 \quad \text{otherwise} \end{cases} \tag{58}$$

where l is the width of the injection band. Repeating the analysis performed in Section 3.3 now leads to

$$\text{mean}(T(L)) = \frac{L}{2kv_0}\frac{1}{1+kD_0/Lv_0}\left(e^{2k} - \frac{\sinh(kl/L)}{kl/L}\right) \tag{59}$$

and

$$\text{var}(T(L)) = \frac{LD_0}{2kv_0^3}\frac{1}{(1+kD_0/Lv_0)^2(1+3kD_0/Lv_0)} \times \times\left\{e^{4k} - \frac{Lv_0}{2kD_0}\left[\left(1+\frac{3kD_0}{Lv_0}\right)\left(\frac{\sinh(kl/L)}{kl/L}\right)^2 - \left(1+\frac{kD_0}{Lv_0}\right)\left(\frac{\sinh(2kl/L)}{2kl/L}\right)\right]\right\} \tag{60}$$

where Eq. (42) has been used to average the moments over the initial condition. From these two equations, it is found that

$$\frac{R}{R_0} = \frac{1+(k/L)(dD_0/dv_0)}{1+kD_0/Lv_0} \times \times\sqrt{\frac{(1+3kD_0/Lv_0)(1+l^2v_0/24LD_0)}{k}} \times \sqrt{\frac{\left(e^{2k} - \frac{\sinh(kl/L)}{kl/L}\right)^2}{e^{4k} - \frac{Lv_0}{2kD_0}\left[\left(1+\frac{3kD_0}{Lv_0}\right)\left(\frac{\sinh(kl/L)}{kl/L}\right)^2 - \left(1+\frac{kD_0}{Lv_0}\right)\left(\frac{\sinh(2kl/L)}{2kl/L}\right)\right]}} \tag{61}$$

In the limit where $l \to 0$, Eqs. (59) to (61) appropriately yield Eqs. (48), (49) and (53).

As shown in Fig. 6a, where Eq. (61) is plotted for a sample set of parameter values, the exponential increase in viscosity can improve the resolution when the injection width is non-zero. This arises because, similarly to the step gradient case (see Section 2.3), the exponential viscosity gradient performs an additional, though gradual, stacking of the initial peak. However, numerical comparisons of the resolution factors for both gradients have shown that the step gradient with $y = 1$ leads to systematically higher resolutions than the exponential viscosity gradient (see Fig. 6b). The best resolutions are thus once again obtained when the increase in viscosity is entirely localized at the injection point.

Electrophoresis 2002, *23*, 1822–1832

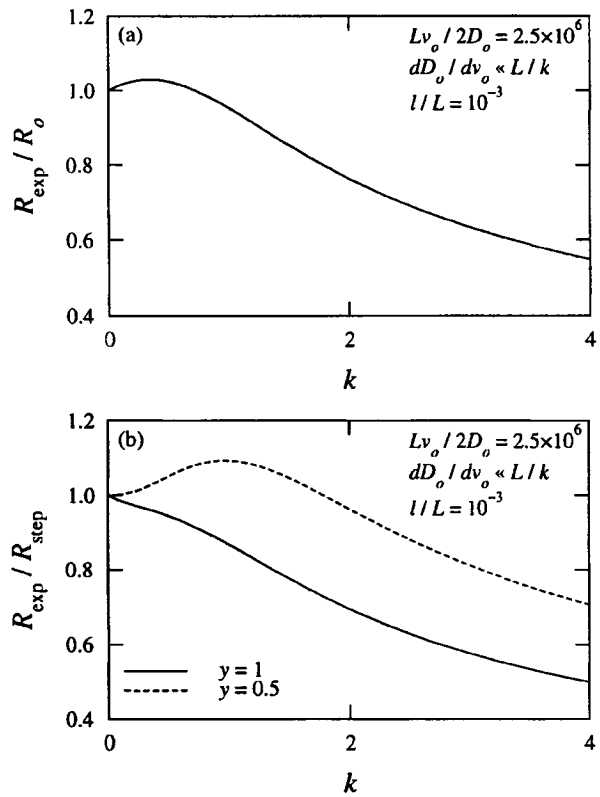


Figure 6. Resolution factor R for the exponential viscosity gradient with a non-negligible injection width. In graph (a), R is divided by its uniform viscosity limit R_0 (see Eq. 61), and in (b), by the resolution factor for the step gradient with the same total variation in viscosity (*i.e.*, with $\tilde{\eta}_X = e^{2k}$). Both the $y = 1$ (step located at the injection point) and $y = 0.5$ (step located halfway between the injection point and the detector) cases are drawn in graph (b). For sufficiently small values of k, the exponential viscosity gradient leads to a better resolution than the uniform viscosity case and the step gradient with $y = 0.5$. However, the step gradient with $y = 1$ still yields the best resolution. (In order to distinguish between the resolution factors for the step and exponential gradients, they are respectively called R_{step} and R_{exp} for this figure only.)

4 Discussion

Qualitatively, the results presented here demonstrate very clearly that one cannot use static gradients to get around the limits imposed by thermodynamics in the case of finishline separation systems. When the final peak width is controlled by diffusion processes, static viscosity gradients cannot improve the resolution. The same conclusion applies to both step and continuous gradients. When the injection width is nonnegligible, a viscosity gradient does help, but the same (and even better results) can be obtained if the extra viscosity is used to further stack the initial zone right at the beginning of the separation. Therefore, there is little point in placing the gradient elsewhere along the path. It is well-known that snapshot systems behave differently; unfortunately, the gradients that are known to improve the performance of such systems cannot be automatically transferred to their finishline cousins.

Furthermore, our analysis revealed an unexpected result: when both the injection width and the diffusion play an important role in determining the final peak width, one should be careful when using a step gradient since placing it too close to the detector will actually reduce the resolution in spite of the stacking that it generates. It is also probable that some continuous gradients may also reduce the resolution if they are placed too close to the detector. In all cases, a better strategy is to locate all sources of stacking as close to the injection position as possible.

An important limiting assumption of our analysis is that both the electrophoretic velocity and the diffusion coefficient are linear (and identical) functions of the local physical property that varies in space. This condition may be violated, for instance, when the conformational properties of the molecules (if they have internal entropy, such as DNA molecules or denatured proteins) are affected by the local conditions and do not adapt instantaneously to them. In such a case, the situation can be quite different from what is depicted in this article.

Nonlinear systems can be exploited in different ways to improve the performance of the system, although not always in a finishline mode. Again, the main problem is that all the molecules visit the exact same path before reaching the detector region. However, if the velocity and the diffusion coefficient are not proportional to each other (*i.e.*, if the velocity and the diffusion coefficient are different nonlinear functions of the local conditions), one can in principle design separation processes that may provide the user with results not normally achievable. In a future article, we will look at the case where the analyte is a long, reptating DNA molecule being electrophoresed in a polymer matrix that possesses a concentration gradient. Both the electrophoretic mobility and the diffusion coefficients are known to be very nonlinear functions of the gel concentration and of the electric field intensity in such systems [1]. Moreover, the presence of the reptation tube means that one can actually exploit hysteretic phenomena where the molecules remember which path they have visited in the recent past [11]. The use of gel concentration gradients during CE of DNA has recently been explored experimentally by Chu *et al.* [8]. Serwer *et al.* [12] have also shown some spectacular results with novel separation ideas entirely based on nonlinear effects in gels.

Nonequilibrium or nonadiabatic systems, where the conformational degrees of freedom of the molecule absorb energy by responding to the spatial or temporal gradients, represent a more interesting situation. Since they do not follow the laws of equilibrium thermodynamics, nothing forbids such systems from improving resolution. This is certainly the case with pulsed-field electrophoresis [3]. Two questions thus come to mind: (i) Can we have a spatial equivalent to pulsed field electrophoresis where the field would vary in space instead of time? (ii) Can varying or pulsed fields (spatial or temporal) improve the resolution of finishline systems? The answer is possibly yes to both questions, but these two topics remain largely unexplored, both experimentally and theoretically.

The method presented here can be useful for a large number of problems, but exact analytical solutions are probably unlikely in most cases. Fortunately enough, it is relatively easy to study first passage time problems numerically. The main problem, however, is that any theoretical study of systems with gradients requires a knowledge of the dependence of the velocity and of the diffusion coefficient upon the different experimental variables, and especially the variable that varies in space and those that are coupled to it. This empirical knowledge is not always available or may be too partial to be useful in practice.

The authors would like to thank Drs. Bernard Tinland and Ben Chu for useful discussions. This work was supported in part by a Research Grant from the Natural Science and Engineering Research Council (NSERC) of Canada to GWS, by an NSERC scholarship to LCM and by an NSERC postdoctoral fellowship to SG.

Received December 10, 2001

5 References

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

[2] Slater, G. W., Desruisseaux, C., Hubert, S. J., Mercier, J.-F., Labrie, J., Boileau, J., Tessier, F., Pépin, M. P., *Electrophoresis* 2000, *21*, 3873–3887.

[3] Burmeister, M., Ulanovsky, L., (Eds.), *Pulsed-Field Gel Electrophoresis: Protocols, Methods, and Theories*, Humana Press, Totowa, NJ 1992.

[4] Slater, G. W., Noolandi, J., *Electrophoresis* 1988, *9*, 643–646.

[5] Izzo, V., Craxi, A., Barbieri, R., *Electrophoresis* 2001, *22*, 29–32.

[6] Chacron, M. J., Slater, G. W., *Phys. Rev. E* 1997, *56*, 3446–3450.

[7] Frumin, L. L., Peltek, S. E., Zilberstein, G. V., *Phys. Rev. E* 2001, *64*, 021902.

[8] Liang, D., Song, L., Quesada, M. A., Tian, Z., Studier, F. W., Chu, B., *Electrophoresis* 2000, *21*, 3600–3608.

[9] Gardiner, C. W., *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer-Verlag, New York 1997.

[10] Slater, G. W., in: Heller, C. (Ed.), *Analysis of Nucleic Acids by Capillary Electrophoresis*, Vieweg, Wiesbaden, Germany 1997, Chapter 2.

[11] Mayer, P., Slater, G. W., Drouin, G., *Electrophoresis* 1994, *15*, 120–127.

[12] Griess, G. A., Rogers, E., Serwer, P., *Electrophoresis* 2000, *21*, 859–864.

6 Appendix: Diffusion in the absence of an electric field

In the absence of an electric field, the Stratonovich SDE describing the evolution of the molecule (Eq. 26) reduces to

$$dx = \sqrt{\frac{2D_0}{\tilde{\eta}(x)}} \circ dW \tag{A1}$$

This equation defines, up to an arbitrary integration constant, a simple change of variable that maps the trajectory of a Wiener process into the trajectory of the diffusing molecule. The evolution of the molecule thus simply corresponds to a Wiener process occurring in a distorted space. In regions of high viscosities, the trajectory of the Wiener process is contracted; in regions of low viscosities, it is dilated.

In the case of the exponential viscosity gradient studied in Section 3, Eq. (A1) yields

$$W = \frac{(e^{kx/L} - 1)L}{k\sqrt{2D}} \tag{A2}$$

once $x = 0$ has been arbitrarily identified with $W = 0$. Using this equation, it is found that the $(-\infty, \infty)$ range in x-space maps to $(-\sqrt{L^2/2k^2D}, \infty)$ in W-space when $k > 0$. In fact, the viscosity approaches zero so rapidly as $x \to -\infty$ that the molecule can feel the presence of a boundary at $x = -\infty$. This must clearly be an artifact, since the viscosity of the medium in which the electrophoresis takes place is never strictly zero in any experimental setup. Under normal circumstances, however, an electric field is applied in such a way that $Lv_0/kD_0 \gg 1$ and, as mentioned in Section 3, the molecule does not feel the presence of this artificial boundary.

8

Conclusion

This thesis contributes to our understanding of free solution electrophoresis of polyelectrolytes, and in particular, the method of End Labelled Free Solution Electrophoresis (ELFSE) for the sequencing of DNA and the analogous method of Free Solution Conjugate Electrophoresis (FSCE) for the characterization of uncharged polymers. The free-draining nature of uniformly charged polyelectrolytes undergoing free solution electrophoresis, where the electric force and the coefficient of friction both scale linearly with the length of the polymer, is a phenomenon that normally leads to co-migration of all lengths of polyelectrolytes in free solution. Breaking of the charge-to-friction symmetry via the conjugation of charged and uncharged molecular species allows for separation in free solution.

Free solution electrophoresis has numerous advantages over gel-based techniques including increased speed and decreased consumption of chemicals, some of which are toxic. The loading of gels and other sieving media into small-diameter capillaries must be done slowly to avoid shear breakage, and once each matrix is loaded it must be left for several minutes or longer in order to become homogeneous. Capillary cleaning is also a time consuming step when using sieving matrices. These problems will only be exacerbated in the future as DNA sequencing moves from capillaries to miniature devices where the narrower channels will make gel loading even more challenging, if not impossible. Free solution electrophoresis, by avoiding the use of gels or other sieving media, has the major advantage of being amenable to miniaturization.

Capillary electrophoresis is responsible for the success of the global effort to sequence the complete human genome, one of the most remarkable achievements of modern science [1, 2]. This effort has been described as a *race to the starting line* for understanding biology at a deeper level. The race from the starting line, however, is not facilitated by today's most advanced systems, which are too slow and too costly. To encourage the research and development necessary to overcome these limitations, the American National Institute of Health (NIH) created a $38 million (US) program in 2004 to fund short and long term research projects, with the aim of enabling the sequencing of individual genomes as part of medical care. According to an NIH News Release [3], the "ability to sequence each person's genome cost-effectively could give rise to more individualized strategies for diagnosing, treating and preventing disease." The report goes on to say that "such information could enable doctors to tailor therapies to each person's unique genetic profile." Use of ineffective medication is expensive, and a bigger problem lies in the use of medication that can be potentially harmful or even fatal for a person of a specific genotype. In view of the myriad of medications available, it is very important to use them as safely and effectively as possible. Quick and affordable clinical DNA testing when needed may present one means of doing so. With its amenability to the miniaturization trend, and its increased speed and decreased cost, ELFSE-based sequencing holds much promise for clinical application.

Before the benefits of ELFSE can be taken advantage of, the biggest challenge to improving ELFSE performance must be overcome. Very long DNA chains have so much charge that their speed is barely affected by a small label and hence the peaks corresponding to longer DNA chains overlap badly and resolution is lost, thereby limiting the maximum size of DNA that can be sequenced. Extending this so-called read length of ELFSE is critical to enable it to be competitive with gel-based techniques; this is the main goal of current ELFSE research. My supervisor, Gary Slater, along with his collaborator Annelise Barron of Northwestern University, head the collaborative group that is the world leader in ELFSE research and development. My work has made important contributions to the efforts of the group. So far, the main approach to extending the read length of ELFSE has been to try and find a larger label, which would result in a greater force of friction, and therefore a bigger effect on the speed of currently unresolved long DNA, increasing their resolution and thereby extending the read length; however, as discussed in Appendix A, it has been difficult to create a large, monodisperse label. Several novel means of extending the read length that do not depend on a larger label were presented in this thesis to help overcome this obstacle.

One means of extending the read length that does not depend on a bigger label relies on labelling both ends rather than simply increasing the size of a single end label. Before the paper presented in Chapter 2, it had not been recognized that labelling both ends would lead to a greater friction, and therefore greater resolving power, than if the size of the label at one end were doubled. This is a result of the end effect predicted by the theory of Long and co-workers [4] for polyelectrolytes; our detailed analysis of this effect on ELFSE illuminated the potential of labelling both ends of the DNA instead of just one. This discovery which was later confirmed for ELFSE in the experimental paper provided in Appendix B, led to Canadian and US patent applications. The experimental paper provided proof of concept; for widespread use however, this technique requires an easy, reproducible means of attaching the label to both ends of all ssDNA strands in a sequencing sample. Work into developing the necessary chemistry is ongoing. A possible avenue for future study would be to investigate positively charged labels. If the neutral label molecule were to be replaced by one with a slight positive charge, although not so much to allow for a collapse of the random coil due to electrostatic interactions between the negative charges of the DNA and the positive charges of the label, then the DNA would be slowed down even more. This would allow the DNA strands more time to separate, thereby increasing the read length. As seen in Chapter 2, however, the location of the positive charges would be expected to have an effect on the mobility. Investigating the effect of the presence, and location, of positive charges on an ELFSE label, borrowing from the theory of Chapter 2, would be an exciting future study. It is important to note that the study of the end effects also provided critical amendments to the theory of FSCE. In the paper, potential errors in the previous method, where the end effects were neglected, were uncovered and more accurate methods for determining the molar mass profile were suggested.

Another means of extending the read length of ELFSE, through use of a controlled electroosmotic flow, was also presented in this thesis (Chapter 3). This flow is due to the counter-ions in the solution filling the capillary, which are attracted to the negative charges of the capillary wall surface. Under the influence of an electric field these charges, which are not bound to the wall as the negative surface charges are, move in the opposite direction to negatively charged DNA. Due to viscous drag forces, the bulk of the solution filling the capillary is dragged along with them in a plug-like flow. If the capillary is uniformly charged and both ends are at the same pressure, this counter-flow (which can drag DNA along with it) is constant [5, 6]. This flow is commonly prevented by coating the capillary walls, however, in Chapter 3 we show how allowing a well-controlled amount of EOF could positively impact on ELFSE performance. It is predicted that this

process could dramatically extend the read length, as well as decrease the time required for sequencing. The resolution of EOF-based ELFSE, however, is very sensitive to the exact EOF value; small changes in the EOF rate can decrease the resolution beyond that of the traditional, EOF-free case. Hence, the capillary coating giving the desired EOF rate must be tested for reliability. A possible extension to this work is to re-examine the predictions for improving ELFSE with EOF, taking the end effect into account. This would adjust the mobility of the conjugates and hence could change the predicted range of EOF that provides better resolution. Since the resolution is sensitive to the EOF rate, the results of this extension, which would effectively combine the work of Chapters 2 and 3, would be quite useful to experimentalists. Another possible extension would be to derive the optimal range of EOF for a positively charged label, taking the end effects into account. Using a positively charged label on both ends, with an optimized EOF rate may be the winning combination to substantially extending the read length.

Potential changes in ELFSE behaviour with next-generation labels and/or high fields were also investigated (Chapter 4). Predictions of the conditions under which segregation of the DNA and label coils may occur, and the resulting impact on ELFSE performance were given. This most recent study indicated that a label providing segregation from DNA via steric repulsion, which had recently been considered a desirable goal, would actually lead to decreased resolution unless used with very high voltages. Also, this work highlighted the potential gains in read length possible with high voltage use. These kinds of findings are critical in guiding the work of ELFSE experimentalists. Since segregation may be followed by stretching of the DNA and/or a deformable label from their equilibrium random coil conformations, a necessary continuation of this study is to investigate the conditions leading to stretching and the resulting impact on resolution, developing predictions for optimal label architecture design and experimental setup. Another interesting future investigation would involve the effect of a positively charged label on deformation of the ELFSE conjugate. The extra force pulling backwards on the DNA due to a positively charged label would also add to the force acting to deform the conjugate. This would be expected to decrease the field (or label and DNA sizes) necessary for deformation; it would be of great interest to have predictions for the optimal conditions for a positively charged label. The theory presented in Chapter 4 would be a good basis for looking at this. It would also be possible to predict the effect of EOF on the resolution of segregated conjugates by using the framework constructed in Chapter 3.

In the future, the long read lengths of ELFSE separations will mean fast analysis times. This is clearly an advantage of the technology, however, the speed of the analytes past the detector

may pose a problem, depending on the detector response time. The effect of next-generation labels and higher field strengths on the analyte speed can be analysed so as to predict the detector requirements. Perhaps these requirements could be lessened by some compensative means such as slowing the analytes down just before detection, for example, by a local field decrease, or a local viscosity increase. This latter idea would need to be assessed for any impact on resolution, making use of the results presented in Chapter 7.

Some more general advancements to the theory of free solution electrophoresis of polyelectrolytes were also developed in this thesis. It was demonstrated that the Nernst-Einstein relation that is commonly used for free solution electrophoresis, actually does not apply under these conditions. This is due to subtle differences in the processes of electrophoretic mobility and diffusion. While for the latter motion, there is a build up of hydrodynamic effects that are unaffected by the electric field, the former is crucially affected by the response of the counter-ions to the electric field, which effectively cancel the hydrodynamic interactions; i.e. the polyelectrolyte coil is free-draining for mobility but not for diffusion. The proper equation for the diffusion coefficient for free solution electrophoresis was given. The findings of this paper were critical for the analysis and theory of free solution electrophoresis, including the theoretical predictions of Chapters 2 through 5.

It was also shown that a gradient in viscosity along the length of the capillary can not be used to improve the resolution in free solution electrophoresis, and hence could not be used to improve the resolution of ELFSE or FSCE. However, under non-ideal conditions, when the initial band loading width is large, a viscosity jump located at the start of the capillary may be used (in addition to a jump in solution conductivity) to increase sample stacking. This theoretical study provided valuable insight into what research paths are expected to be fruitful.

This thesis provided valuable advancements in general to the theory of free solution electrophoresis, and more specifically, to the ELFSE and FSCE techniques. It is hoped that these advancements will play a role in improving free solution capillary electrophoresis performance, helping to guide future work. This future work may be successful in enabling an ELFSE-based, cost-effective clinical tool that could hopefully be used with prudence to improve medical care.

References

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

[2] JC Venter et al. ***Science*** **291**, 1304–1351 (2001).

[3] *NIH News Release, Oct. 14, available at www.genome.gov/12513210.*

[4] D Long, AV Dobrynin, M Rubinstein, A Ajdari. ***Journal of Chemical Physics*** **108**, 1234–1244 (1998).

[5] S Ghosal. ***Electrophoresis*** **25**, 214–228 (2004).

[6] D Sinton, C Escobedo-Canseco, L Ren, D Li. ***Journal of Colloid and Interface Science*** **254**, 184–189 (2002).

Appendix A

End-Labeled Free-Solution Electrophoresis of DNA (A Review)

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

This invited review article covers both ELFSE and FSCE technology, from a theoretical and experimental standpoint. It is the most comprehensive introduction to ELFSE available.

Note that our calculation of the critical electric field for segregation to occur in subsection 3.1 was a rough one; a more detailed estimate is presented in Chapter 4.

This review article was written in collaboration with Annelise Barron's experimental group at Northwestern University. The Ottawa group collaborated on the theoretical segments; I made contributions to sections 3.1, 3.2, 5.1 and 6.3.

Electrophoresis 2005, *26*, 331–350

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.

Electrophoresis 2005, *26*, 331–350

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:

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

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.

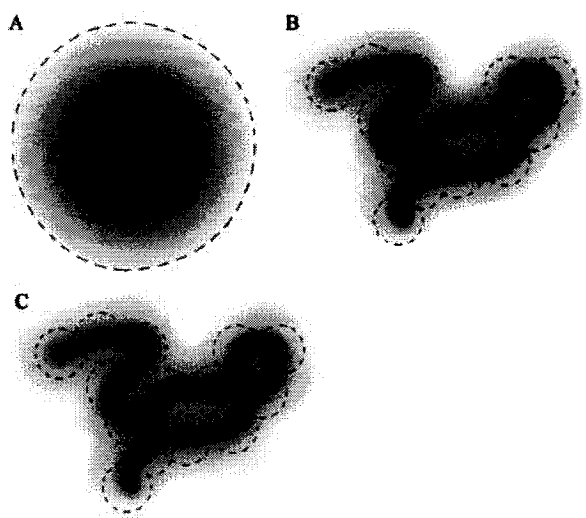


Figure 1. The three regimes of free-solution DNA electrophoresis (the DNA molecule is surrounded by a cloud of counterions). (A) The radius of gyration R_g of the DNA coil is smaller than the Debye length λ_D describing the extent of the cloud. In this case, the DNA is not free-draining, and we can separate 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:

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

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 $\alpha = 100$ and $\alpha = 250$ 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

Electrophoresis 2005, *26*, 331–350

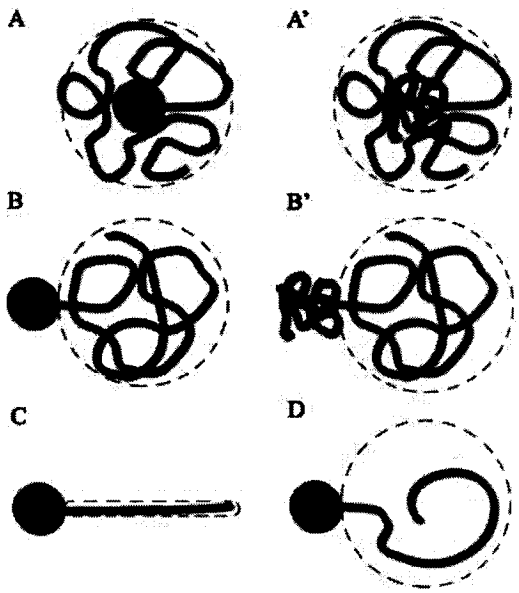


Figure 2. Schematic diagram showing various ELFSE regimes for a spherical drag-tag (A, B, C, D) and an extended polymeric drag-tag (A' + B'); the DNA is shown in red and the label in black. (A+A') The drag-tag is part of the random coil conformation of the DNA molecule. (B+B') The drag-tag and the DNA 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 $\mu = 0$ 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:

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

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

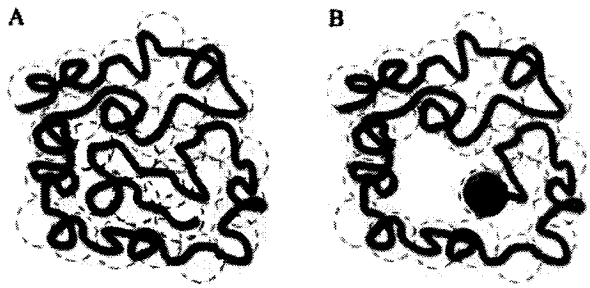


Figure 3. Schematic diagram showing the blob theory as applied to a ssDNA molecule (in red) attached to an extended polymeric label (A) or a spherical label (B). In both cases, we form blobs having identical hydrodynamic radii. The hybrid molecule is then seen as a chain of blobs; the latter thus play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

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

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-

Electrophoresis 2005, *26*, 331–350

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:

$$t(M_c) - t(M_c + 1) \cong -\frac{\partial t(M_c)}{\partial M_c} = \frac{L}{\mu_0 E}\frac{\alpha_1 M_u}{M_c^2} \tag{25}$$

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_0 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:

$$R = R_0 \frac{M_c^{1/2}(M_c + \alpha_1 M_u)^{5/4}}{\alpha_1 M_u} \tag{30}$$

where $R_0 = 4(\ln(2)D_1/\mu_0 EL)^{1/2}$ 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, $\mu_0 = 3 \times 10^{-4}$ 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.

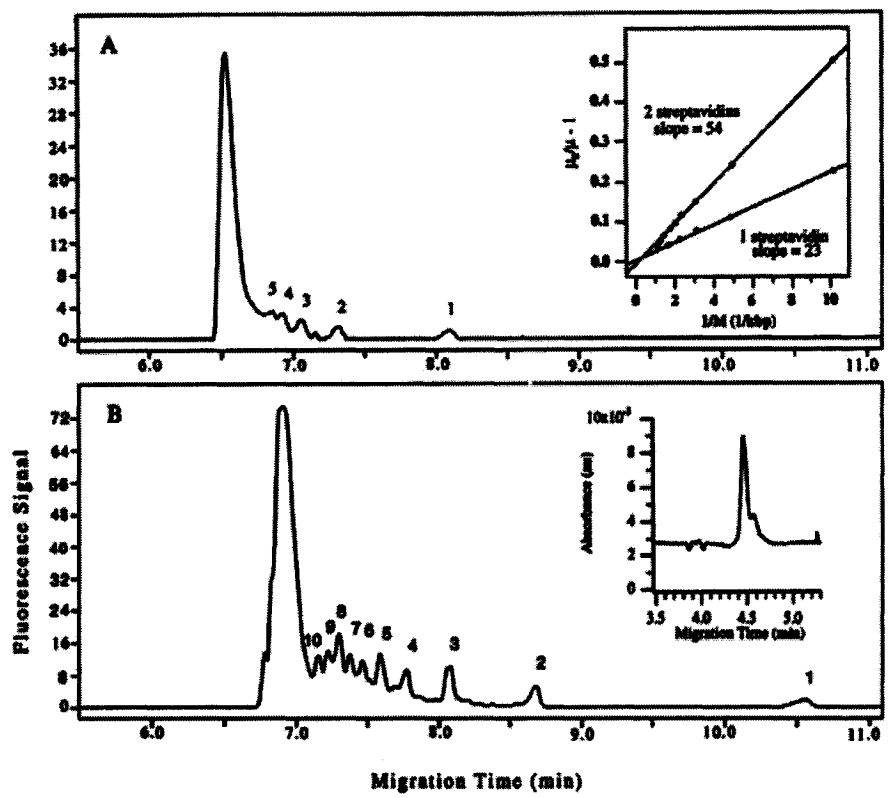


Figure 4. Separation of a 100 bp dsDNA ladder with (A) one or (B) two streptavidin molecules used as ELFSE drag-tags. The peaks marked 1–10 represent the M = 100 to M = 1000 bp dsDNA fragments, respectively. Insert of A: plot of $\mu_0/\mu-1$ *vs.* $1/M$. Insert of B: polydispersity of the streptavidin as measured by CE. Reprinted from [8], with permission.

Electrophoresis 2005, *26*, 331–350

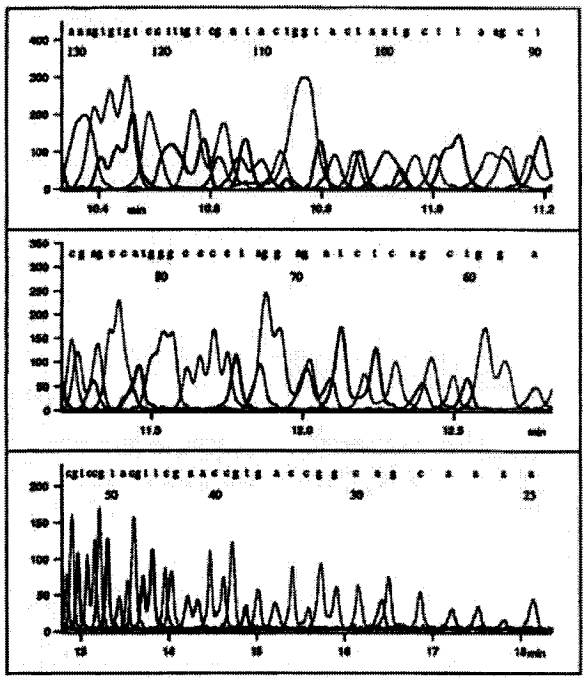


Figure 5. Separation of DNA sequencing reaction products using a purified streptavidin drag-tag and a dynamically coated capillary. The *Y*-axis represents the intensity of the fluorescence signal and the *X*-axis gives the elution time in minutes. The numbers above the peaks give the number of nucleotides (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.

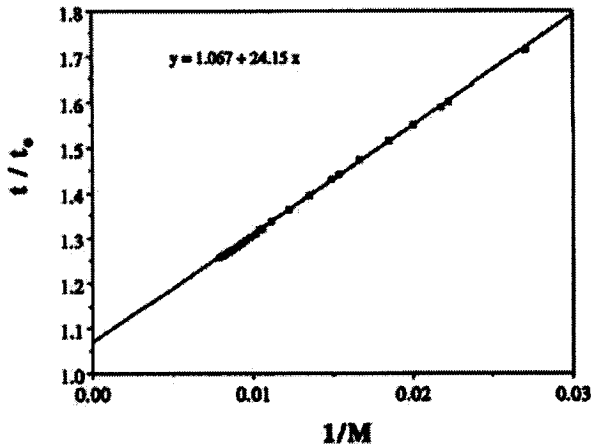


Figure 6. Ratio $t(M)/t_0$ *vs.* inverse ssDNA molecular size $1/M$, for the electropherogram presented in Fig. 5. Here, $t(M)$ is the elution time of the labeled ssDNA molecule of size M (in bases), while t_0 is the elution time of the unlabeled ssDNA molecules. The straight-line fit, predicted by theory, 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

Electrophoresis 2005, *26*, 331–350

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.

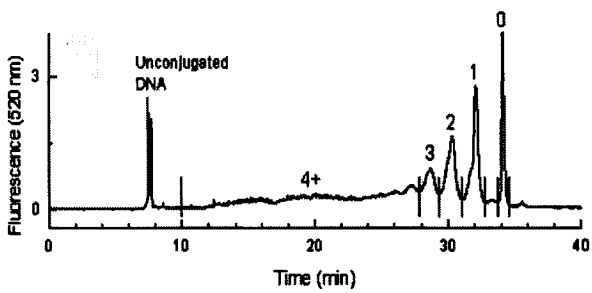


Figure 7. Capillary electrophoretic analysis of a 22-base fluorescently labeled DNA conjugated to a 337 amino acid protein polymer with the repeating "P1" sequence $(GAGQGSA)_{48}G$. The numbered peaks refer to conjugate species with 0, 1, 2, 3, 4 or more deamidated glutamines; this polydispersity renders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

Electrophoresis 2005, *26*, 331–350

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:

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

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

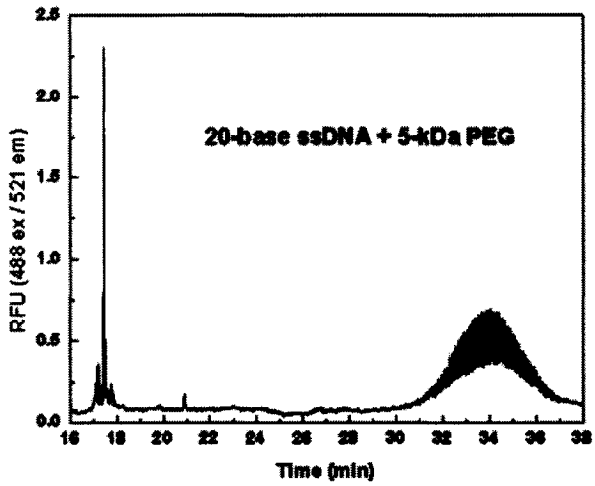


Figure 8. FSCE separation of a monodisperse, fluorescently labeled, 20-base ssDNA "engine" conjugated to a PEG with nominal molar mass of 5 kDa and a polydispersity index of 1.01. The series of peaks between 30 and 37 min represent PEG chains with different numbers of monomer units, conjugated to the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

Appendix B

Experimental Confirmation of the End Effect in ELFSE

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

This paper presents the experimental confirmation of the end effects in ELFSE, predicted in the theoretical paper presented in Chapter 2. Robert Meagher of Northwestern University was primarily responsible for the experiments undertaken by the Northwestern University group led by Annelise Barron. Gary Slater and myself provided assistance with the data analysis.

1702 *Electrophoresis* 2006, *27*, 1702–1712

Robert J. Meagher[1]
Laurette C. McCormick[2]
Russell D. Haynes[1,3]
Jong-In Won[1*]
Jennifer S. Lin[1]
Gary W. Slater[2]
Annelise E. Barron[1,3]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA
[2]Département de Physique, Université d'Ottawa, Ottawa, ON, Canada
[3]Department of Chemistry, Northwestern University, Evanston, IL, USA

Received July 28, 2005
Revised November 7, 2005
Accepted November 8, 2005

Research Article

Free-solution electrophoresis of DNA modified with drag-tags at both ends

In end-labeled free-solution electrophoresis (ELFSE), DNA molecules are labeled with a frictional modifier or "drag-tag", allowing their size-based electrophoretic separation in free solution. Among the interesting observations from early work with dsDNA using streptavidin as a drag-tag was that the drag induced by including a streptavidin label at both ends was significantly more than double that from a single streptavidin (Heller, C. *et al.*, *J. Chromatogr. A* 1998, *806*, 113–121). This finding was assumed to be in error, and subsequent work focused on experiments in which only a single drag-tag is appended to one end of the DNA molecule. Recent theoretical work (McCormick, L. C., Slater, G. W., *Electrophoresis* 2005, *26*, 1659–1667) has examined the contribution of end-effects to the free-solution electrophoretic mobility of charged-uncharged polymer conjugates, reopening the question of enhanced drag from placing a drag-tag at both ends. In this study, this effect is investigated experimentally, using custom-synthesized ssDNA oligonucleotides allowing the attachment of drag-tags to one or both ends, as well as dsDNA PCR products generated with primers appropriate for the attachment of drag-tags at one or both ends. A range of sizes of drag-tags are used, including synthetic polypeptoid drag-tags as well as genetically engineered protein polymer drag-tags. The enhanced drag arising from labeling both ends has been confirmed, with 6–9% additional drag for the ssDNA and 10–23% additional drag for the dsDNA arising from labeling both ends than would be expected from simply doubling the size of the drag-tag at one end. The experimental results for ssDNA labeled at both ends are compared to the predictions of the recent theory of end-effects, with reasonably good quantitative agreement. These experimental findings demonstrate the feasibility of enhancing ELFSE separations by labeling both ends of the DNA molecule, leading to greater resolving power and a wider range of applications for this technique.

Keywords: DNA sequencing / Drag-tags / End-labeled free-solution electrophoresis / Genotyping / Molecular end-effect DOI 10.1002/elps.200500554

1 Introduction

1.1 General

Size-based separations of DNA for applications such as DNA sequencing and genotyping are frequently accomplished by electrophoresis in a polymeric sieving matrix, examples of which include crosslinked gels and highly entangled solutions of linear polymers [1]. Although this technique is a workhorse of modern molecular biology, the sieving matrix imposes limitations on the speed of separation, and electric field-induced band-broadening and molecular orientation effects lead to a reduced ability to separate larger DNA fragments [1–4]. Additionally, crosslinked gels and viscous polymer solutions are problematic to load into miniaturized microfluidic devices currently being developed for DNA sequencing, PCR product sizing, and other electrophoretic separations [5–9].

A variety of alternative DNA separation modes have been proposed for use in capillaries and microfluidic devices [10], including entropic trapping [11, 12], separation in ultradilute polymer solutions [13] or in microfabricated

Correspondence: Professor Annelise E. Barron, Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road – Tech E136, Evanston, IL 60208, USA
E-mail: a-barron@northwestern.edu
Fax: +1-847-491-3728

Abbreviations: ELFSE, end-labeled free-solution electrophoresis; **NMEG**, *N*-methoxyethylglycine; **Sulfo-SMCC**, sulfosuccinimidyl 4-*N*-maleimidomethyl cyclohexane-1-carboxylate; **TCEP**, tris(2-carboxyethylphosphine)

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

Electrophoresis 2006, *27*, 1702–1712

arrays of posts or other obstacles [14, 15]. One exciting approach that has received considerable attention is end-labeled free-solution electrophoresis (ELFSE) [16–20]. In this approach, DNA is modified end-on with an uncharged, monodisperse, polymeric end-label, or "drag-tag" to create a charged-uncharged polymer conjugate. During electrophoresis in free solution, the drag-tag imparts the bioconjugate with a fixed amount of additional hydrodynamic friction. The additional friction modifies the electrophoretic mobility of the DNA-drag-tag conjugates in a size-dependent fashion: Conjugates comprising small DNA fragments migrate more slowly than conjugates with large DNA fragments, and thus a size-based separation can be accomplished in the absence of a sieving matrix.

The theoretical principles and experimental demonstrations of ELFSE have been recently reviewed [20]. In the first experimental demonstration of ELFSE, streptavidin was used to label dsDNA restriction fragments that had been biotinylated at one or both ends [18]. The efficiency of this separation was limited primarily by the inherent polydispersity of the streptavidin label, as well as by interactions between the streptavidin and the capillary walls. One of the interesting results of this study, however, was that the amount of hydrodynamic drag associated with adding a streptavidin label to both ends of the DNA was observed to be significantly more than twice the friction for adding streptavidin to one end only. Whereas a single streptavidin provided friction equivalent to an additional 23 bp of DNA, two streptavidins provided the friction of an additional 54 bp, 17% greater than would be expected from simply doubling the amount of friction from a single streptavidin. The implications of this finding were not fully appreciated at the time, and, being attributed to experimental error, this effect was not explored further.

In later work, a gel-purified streptavidin was used to label ssDNA sequencing fragments generated using a 5'-biotinylated primer [19]. Using the more homogeneous streptavidin as a drag-tag at the 5' end of the sequencing fragments, and employing a more effective wall-coating agent, approximately 110 bases of the four-color sequencing reaction were separated by ELFSE. Although these initial results were promising, the main limitation preventing the further use of ELFSE has been the lack of suitable large, water-soluble, monodisperse drag-tags with appropriate chemical functionality for unique attachment to DNA. More recently, progress has been made with the development of novel drag-tags consisting of long, repetitive, genetically engineered polypeptides (or "protein polymers") [21, 22], or linear or branched polyamides synthesized by solid-phase techniques [23–25]. A variety of these new drag-tags have been used in this study to revisit the potential for performing ELFSE separations of DNA molecules with drag-tags at each end.

1.2 Theory of end-effects in ELFSE

The standard theory of ELFSE was developed through investigations into the electrophoretic mobility of polymers with nonuniform charge distributions. For the case of the migration of a DNA-drag-tag conjugate, with a charged DNA segment consisting of M_C charged monomers and an uncharged drag-tag consisting of M_U uncharged monomers, the mobility μ is given by a weighted average of the electrophoretic mobilities of the charged and uncharged monomers:

$$\mu = \mu_0 \frac{M_C}{M_C + \alpha_1 M_U} \tag{1}$$

where μ_0 is the mobility of the charged monomers (*i.e.*, the free-solution mobility of DNA). (The uncharged monomers have zero electrophoretic mobility, and thus do not appear in the numerator of Eq. (1).) The parameter α_1 reweights the number of uncharged monomers M_U to reflect differences in persistence length and other hydrodynamic properties. The product $\alpha_1 M_U$, referred to as α, describes the total friction provided by the drag-tag, in terms of the number of additional uncharged monomers of DNA that would add equivalent friction. Thus, in the experiments described previously [18], a single streptavidin drag-tag provided $\alpha = 23$, *i.e.*, an amount of friction equivalent to 23 uncharged bp of DNA, whereas two streptavidins gave $\alpha = 54$. Notably, Eq. (1) cannot adequately explain the more than doubling of α arising from using two drag-tags.

The weighting of the individual monomer units in constructing the average in Eq. (1) was recently re-examined theoretically [26]. Whereas previous theory assumed that each monomer unit (after rescaling the uncharged monomers by α_1) contributes equally to the electrophoretic mobility of the composite molecule, more recent theory has taken into account end-effects originally described by Long *et al.* [27]. According to this theory, monomer units near either end of the polymer chain have greater influence than monomer units near the middle in determining the electrophoretic mobility of the composite molecule. This can be expressed by including a weighting factor ψ in the calculation of the mobility. For the case of ELFSE, with M_C charged monomers conjugated end-on to M_U uncharged monomers, and scaling M_U by the factor α_1 such that the total number of monomers is effectively $N = M_C + \alpha_1 M_U$, the weighted average mobility is expressed as

$$\mu = \frac{1}{N}\int_0^{M_C} \mu(n)\Psi\left(\frac{n}{N}\right)dn \quad (2)$$

where the index of integration, n, represents the position of a charged monomer unit in the chain. The ratio n/N, which appears as the argument of the weighting function ψ, ranges from 0 to 1, and represents the relative position of a given monomer unit in the chain. The limits of integration are written from 0 to M_C (rather than 0 to N) since the uncharged monomers ($n = M_C + 1..N$) have zero electrophoretic mobility, and only the charged monomers contribute to the total. Making the further substitution that for charged DNA monomers, the mobility $\mu(N) = \mu_0$, and using the definition $N = M_C + \alpha_1 M_U$, the mobility of the composite molecule can be written as

$$\mu = \frac{\mu_0}{M_C + \alpha_1 M_U}\int_0^{M_C} \Psi\left(\frac{n}{M_C + \alpha_1 M_U}\right)dn \quad (3)$$

The normalized weighting function $\psi(n/N)$ of a Gaussian polymer chain was found in [26] to be well represented by the following function:

$$\Psi\left(\frac{n}{N}\right) \approx -0.65 + 0.62\left(\frac{n}{N}\right)^{-\frac{1}{4}} + 0.62\left(1 - \frac{n}{N}\right)^{-\frac{1}{4}} \quad (4)$$

Equation (4) is a well behaved, easily calculated (and easily integrated) function for $0<(n/N)<1$, and is depicted in Fig. 1 of [26]. Using this functional form in Eq. (3) allows the straightforward calculation of the electrophoretic mobility for any composite molecule consisting of a DNA chain linked end-on to an uncharged drag-tag chain, provided that the scaling factor α_1 is known for a given set of experimental conditions. For the slightly more complicated case of a charged DNA chain with uncharged drag-tags at *both ends* of the DNA chain, Eqs. (2) and (3) need only be modified by changing the limits of integration, and the total number of effective monomer units N. For the case of a DNA chain consisting of M_C charged monomers, with identical drag-tags consisting of M_U uncharged monomers at each end, the total number of effective monomers is now $N = M_C + 2\alpha_1 M_U$. With this change, and inserting the appropriate limits of integration, the mobility becomes

$$\mu = \frac{\mu_0}{M_C + 2\alpha_1 M_U}\int_{\alpha_1 M_U}^{\alpha_1 M_U + M_C} \Psi\left(\frac{n}{M_C + 2\alpha_1 M_U}\right)dn \quad (5)$$

Besides providing a more complete analysis of the electrophoretic mobility of ELFSE conjugates, and improving the quantitative analysis of previous data from the molar mass profiling of PEG [28], the theory of end-effects makes useful predictions for enhancing the performance of DNA sequencing and other separations using ELFSE. The $\psi(n/N)$ function in Eq. (4) has its maxima near the ends of the molecule, indicating that the chain ends are

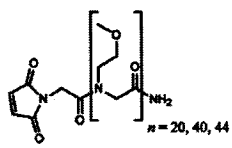
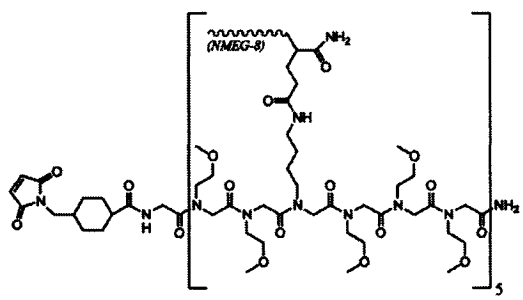
Drag-tag	Structure	References
NMEG-20 NMEG-40 NMEG-44		[23-25]
Branched NMEG-70		[31]
P1-169	H_2N - (Gly Ala Gly Gln Gly Ser Ala)$_{24}$ Gly - COOH	[21, 22]
P2-127	H_2N-(GAGTGSA)$_4$-GAGTGRA-(GAGTGSA)$_7$-GAGTGRA-(GAGTGSA)$_5$-G-COOH	[21]

Figure 1. Structures and code names for the six different drag-tag molecules used in this study. P1–169 and P2–127 drag-tags had maleimide functionalites added to their N-termini by activation with sulfosuccinimidyl 4-*N*-maleimidomethyl cyclohexane-1-carboxylate (Sulfo-SMCC), as described in [22].

Electrophoresis 2006, *27*, 1702–1712

weighted more heavily in determining the electrophoretic mobility of the composite molecule. The heavier weighting of the chain ends implies that adding an uncharged drag-tag to each end of a DNA molecule provides more than twice the drag of using a single drag-tag of the same size at one end of the DNA molecule. This is consistent with the initial experimental observations using streptavidin as a drag-tag [18]. Moreover, since the production of very large, totally monodisperse drag-tag molecules has thus far been problematic [22, 29], the effect might be exploited to provide sufficient drag for high-efficiency separations by using two smaller (and more monodisperse) drag-tags, rather than one larger drag-tag. In this study, we provide experimental confirmation of this effect using both short ssDNA oligos and larger dsDNA PCR products, with drag-tags of varying sizes at one or both ends of the DNA molecules.

2 Materials and methods

2.1 Chemicals

Tris(2-carboxyethylphosphine) (TCEP) and maleimide were purchased from Acros Organics (Morris Plains, NJ, USA). Sulfosuccinimidyl 4-*N*-maleimidomethyl cyclohexane-1-carboxylate (Sulfo-SMCC) was purchased from Pierce (Rockford, IL, USA). Buffer salts Tris (free base), *N*-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS), and EDTA were purchased from Amresco (Solon, OH, USA). POP-6 polymer solution was purchased from Applied Biosystems (Foster City, CA, USA). All water was purified using an E-Pure system from Barnstead (Boston, MA, USA) to a minimum resistivity of 17.8 MΩ·cm.

2.2 Drag-tag molecules

Six different drag-tag molecules were used in this study. Three were linear *N*-methoxyethylglycine (NMEG) oligomers of length 20, 40, or 44 monomers, produced by a solid-phase submonomer synthetic protocol [30], capped with an *N*-terminal maleimide, and purified to monodispersity by RP-HPLC as described previously [23–25]. Another drag-tag used was a monodisperse branched molecule consisting of a 30mer poly(NMEG) backbone with five octamer oligo(NMEG) branches, also described previously [31]. The final two drag-tags were repetitive protein polymers of length 127 and 169 amino acids, produced using the controlled cloning technique [21], and activated at the *N*-termini using the heterobifunctional cross-linker Sulfo-SMCC by reacting the protein polymers with a ten-fold molar excess of Sulfo-SMCC for 1 h at room temperature and pH 7.2, and then removing excess cross-linker by gel filtration as described previously [22, 29]. The structures and short names of the drag-tags are shown in Fig. 1. The NMEG-20 and NMEG-40 drag-tags were used for the studies of ssDNA, whereas the larger tags were used for the studies of dsDNA. All of the drag-tags used are hydrophilic, water-soluble molecules. Following the maleimide activation of the *N*-termini, the NMEG drag-tags are charge-neutral, whereas the P1–169 has a net charge of –1 (from deprotonation of the C-terminus), and the P2–127 (with two cationic arginine residues) has a net charge of +1.

2.3 Production of ssDNA conjugates

Two poly(dT) oligonucleotides of length 20 and 40 bases were purchased from Integrated DNA Technologies (Coralville, IA, USA). The oligos were modified at the 5'-end with a thiol linker that has a 6-carbon spacer, and at the 3'-end with a thiol linker having a 3-carbon spacer. The oligos were also modified internally with a fluorescein-dT base near the middle of the chain. These dithiolated, fluorescently labeled oligos (referred to as T20-dithiol and T40-dithiol) are shown schematically in Table 1.

Table 1. Oligonucleotides used for producing ssDNA conjugates with drag-tags at one or both ends

Oligonucleotide	Sequence
T20-dithiol	X_1 TTTTTTTTTX_2 TTTTTTTTT X_3
T40-dithiol	X_1 TTTTTTTTT TTTTTTTTTX_2 TTTTTTTTT TTTTTTTTT X_3

X_1, 5'-thiol linker with 6-carbon spacer; X_2, internal fluorescein-dT base; X_3, 3'-thiol linker with 3-carbon spacer

The thiol linkers on the DNA oligos were reduced using TCEP. To accomplish this reduction, 400 pmol of the dithiolated ssDNA (either T20-dithiol or T40-dithiol) was mixed with a 40:1 molar excess of TCEP, in a total volume of 10 μL of sodium phosphate buffer (100 mM, pH 7.2). This mixture was incubated at 40°C for 2 h. The reduced DNA was then split into aliquots of 10 pmol each prior to the addition of the drag-tag. To one aliquot, a large excess of maleimide (5 nmol) was added, capping the reduced thiols, and creating ssDNA molecules with no drag-tag (except the maleimide). To another aliquot, a large excess of drag-tag (1 nmol of either NMEG-20 or NMEG-40) was added, such that the majority of ssDNA molecules would have polymeric drag-tags at both ends. The other aliquots were treated with different amounts of drag-tag, from 50 to 200 pmol, with the intent of creating mixtures containing appreciable amounts of DNA with zero, one, or

two drag-tags. After reacting for approximately 90 min, an excess of maleimide (5 nmol) was added to these reactions to cap any remaining free thiols. The reactions were incubated in the dark at room temperature for at least 4 h prior to CE analysis.

2.4 Production of dsDNA conjugates

Oligonucleotides used as PCR primers were purchased from Integrated DNA Technologies, and are shown schematically in Table 2. The oligonucleotides consist of an M13 forward primer with a 5'-thiol linker and an internal fluorescein-dT base, and a set of M13 reverse primers, with or without 5'-thiol linkers, designed to produce dsDNA products of 75, 100, 150, or 200 bp in size when used in a PCR reaction with the forward M13 primer.

PCR reactions were performed using Pfu Turbo polymerase (Stratagene, La Jolla, CA, USA). Eight reactions were carried out with 20 pmol of the fluorescently labeled, thiolated M13 forward primer, and 20 pmol of each of the M13 reverse primers shown in Table 2, in a total volume of 20 μL. M13mp18 control DNA from a sequencing kit (0.2 μL) (Amersham Biosciences, Piscataway, NJ, USA) was used as a template. The M13 template was PCR-amplified with 32 cycles of denaturation at 94°C for 30 s, followed by annealing at 54°C for 30 s and extension at 72°C for 60 s. Products were analyzed by 2.5% agarose gel electrophoresis to confirm the sizes of the dsDNA amplicons, and the products were stored at –20°C until subsequent use.

Thiolated PCR products were reduced using a large excess of TCEP. To do this, 7 μL of PCR product was mixed with 0.7 μL of 1 M TCEP (in 1 M Tris buffer), plus an additional 0.35 μL of 1 M Tris, resulting in a solution of pH ~ 5. This mixture was incubated for 2–2.5 h at 40°C. Excess TCEP as well as PCR reaction components was removed using QIAquick PCR purification spin columns (QIAgen, Valencia, CA, USA) according to the manufacturer's instructions, with elution of the purified DNA in 30 μL of 100 mM sodium phosphate buffer, pH 7.2.

Table 2. Oligonucleotides used as PCR primers for producing dsDNA conjugates with drag-tags at one or both ends.

Oligonucleotide	Sequence
M13-Forward	X_1 CCX_2TTTAGGG TTTTCCCAGT CACGACGTTG
75-Reverse	GAGTCGACCT GCAGGCATGC
75-Reverse-T	X_1 GAGTCGACCT GCAGGCATGC
100-Reverse	GAGCTCGGTA CCCGGGGATC
100-Reverse-T	X_1 GAGCTCGGTA CCCGGGGATC
150-Reverse	GCGGATAACA ATTTCACACA
150-Reverse-T	X_1 GCGGATAACA ATTTCACACA
200-Reverse	CCAGGCTTTA CACTTTATGC
200-Reverse-T	X_1 CCAGGCTTTA CACTTTATGC

X_1, 5'-thiol linker with 6-carbon spacer; X_2, internal fluorescein-dT base

The purified PCR products (with one or two reduced thiols, depending on the reverse primers used) were split into multiple aliquots, and treated with one of four maleimide-activated drag-tags: NMEG-44, branched NMEG-70, P1–169, or P2–127. The amounts of drag-tag were sufficient in most cases to produce significant quantities of DNA with one or two drag-tags. Additional aliquots were treated with excess maleimide, to simply cap the reduced thiols and prevent further reaction or dimerization.

2.5 CE analysis of conjugates

Free-solution CE analysis was performed using an Applied Biosystems Prism 3100 Genetic Analyzer (Applied Biosystems), using an array of 16 fused-silica capillaries with inner diameter of 50 μm and a total length of 47 cm (36 cm to the detector). The running buffer was 89 mM Tris, 89 mM TAPS, 2 mM EDTA, pH 8.5, and 1% v/v POP-6 polymer solution to act as a wall-coating agent, with the adsorbed poly(dimethylacrylamide) effectively suppressing the EOF [32]. (The resulting polymer concentration is very low, and does not lead to any size-based sieving of the DNA.) Samples were diluted in water prior to analysis, to provide signals of appropriate strength for the fluorescence detector. The ssDNA samples were analyzed at 55°C, whereas dsDNA samples were analyzed at 25°C to prevent denaturation. Samples were introduced into the capillaries by electrokinetic injection at 1 kV (22 V/cm) for 2–20 s. Separations were carried out at 15 kV (320 V/cm). The fluorescein label of the DNA was detected in the "G" channel of ABI Dye Set E5, with λ_{max} 530 nm.

3 Results

3.1 Analysis of ssDNA conjugates

The experimental protocol in which ssDNA was mixed with different amounts of maleimide-activated drag-tag allowed the successful production of species with zero, one, or two drag-tags, which were easily separated and identified by free-solution CE analysis. This is illustrated in Fig. 2 for the case of the T40-dithiol DNA with NMEG-40 drag-tags. As seen in Fig. 2A, DNA with no drag-tag

Electrophoresis 2006, *27*, 1702–1712

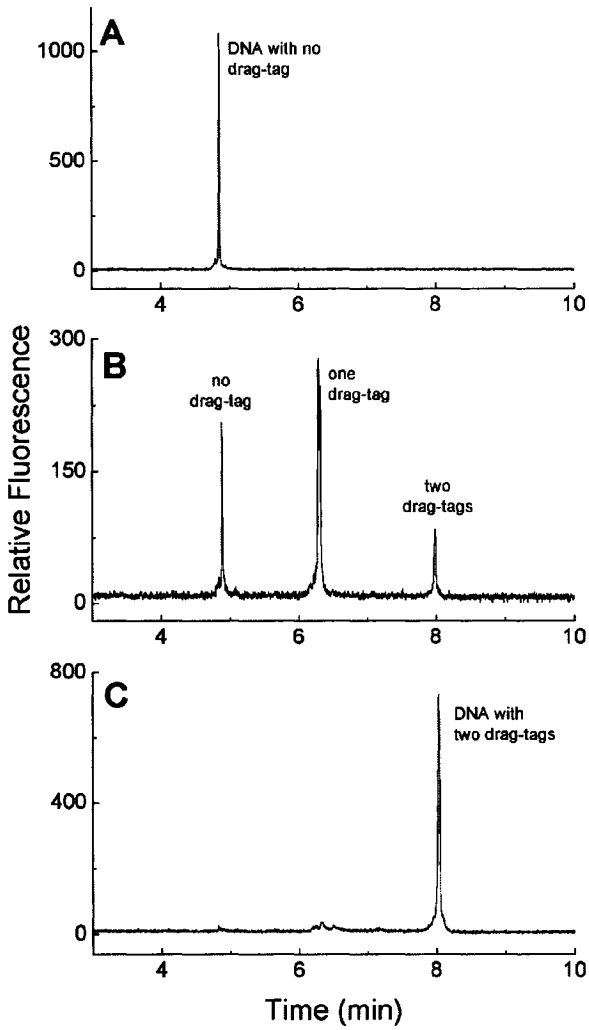


Figure 2. T40-dithiol DNA (A) capped at both ends with excess maleimide to create unlabeled ssDNA, (B) mixed with a 15:1 molar excess of NMEG-40 drag-tag followed by excess maleimide to create a mixture of unlabeled ssDNA and ssDNA with one or two drag-tags, and (C) mixed with a 100:1 molar excess of NMEG-40 drag-tag to create doubly labeled ssDNA. Samples were analyzed on an ABI 3100 capillary array instrument in 47 cm capillaries (36 cm to detector) in 89 mM Tris, 89 mM TAPS, 2 mM EDTA buffer, pH 8.5, with 1% v/v POP-6 polymer as a dynamic coating. Samples were injected electrokinetically at 22 V/cm for 3 s (A) or 2 s (B and C), and run at a field strength of 320 V/cm, with a current of 15 μA *per* capillary.

migrated as a single sharp peak with an electrophoretic mobility $\mu_0 = 3.9 \times 10^{-4}$ cm^2/V·s. Adding a 5- to 20-fold molar excess of the drag-tag to the DNA resulted in mixtures containing significant amounts of DNA with zero, one, or two drag-tags, as shown in Fig. 2B. Adding the drag-tag in a much larger molar excess (100-fold, relative to the DNA) led to nearly complete reaction of both ends of the DNA, again resulting in a single sharp peak as seen in Fig. 2C. Residual TCEP, present at 40-fold excess during the reduction, interferes somewhat with the reaction of the free thiols with the maleimide-activated drag-tags, and it was found that a significantly greater than 40-fold molar excess of drag-tag was necessary to achieve complete derivatization of both ends of the DNA. Species that were identified as ssDNA with one drag-tag typically appeared as a doublet of closely spaced peaks, as with the middle peak in Fig. 2B. The reason for this was not immediately obvious, but one possibility is that slight differences in electrophoretic mobility arise from labeling at the 5'- end or 3'- end of the DNA molecule, since the thiol linkers at the two ends are of different lengths.

In the optimized protocol, excess maleimide was used to cap any remaining unreacted thiols. We did this because, in initial attempts to produce mixtures comprising significant amounts of DNA with zero or one drag-tag, additional peaks would appear at characteristic spots in the electropherogram, particularly between the peaks for DNA with one and two drag-tags, and trailing the peak for DNA with two drag-tags. The extra peaks would be absent when the samples were first analyzed, but would grow in magnitude over the course of hours to days after the reduction of the DNA and reaction with the drag-tags. Although the extra peaks were never conclusively identified, it was hypothesized that they resulted from reoxidation of some of the residual free thiols to form disulfides. The addition of excess maleimide about 2 h after the addition of the drag-tag effectively prevented this problem, as the maleimide rapidly reacts with any remaining free thiols. The capping of both ends of the dithiolated DNA with this small molecule was found to induce a small, almost negligible mobility shift of 2–3 s relative to reduced, uncapped dithiolated DNA (data not shown), corresponding to an additional drag for the maleimide moiety equivalent to ~0.1 bases of DNA, as calculated in Section 4.

For each drag-tag (NMEG-20 or NMEG-40), samples consisting of both sizes of DNA (T20-dithiol or T40-dithiol) with zero, one, or two drag-tags were pooled to create mixtures containing multiple species, which were then separated and analyzed by CE. Run-to-run and capillary-to-capillary variabilities in migration time were generally quite low (approximately ± 1%), allowing easy identification of peaks in the pooled samples by comparing to the migration times of the individual components prior to pooling. CE analyses of these pooled mixtures are shown in Fig. 3, along with the peak assignments. A simple visual inspection confirms the general predictions of the end-

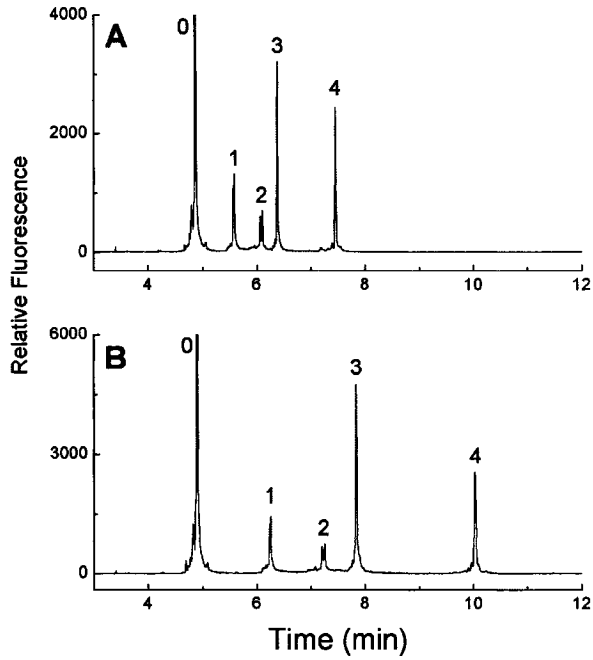


Figure 3. CE analysis of mixtures of 20mer and 40mer DNA with (A) NMEG-20 drag-tag, and (B) NMEG-40 drag-tag. Analysis conditions are the same as Fig. 2, except the injection was 22 V/cm for 15 s. Running current was 15 μA *per* capillary. Peak assignments for both (A) and (B) are: 0 = maleimide-capped DNA (no drag-tag); 1 = 40mer DNA with one drag-tag; 2 = 20mer DNA with one drag-tag; 3 = 40mer DNA with two drag-tags; 4 = 20mer DNA with two drag-tags.

effects theory: 20mer DNA with two 20mer drag-tags (Fig. 3A, Peak 4) migrates more slowly than 20mer DNA with one 40mer drag-tag (Fig. 3B, Peak 2), and likewise for the 40mer DNA (compare Fig. 3A, Peak 3 and Fig. 3B, Peak 1).

The apparent overall frictional parameter $\alpha = \alpha_1 M_U$ (as given by Eq. 1) could be computed directly from the peak times in Fig. 3. The α value calculated through use of Eq. (1), which neglects the end-effect, is termed the "apparent" α value so as to distinguish it from that determined using Eqs. (3) and (5), which account for the end-effect, as will be discussed in Section 4. The apparent α values, which qualitatively display the trend expected from the end-effects theory, are shown in Table 3. It is evident that two drag-tags give more than double the drag of a single tag, with roughly 6–9% enhancement for two drag-tags on ssDNA *versus* the expected drag for a single tag of twice the size. These experimental results will be analyzed quantitatively in Section 4, using the more detailed theory taking end-effects into account.

Table 3. Apparent frictional parameter α for ssDNA with one or two drag-tags calculated from peak times in Fig. 3, with correction made for the slight mobility shift arising from the maleimide capping

DNA length	Drag-tag	Apparent α	Error (±)	Ratio $(\alpha_{(2)}/2\alpha_{(1)})$
20	NMEG-20 (one)	5.1	0.07	1.07
	NMEG-20 (two)	10.9	0.1	
20	NMEG-40 (one)	9.7	0.1	1.09
	NMEG-40 (two)	21.2	0.2	
40	NMEG-20 (one)	6.1	0.08	1.06
	NMEG-20 (two)	12.9	0.2	
40	NMEG-40 (one)	11.2	0.2	1.09
	NMEG-40 (two)	24.5	0.3	

Final column gives the ratio of the drag for a tag at each end *vs.* the expected drag for a single tag of twice the size. Error margins on experimentally determined α values assume an uncertainty of ±0.05 min in peak times, which reflects the run-to-run and capillary-to-capillary variability observed with the instrument.

It is also clear from the results in Table 3 that the apparent α for a given size of drag-tag depends on the size of the DNA. For example, two NMEG-20 drag-tags on the 20mer DNA give α = 10.9, whereas the same two NMEG-20 drag-tags on the 40mer DNA give α = 12.9 – a difference of 18%. This is in agreement with the end-effects theory: For a drag-tag of a fixed size on one or both ends, a longer DNA molecule means that the drag-tag monomers are relatively closer to the chain end (*n*/*N* closer to 0 and/or 1), thereby giving the drag-tag monomers a heavier weighting in determining the mobility of the conjugate. Thus, the apparent α value for a given drag-tag on one or both ends of the DNA increases as the DNA chain length increases.

3.2 Analysis of dsDNA conjugates

dsDNA conjugate molecules were produced by performing PCR using a thiolated forward primer and normal (unthiolated) reverse primer (for production of dsDNA conjugates with a drag-tag at one end only), or using thiolated forward and reverse primers (for production of dsDNA conjugates with drag-tags at both ends). A large excess of TCEP was used for reduction of the thiols after the PCR reaction. Since TCEP is supplied as an HCl salt, the use of a large excess results in an acidification of the PCR buffer. To compensate for this, and to prevent long-term exposure of the DNA to very acidic conditions, additional 1 M Tris was added to the reduction mixture,

Electrophoresis 2006, 27, 1702–1712

resulting in a more acceptable pH. Following the reduction, the PCR products were purified using QIAquick spin columns, which effectively remove residual buffer salts, surfactants, enzyme, and reducing agents left over from the PCR reaction and reduction, which might otherwise interfere with reaction with the drag-tags.

The drag-tags used for the dsDNA conjugates were two moderately large synthetic polypeptoids (linear NMEG-44 and branched NMEG-70), and two protein polymers produced by genetic engineering of *Escherichia coli*. The branched NMEG-70 and the P1–169 drag-tags have been described previously for the separation of denatured (single-stranded) PCR products of sizes similar to those described here [22, 31]. In this study, CE analysis was performed at room temperature with no denaturants in the buffer, ensuring that the DNA remained in its double-stranded state. Keeping the DNA in its double-stranded state allows for the easy incorporation of a drag-tag at both ends, which was expected to generate more than twice the drag of a single drag-tag, allowing the separation of a wider size range of dsDNA molecules.

The concentration of the DNA purified with the QIAgen spin column was too low for accurate measurement of absorbance at 260 nm, and thus the molar ratios of DNA to drag-tag are not known precisely. The amounts of drag-tag were generally sufficient to produce significant amounts of product with zero and one drag-tag (for products with only the forward primer thiolated), and zero, one, and two drag-tags (for PCR products with both primers thiolated). Typical electropherograms for two sizes of DNA (100 bp and 200 bp) with the P2–127 protein polymer are shown in Fig. 4. In each case, the migration time of the "free" DNA (with no drag-tag) is around 6.2 min. In panels (A) and (C), which show PCR products generated with only a thiolated forward primer, the free DNA peak is followed by a single peak, corresponding to DNA with a single drag-tag. In panels (B) and (D), which show PCR products generated with both forward and reverse thiolated primers, there is an additional peak 1–2 min later, corresponding to DNA with a drag-tag at both ends. Note also in panels (B) and (D) that, for the products generated with both primers thiolated, there are two closely spaced peaks migrating around the same time as the product with one drag-tag in panels (A) and (C). As with the split peaks for the ssDNA conjugates with one drag-tag, the exact cause of this phenomenon is unknown, but it was observed for all sizes of dsDNA with all of the drag-tags, and may result from slight differences in electrophoretic mobility arising from labeling at either end of the DNA molecules.

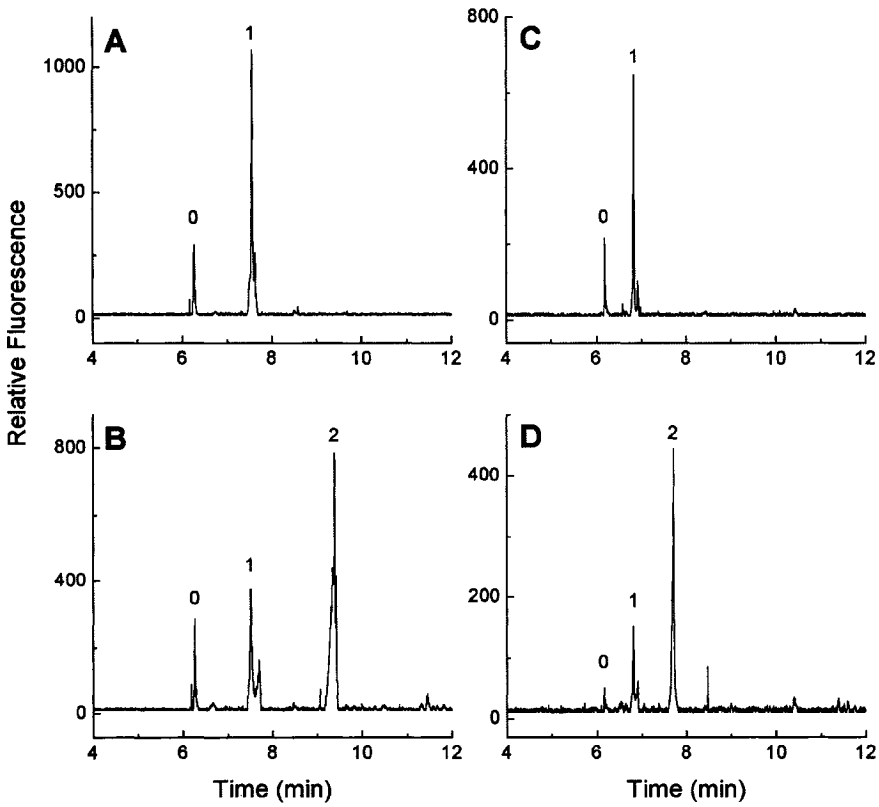


Figure 4. Electropherograms of dsDNA conjugated to P2–127 drag-tag. (A) 100-bp PCR product with forward primer thiolated, (B) 100-bp PCR product with both primers thiolated, (C) 200-bp PCR product with forward primer thiolated, and (C) 200-bp PCR product with both primers thiolated. Analysis conditions were the same as Fig. 2, except the run temperature was 25°C and the injection was 1 kV for 20 s. Peaks labeled 0, 1, and 2 refer to DNA species with zero, one, or two drag-tags, respectively.

The P1–169 and P2–127 protein polymers used here as drag-tags were not entirely monodisperse [22], leading to some additional peak broadness. The additional broadness is most noticeable with the smaller sizes of DNA, and is more pronounced for the species with two drag-tags. Both of these effects are as expected. Sharper peaks for larger sizes of DNA conjugated to impure drag-tags (including P1–169) were reported in [22], and are also in line with theory presented in [33]. The conjugation of a polydisperse drag-tag to both ends of a DNA molecule leads to a large number of possible combinations, each with slightly different electrophoretic mobility, which is apparent as additional peak broadness. The NMEG-44 and branched NMEG-70 drag-tags, both of which were purified to near monodispersity by RP-HPLC, generate cleaner, sharper peaks than the protein polymer drag-tags (data not shown).

Alpha values were calculated from the peak migration times of each species. In previous ELFSE literature, the relative mobilities of unlabeled and labeled DNA (μ_0/μ) would be plotted with respect to $1/M_C$, resulting in a straight line with slope α [18, 19]. This approach neglects the end-effects theory, which predicts a different overall value of α for each size of DNA. In this case, such plots are still essentially linear (not shown), and can be used to give an average apparent value of α for each drag-tag, as given in Table 4. (Note that the average α values determined by the linear fit of μ_0/μ *vs.* $1/M_C$ are not necessarily equal to the arithmetic average of the individual α values calculated for each size of DNA.) As indicated by the right-most ("Ratio") column in Table 4, the average α for two drag-tags is noticeably greater (10–23%) than twice α for a single-drag-tag, for these dsDNA species.

Table 4. Apparent frictional parameter α for dsDNA with one or two drag-tags, averaged for all sizes of DNA

Drag-tag	Average α	Ratio $(\alpha_{(2)}/2\alpha_{(1)})$
NMEG-44 (one)	12.7	1.10
NMEG-44 (two)	28.0	
Branched NMEG-70 (one)	17.0	1.22
Branched NMEG-70 (two)	41.6	
P1–169 (one)	27.2	1.13
P1–169 (two)	61.7	
P2–127 (one)	19.9	1.23
P2–127 (two)	48.8	

Final column gives the ratio of the drag for a tag at each end *vs.* the expected drag for a single tag of twice the size.

4 Discussion

The results we obtained for the analysis of ssDNA conjugates with poly(NMEG) drag-tags can be compared directly to the predictions from the end-effect theory presented in Eqs. (3) and (5). To take the end-effect into account, the weighting function presented in Eq. (4) is used. The parameter α_1 for scaling the uncharged monomers can be calculated using the end-effect theory, but we must first account for the slight additional drag arising from the maleimide moiety added to cap any unreacted thiols. To find the drag α_m associated with a single maleimide cap, the following equation was solved (using Maple):

$$t = \frac{t_0(M_c + 2\alpha_m)}{\int_{\alpha_m}^{\alpha_m + M_c} \Psi\left(\frac{n}{M_c + 2\alpha_m}\right) dn} \quad (6)$$

where t_0 is the arrival time of the uncapped DNA, and t is the arrival time of the DNA capped on each end with maleimide. For the 20-base DNA, α_m was found to be 0.035, while for the 40-base DNA it was found to be 0.052. Since the end-effect theory was derived for long Gaussian chains, it is assumed that the α_m value found for the larger DNA chain more closely represents the true value.

Note that the fluorescein-dT base near the middle of the chain likely exerts some effect on the mobility, as the fluorescein carries a -2 charge, and the dye along with the spacer arm linking it to the dT base likely add some hydrodynamic friction. To properly account for this effect would require a dithiolated oligonucleotide with no fluorescein, which would be undetectable with the CE instrument used for the analysis. The effect of the fluorescein is likely moderated by its position near the middle of the DNA chain (and hence its lower weight in determining the electrophoretic mobility). Additionally, the experimental determinations of α were made by comparing mobilities of drag-tag-labeled and free DNA, all of which were labeled identically with fluorescein. The impact on the results is expected to be minimal, and thus the contributions of the fluorescein as well as the thiol linkers present on all of the DNA species are ignored.

For DNA with one drag-tag and one maleimide cap, α_1 for the drag-tag can be found by solving Eq. (7)

$$t_1 = \frac{t_0(M_c + \alpha_m + \alpha_1 M_u)}{\int_{\alpha_m}^{\alpha_m + M_c} \Psi\left(\frac{n}{M_c + \alpha_m + \alpha_1 M_u}\right) dn} \quad (7)$$

where t_0 is the arrival time of the DNA with no drag-tag (after correcting for the presence of maleimide caps on each end), and t_1 is the arrival time of the DNA with one maleimide cap and one drag-tag. The calculated values of α_1 are presented in Table 5. Note that the closely

Electrophoresis 2006, *27*, 1702–1712

Table 5. Values of α_1 for NMEG drag-tags calculated from experimental data for ssDNA, taking into account the theory of end-effects

DNA length (M_C)	Drag-tag length (M_U)	α_1
20	20	0.19
	40	0.21
40	20	0.20
	40	0.21

spaced doublet for the arrival time of these singly labeled molecules was averaged for the results presented in Table 5; using either the faster or slower times resulted in α_1 values that differed from the average by a negligible amount. Note that the values of α_1 increase slightly with increasing size of the conjugate. For a given class of polymer, α_1 is expected to be a constant that is related to the chemical structures of the components and the experimental conditions (*i.e.* monomer size and Kuhn length, ionic strength of the buffer). The slight variation among the conjugates is likely due to the fact that the DNA and the drag-tags are too small to be perfectly Gaussian in conformation, which is an underlying assumption for the theory of ELFSE. Since the largest molecules are expected to be the closest to being Gaussian in conformation, we use the corresponding value of $\alpha_1 = 0.21$ to represent the true value for the poly(NMEG) drag-tags under the current experimental conditions.

Using the end-effect theory, the predicted arrival time for DNA with two drag-tags is

$$t_2 = \frac{t_0(M_c + 2\alpha_1 M_u)}{\int_{\alpha_1 M_u}^{\alpha_1 M_u + M_c} \Psi\left(\frac{n}{M_c + 2\alpha_1 M_u}\right) dn} \quad (8)$$

Equations (7) and (8) can now be used to predict the ratio of the mobilities of a bioconjugate with two drag-tags to the mobility of a conjugate with one drag-tag of twice the size, $\mu_2/\mu_1 = t_1/t_2$. The values predicted from Eqs. (7) and (8), using $\alpha_1 = 0.21$, are given in Table 6, along with the experimentally observed values, for the cases of 20mer or 40mer DNA with either a single 40mer drag-tag, or two 20mer drag-tags. The experimental results are closer to the value of 1, which is that predicted by the simple theory in Eq. (1) that neglects end-effects. The experimental value for the 40mer DNA is closer to the values predicted by the end-effect theory; this may be because the larger chains more closely approximate Gaussian coils, and are thus more appropriate test cases for the theory.

The quantitative end-effect theory is not directly applicable to the dsDNA data presented here. Although the dsDNA products are significantly longer, dsDNA is also considerably stiffer, with a much longer persistence length than ssDNA. Thus, even the longer dsDNA products are more likely to resemble stiff rods or cylinders, rather than random coils. Even with such a geometry, there is still likely an end-effect, which is dramatically illustrated by the experimental measurements of α presented in Table 4. Since the dsDNA-drag-tag conjugates are not likely to even approximate Gaussian coils, application of the theory used for the ssDNA conjugates is not appropriate.

Table 6. Mobility ratio μ_2/μ_1 for two 20mer drag-tags (μ_2) *vs.* one 40mer drag-tag (μ_1)

DNA length (M_C)	Predicted μ_2/μ_1	Experimental μ_2/μ_1
20	1.08	1.03
40	1.05	1.03

The drag enhancement for placing a drag-tag at each end of dsDNA is noticeably larger than was observed for placing a drag-tag at each end of ssDNA. This could simply be a function of the specific sizes of DNA and drag-tags that were chosen for study, but it may also be the result of the stiff rod-like structure of the dsDNA. Because the dsDNA molecules studied here are relatively short, the ends of the dsDNA molecule are more often on the "outside" of the chain, as opposed to a true Gaussian coil for which the chain ends may occupy positions in the interior of the coil. In addition, there may be a greater degree of hydrodynamic segregation between the rod-like dsDNA and the random coil drag-tags. Detailed theoretical analysis is required to determine if these simple arguments can explain the larger end-effect observed for dsDNA in these experiments.

The enhanced drag arising from placing a drag-tag at both ends of DNA leads to interesting new possibilities for sequencing and genotyping by ELFSE. The separation capacity of ELFSE is tied directly to the amount of friction generated by the drag-tag, and previous efforts have been focused on creating larger drag-tags to generate more friction. The possibility of including a drag-tag at both ends extends the range of separations that are possible with existing drag-tags. This is particularly important as the production of very large, totally monodisperse protein polymer drag-tags has proven difficult [22, 29]. The direct application of this technique to DNA sequencing would be difficult with current commercially available dye terminator chemistry, which presents no convenient functional group for attaching a second drag-tag at the 3'-end of the sequencing fragment. The dithiolated ssDNA oligos used in this study were custom-synthesized at considerable expense, and further appli-

cation to separation of ssDNA will likely require new developments in sequencing chemistry. The application of labeling both ends to the separation of dsDNA generated by PCR is more straightforward, given the wide availability of custom-synthesized DNA primers with a variety of functional groups and linkers that can be incorporated at the 5'-end.

In conclusion, this study has provided verification of an important and interesting prediction of the new theory of end-effects in ELFSE separations. Using both custom-synthesized ssDNA oligonucleotides and larger dsDNA products generated by PCR, labeled at one or both ends with a variety of drag-tags, it has been shown that the drag induced by labeling both ends is more than double the drag arising from a single drag-tag at one end, and is also larger than the drag that would arise from a single drag-tag of twice the size at one end. The effect is significant, with drag (α) enhanced by 6–9% for the ssDNA and by 10–23% for the dsDNA in the size range tested with the available drag-tags. This enhanced drag from double end-labeling could potentially be useful for various types of ELFSE separations such as DNA sequencing, if a suitable experimental approach can be developed for incorporating a drag-tag on each end of the DNA prior to analysis.

While the experimental data qualitatively show the trends expected from the end-effect, the data for ssDNA differ somewhat with the quantitative theoretical predictions. The agreement between the experimental data and the theory is closer for the larger molecules studied, which may be because these molecules are more Gaussian in conformation. Ideally, much larger (and hence more Gaussian) ssDNA and drag-tags could be used to test the end-effect theory; however, the direct synthesis of very long, doubly thiolated, and internally fluorescently labeled oligonucleotides, followed by purification to monodispersity, would be low-yielding and cost-prohibitive. Further tests with much larger ssDNA will require a different approach to creating long ssDNA with functional groups on each end appropriate for conjugation of a drag-tag.

The work was supported, in part, by the National Institutes of Health (NIH) of the USA (Grant No. NHGRI R01 HG002918–01), as well as a University of Ottawa Admission Scholarship to LM. 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.

5 References

[1] Viovy, J. L., *Rev. Mod. Phys.* 2000, *72*, 813–872.
[2] Slater, G. W., Kenward, M., McCormick, L. C., Gauthier, M. G., *Curr. Opin. Biotechnol.* 2003, *14*, 58–64.
[3] Meistermann, L., Tinland, B., *Phys. Rev. E* 1998, *58*, 4801–4806.
[4] Pluen, A., Tinland, B., Sturm, J., Weill, G., *Electrophoresis* 1998, *19*, 1548–1559.
[5] Medintz, I. L., Paegel, B. M., Mathies, R. A., *J. Chromatogr. A* 2001, *924*, 265–270.
[6] Woolley, A. T., Mathies, R. A., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 11348–11352.
[7] Woolley, A. T., Mathies, R. A., *Anal. Chem.* 1995, *67*, 3676–3680.
[8] Giordano, B. C., Ferrance, J., Swedberg, S., Huhmer, A. F. R., Landers, J. P., *Anal. Biochem.* 2001, *291*, 124–132.
[9] Chiesl, T. N., Shi, W., Barron, A. E., *Anal. Chem.* 2005, *77*, 772–779.
[10] Slater, G. W., Desruisseaux, C., Hubert, S. J., Mercier, J. F. *et al.*, *Electrophoresis* 2000, *21*, 3873–3887.
[11] Han, J., Craighead, H. G., *Science* 2000, *288*, 1026–1029.
[12] Han, J. Y., Craighead, H. G., *Anal. Chem.* 2002, *74*, 394–401.
[13] Barron, A. E., Soane, D. S., Blanch, H. W., *J. Chromatogr. A* 1993, *652*, 3–16.
[14] Chou, C. F., Austin, R. H., Bakajin, O., Tegenfeldt, J. O. *et al.*, *Electrophoresis* 2000, *21*, 81–90.
[15] Volkmuth, W. D., Austin, R. H., *Nature* 1992, *358*, 600–602.
[16] Noolandi, J., *Electrophoresis* 1992, *13*, 394–395.
[17] Mayer, P., Slater, G. W., Drouin, G., *Anal. Chem.* 1994, *66*, 1777–1780.
[18] Heller, C., Slater, G. W., Mayer, P., Dovichi, N. *et al.*, *J. Chromatogr. A* 1998, *806*, 113–121.
[19] Ren, H., Karger, A. E., Oaks, F., Menchen, S. *et al.*, *Electrophoresis* 1999, *20*, 2501–2509.
[20] Meagher, R. J., Won, J. I., McCormick, L. C., Nedelcu, S. *et al.*, *Electrophoresis* 2005, *26*, 331–350.
[21] Won, J. I., Barron, A. E., *Macromolecules* 2002, *35*, 8281–8287.
[22] Won, J. I., Meagher, R. J., Barron, A. E., *Electrophoresis* 2005, *26*, 2138–2148.
[23] Vreeland, W. N., Barron, A. E., *Abstr. Pap. Am. Chem. Soc.* 2000, *219*, 555–556.
[24] Vreeland, W. N., Slater, G. W., Barron, A. E., *Bioconjugat. Chem.* 2002, *13*, 663–670.
[25] Vreeland, W. N., Meagher, R. J., Barron, A. E., *Anal. Chem.* 2002, *74*, 4328–4333.
[26] McCormick, L. C., Slater, G. W., *Electrophoresis* 2005, *26*, 1659–1667.
[27] Long, D., Dobrynin, A. V., Rubinstein, M., Ajdari, A., *J. Chem. Phys.* 1998, *108*, 1234–1244.
[28] Vreeland, W. N., Desruisseaux, C., Karger, A. E., Drouin, G. *et al.*, *Anal. Chem.* 2001, *73*, 1795–1803.
[29] Won, J. I., Meagher, R. J., Barron, A. E., *Biomacromolecules* 2004, *5*, 618–627.
[30] Zuckermann, R. N., Kerr, J. M., Kent, S. B. H., Moos, W. H., *J. Am. Chem. Soc.* 1992, *114*, 10646–10647.
[31] Haynes, R. D., Meagher, R. J., Won, J. I., Bogdan, F. M., Barron, A. E., *Bioconjug. Chem.* 2005, *16*, 929–938.
[32] Doherty, E. A. S., Berglund, K. D., Buchholz, B. A., Kourkine, I. V. *et al.*, *Electrophoresis* 2002, *23*, 2766–2776.
[33] McCormick, L. C., Slater, G. W., Karger, A. E., Vreeland, W. N. *et al.*, *J. Chromatogr. A* 2001, *924*, 43–52.