Published in final edited form as:

Comput Chem Eng. 2014 May 7; 64: 63–70. doi:10.1016/j.compchemeng.2014.01.012.

Modeling and Global Optimization of DNA separation

Max A. Fahrenkopf^a, B. Erik Ydstie^{a,b}, Tamal Mukherjee^b, and James W. Schneider^a

^aDepartment of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

^bDepartment of Electrical & Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

Abstract

We develop a non-convex non-linear programming problem that determines the minimum run time to resolve different lengths of DNA using a gel-free micelle end-labeled free solution electrophoresis separation method. Our optimization framework allows for efficient determination of the utility of different DNA separation platforms and enables the identification of the optimal operating conditions for these DNA separation devices. The non-linear programming problem requires a model for signal spacing and signal width, which is known for many DNA separation methods. As a case study, we show how our approach is used to determine the optimal run conditions for micelle end-labeled free-solution electrophoresis and examine the trade-offs between a single capillary system and a parallel capillary system. Parallel capillaries are shown to only be beneficial for DNA lengths above 230 bases using a polydisperse micelle end-label otherwise single capillaries produce faster separations.

Keywords

Global Optimization; Nonlinear Programming; DNA electrophoresis; DNA Separation; End-labeled free-solution electrophoresis; ELFSE

1. Introduction

Global optimization is a tool of great utility for many applications in the chemical and process industries and beyond. Such applications include water treatment (Karuppiah & Grossmann, 2006), protein folding (Floudas et al., 1999), pooling and blending (Adhya et al., 1999), robust process control (Van Antwerp et al., 1999), and the prediction of phase and chemical equilibrium (McKinnon & Mongeau, 1998) just to name a few with more details found in a review by Floudas et al. (2005). The growing demand for rapid DNA analysis technology and the existence of closed form algebraic models that accurately describe several DNA analysis processes (Viovy, 2000; Dorfman et al., 2012) leads to a situation that necessitates the application global optimization to identify the proper DNA analysis designs.

© 2014 Elsevier Ltd. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The algebraic models that describe the many different DNA analysis processes vary significantly in complexity and structure. Given a general non-convex, non-linear programming (NLP) problem, the task of rigorously and reliably finding the global minimum is a challenging problem. Non-convex functions may have multiple local minima. Rigorous methods for solving these problems to global optimality rely on spatial branch and bound techniques to manually search the feasible region of the problem (Adjiman et al., 1998; Tawarmalani & Sahinidis, 2005, 2002). A spatial branch and bound algorithm is designed to reduce the gap between the upper bound, the smallest of the known local solutions, and the lower bound, the solution to a convex relaxation of the original problem. The efficiency of state-of-the-art global solvers α BB (Adjiman et al., 1998), BARON (Tawarmalani & Sahinidis, 2005; Sahinidis, 2012), LindoGLOBAL (Lin & Schrage, 2009), etc. is realized by utilizing tight convex relaxations for known functional forms such as bilinears, xy , (McCormick, 1976; Al-Khayyal & Falk, 1983), linear fractionals, x/y , (Quesada & Grossmann, 1995; Tawarmalani & Sahinidis, 2001) and concave univariates such as a square root function, $\sqrt{x}$, (Smith & Pantelides, 1999). Symbolic reformulation (Smith & Pantelides, 1999) and reformulation-linearization techniques for polynomial terms (Sherali & Tuncbilek, 1992), for instance, can also be used to restructure the original problem to further ensure a tight convex relaxation. With these principles in mind, global optimization becomes tractable for many applications including the one presented in this paper for optimal design of DNA length based separation devices.

Many rapid DNA analysis technologies utilize length based separation in capillary electrophoresis and microfluidic devices. Short-tandem repeat analysis (Butler, 2001; Shi, 2006), mutation detection for cancer diagnosis (Wu et al., 1997; Fraga et al., 2002; Albrecht et al., 2013) and sequencing (Sanger et al., 1977; Fredlake et al., 2008; Albrecht et al., 2011) are some examples of rapid DNA analysis using length based separation. Typically length based DNA separation is completed using gel electrophoresis, but gels are inherently slow due in part to the large amount of friction they impart onto the DNA. A Gel-free alternative called end-labeled free-solution electrophoresis (ELFSE) (Ren et al., 1999; Meagher et al., 2008) (also referred to as free-solution conjugate electrophoresis (FSCE) (Albrecht et al., 2011, 2013)) are much faster, resolving DNA lengths up to 265 bases in 30 minutes, with single-base resolution. These methods require that a highly monodisperse drag-tag be attached to the end of the DNAs in the sample. A variation of this technique is micelle-ELFSE electrophoresis (Grosser et al., 2007; Savard et al., 2008), which instead uses ensembles of transiently end-attached surfactant micelles as drag-tags. Here, stochastic variations in micelle size provide a highly uniform drag, despite a polydispersity of micelle size in the overall sample. Micelle-ELFSE allows the size of the drag-tag to be chosen simply by using buffers with a desired micelle size, without further chemical modification. This presents an opportunity to decrease run times by performing separations of longer DNA using larger drag-tags, and those for shorter DNA using shorter drag-tags. Another design trade-off is presented by the choice of electric field and capillary length. Use of high electric fields will give fast run times, but will cause excessive band broadening due to insufficient micelle size-sampling.

Here, we use global optimization methods to specify optimal micelle sizes (and other process parameters) for single- and multi-channel DNA separation by micelle-ELFSE. The framework presented here could easily be modified to optimally configure conventional ELFSE separations, which are often limited by band broadening due to wall adsorption. While optimal design of generic electrophoretic separation devices has been reported previously (Pfeiffer et al., 2004), we believe this is the first mathematical optimization of a DNA separation system. Such optimization is required due to the nonlinear and coupled nature of the ELFSE working equations. Optimal operating conditions for rapid, accurate DNA separations are identified using the global optimization solver BARON (Tawarmalani & Sahinidis, 2001; Sahinidis, 2012) implemented through the modeling language GAMS. Results obtained from optimization present the trade-offs of different experimental designs and the optimization frame work allows for rapid assessment of hypothetical constructs. We close with some conclusions about this approach and some future work.

2. Minimum run time under resolution constraint

Length based separation of DNA is successful when the DNA length of interest is resolved from the other lengths. A typical apparatus for completing this task is a capillary electrophoresis system with a diagram shown in figure 1. In capillary electrophoresis charge molecules separate over some distance l_D driven by an electric potential difference V_{app} . The charged molecules separate as they migrate at different rates down the capillary at rate dictated by their electrophoretic mobility

$$u = \mu E \quad (1)$$

where u is the analyte velocity, μ is the electrophoretic mobility and E is the applied electric field given by

$$E = \frac{V_{app}}{l_C} \quad (2)$$

where V_{app} is the applied voltage and l_C is the total length of the capillary. Molecules with naturally differing electrophoretic mobilities μ can be separated in free-solution capillary electrophoresis without the addition of a separation matrix. Many interesting molecules such as DNA, however, have electrophoretic mobilities μ that scale independently of length (Viovy, 2000). The length independent scaling can be broken with the addition of separation matrix to the capillary. Gel electrophoresis is commonly used to separate DNA but recent advances in end-labeled free-solution electrophoresis has identified an alternatives means of breaking the length independent scaling of the electrophoretic mobility (Ren et al., 1999; Meagher et al., 2008; Albrecht et al., 2011). The addition of a uncharged drag tag to DNA acts as a molecular parachute and has the advantage of significant speed-up over typical gel electrophoresis runs.

Separation of DNA using capillary electrophoresis is a semi-batch process, i.e. DNA is first injected as one plug, then the electric field is applied and the analytes migrate down the

capillary and separate according to their differing mobilities μ and create separate concentration bands. When the concentration bands are detected they are observed as Gaussians with some full-width at half-maximum W_i and mean migration time t_i . The width of the Gaussian signal is the result of various disturbances on the velocity of each analyte i . The quality of a length based separation is quantified by the resolution factor for each analyte R_i . The resolution between two bands is the ratio of the Gaussians' broadness to their spacing

$$R_i = \frac{W_{i+1} + W_i}{2|t_{i+1} - t_i|} \quad (3)$$

where W_i is the full width at half-maximum of the Gaussian, and t_i is the mean migration time for the analyte i . For a Gaussian profile, the full width at half-maximum is described in terms of its variance, σ_i^2 , such that $W_i = 2\sqrt{2\ln(2)\sigma_i^2}$. When the resolution factor in Eq. (3) is less than 1.5 then two concentration bands are considered resolved from each other.

The run conditions such as V_{app} , l_C and properties of the separation matrix, such as gel concentration, directly determine both the run time and the resolution for each analyte i . The separation is complete when all analytes i are resolved. The optimal run conditions for length-based separations using capillary electrophoresis can be found by solving the optimization problem Eq. (4)

$$\begin{aligned} \min_{\mathbf{z}} \quad & t_{run}(\mathbf{z}) = \max_{i \in I} \{t_i(\mathbf{z})\} \\ \text{s.t.} \quad & R_i(\mathbf{z}) \leq 1.5, \quad \forall i \in I \\ & R_i(\mathbf{z}) = \frac{W_{i+1}(\mathbf{z}) + W_i(\mathbf{z})}{2|t_{i+1}(\mathbf{z}) - t_i(\mathbf{z})|} \\ & W_i(\mathbf{z}) = 2\sqrt{2\ln(2)\sigma_i^2(\mathbf{z})} \\ & \sigma_i^2(\mathbf{z}) = s_i(\mathbf{z}) \\ & t_i(\mathbf{z}) = h_i(\mathbf{z}) \\ & \mathbf{g}(\mathbf{z}) \leq 0 \\ & \mathbf{z} \in \mathbb{R}^n \end{aligned} \quad (4)$$

where $s_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is the model for the variance generation during the separation, $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is the model for the migration time and $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m < n$ are the constraints on the system and the states $\mathbf{z}$. The run time is the longest migration time of all the analytes, $t_{run} = \max_{i \in I} \{t_i\}$ which is typically known to correspond to either the smallest or largest analyte depending on the separation method.

The models in the optimization problem Eq. (4) define two important features of the DNA separation problem: t_i the migration time and σ_i^2 the variance generated for each analyte i . Both of these functions dependent of the separation mode used. In the next section, we will discuss some of the sources of variance that occur during capillary electrophoresis.

3. Minimum run time DNA separation using micelle end-labeled free-solution electrophoresis

The optimization framework developed in the previous section can be utilized to find the optimal run conditions for micelle end-labeled free-solution electrophoresis (ELFSE). This problem is of interest to the authors as it was recently discovered by Schneider and co-workers (Grosser et al., 2007; Savard et al., 2008) to be a promising alternative to gel electrophoresis for rapid length based separation of DNA. The optimization model in Eq. (4) requires definition of migration time t_i and the variance σ_i^2 for each analyte i . In the next section we will discuss the micelle ELFSE model in detail and how it fits into our optimization framework.

In free-solution, DNA undergoes electrophoresis at a rate independent of its length. End-labeled free-solution electrophoresis (ELFSE) is a length based separation method that applies additional hydrodynamic friction to each DNA length through an end-attached drag tag which breaks the length independent scaling of DNA electrophoresis. From Eq. (1), the electrophoretic mobility dictates the velocity for each DNA length L . The electrophoretic mobility of the DNA drag-tag complex is given by

$$\mu = \mu_0 \left(\frac{L}{L + \alpha} \right) \quad (5)$$

where μ_0 is the free-solution mobility of DNA, L is the length of DNA in terms of bases and α is the number of DNA bases that have an equivalent hydrodynamic drag to the drag tag (Desruisseaux et al., 2001; Ren et al., 1999). The migration time for each DNA length L is then given by the ratio of capillary length to the velocity $t = l_D/u$, or

$$t = \frac{l_D}{\mu_0 E} \left(1 + \frac{\alpha}{L} \right) \quad (6)$$

where l_D is the length to the detector and E is the applied electric field strength. The migration time is function of capillary length l_D , applied voltage V_{app} and the “drag tag size” α which define the state variables $\mathbf{z} = [l_D, V_{app}, \alpha]^T$ for this specific problem and the free-solution mobility μ_0 is taken to be a constant parameter, $\mu_0 = 2.7 \times 10^{-4} \text{ cm}^2/\text{V}\cdot\text{s}$, (Istivan, 2012) for some implicit temperature, salt concentration and salt type (Stellwagen et al., 1997) not defined in the model.

Modeling the migration time alone does not give a complete description of the separation process; we must also define a model for the variance generated during the separation. The physics behind many velocity disturbances are modeled as random walks that are statistically independent from each other (Grossman & Colburn, 1992; Giddings, 1991; Dorfman et al., 2012). As independent random walks, the variance generated from each disturbance can be summed to yield the total variance $\sigma_x^2 = \sum \sigma_{x,\text{source}}^2$. For ELFSE using capillary electrophoresis, the significant sources of variance are injection variance, diffusion,

and drag tag polydispersity resulting in the total spatial variance (Ren et al., 1999; Meagher et al., 2008; Albrecht et al., 2011)

$$\sigma_x^2 = \sigma_{x, inj}^2 + \sigma_{x, diff}^2 + \sigma_{x, poly}^2 \quad (7)$$

Joule heating, wall adsorption and other smaller effects are neglected. It is important to note that the concentration bands are observed at a fixed position l_D away from injection and propagate in time as they pass the “finish-line” detector. The spatial variance is therefore observed as temporal variance under the transformation

$$\sigma_t^2 = \left(\frac{\partial t}{\partial x} \right)^2 \sigma_x^2 = \frac{\sigma_x^2}{u^2}. \quad (8)$$

The variance defined in the optimization problem Eq. (4) is, in fact, the total temporal variance for each analyte i , $\sigma_i^2 = \sigma_{t,i}^2$.

Injection variance $\sigma_{x, inj}^2$ is the initial variance of the concentration band at the beginning of the separation and is modeled as the variance from a rectangular plug

$$\sigma_{x, inj}^2 = \frac{l_w^2}{12} \quad (9)$$

where l_w is the width of the plug. Injection widths l_w can be controllable to a certain degree in capillary electrophoresis and in microfluidic devices they depend on the design on the injection scheme. For this work we assume $l_w = 0.1\% l_C$ where l_C is the total length of the capillary, typically $\delta l = 10$ cm longer than the length to the detector l_D , i.e. $l_C = l_D + \delta l$. This relationship gives reasonable estimates for the injection width and reflects the fact that longer capillaries require more initial analyte for detection downstream (Istivan, 2012).

Diffusion variance $\sigma_{x, diff}^2$ follows from thermal agitations causing stochastic motion in each DNA molecule of length L and is modeled by

$$\sigma_{x, diff}^2 = 2Dt \quad (10)$$

where D is the diffusion coefficient of the DNA of length L (Grossman & Colburn, 1992) and t is the migration time given by Eq. (6). The diffusion coefficients can be typically found in the literature, measured experimentally, or, if the analyte is a polymer, the diffusion coefficient can be estimated from well known scaling relations (Rouse, 1953; Zimm, 1956; Viovy, 2000). The scaling $D = D_1 / \sqrt{L + \alpha}$ works well for single-stranded DNA with $D_1 = 3.2 \times 10^{-6} \text{ cm}^2/\text{s}$ (Ren et al., 1999).

Micelle drag tag polydispersity $\sigma_{t,poly}^2$ is simplest to model using the temporal variance rather than the spatial variance. The relationship between migration time and the drag tag size α is seen in Eq. (6) to be linear thus it follows that

$$\sigma_{t,poly}^2 = \left(\frac{l_D}{\mu_0 E} \right)^2 \frac{\sigma_\alpha^2}{L^2} \quad (11)$$

where σ_α^2 is the variance in α . When the drag tag is a micelle, we need to account for the surfactant and micelle trading dynamics. Micelles are dynamic aggregations of surfactant molecules. Micelles quickly exchange surfactant molecules and will decompose completely after some time and later reform another micelle. For instance non-ionic micelles formed of Triton X-100 exchange surfactant molecules on the order of milliseconds (Rillaerts & Joos, 1982) and demicellizate on the order of seconds (Patist et al., 2002). This effect results in DNA sampling multiple micelles that they, themselves, change over time. A simple model can be derived by assuming that each of n micelle samples are independent of each other such that the variance due to micelle trading is $\sigma_{micelle}^2 = n\sigma_\alpha^2$ where the micelle trading occurs at a constant frequency ν , so that $n = \nu t$. Additionally, micelle polydispersity is known to be function of size (Mukerjee, 1972) so we further assume that the relative standard deviation $RSD = \sigma_{micelle}/\alpha$ is a weak function of α and DNA length L and is treated as a constant. Ultimately this results in the micelle polydispersity producing a standard deviation in equivalent micelle drag tag size $\sigma_{micelle}$ which depends linearly on equivalent micelle drag tag size α , which has been shown to fit micelle ELFSE data accurately (Istivan, 2012). Defining $B' = RSD^2/\nu$ we find the temporal variance due to micelle polydispersity to be of the form

$$\sigma_{t,poly}^2 = B' \left(\frac{\alpha}{L+\alpha} \right)^2 t \quad (12)$$

where t is the migration time of DNA length L and B' is the fractional polydispersity coefficient. We can convert to a spatial variance using $\sigma_x^2 = u^2 \sigma_t^2$

$$\sigma_{x,poly}^2 = B' \left(\frac{\alpha}{L+\alpha} \right)^2 u l_D \quad (13)$$

where u is the velocity of DNA length L and l_D is the capillary length to the detector.

Equipped with the models for migration time and variance generation for each DNA length we can formulate the optimization problem to find the drag tag size α , the capillary length to the detector l_D , and the applied voltage V_{app} that yield the minimum run time for resolving DNA from lengths $L_0 = 26$ bases to the length of read L which is varied from 30, 40, ..., 500 to show the trade offs between read frame size and run time. The optimization problem Eq. (4) can now be restated as

$$\begin{aligned}
\min \quad & t_{run} = \frac{l_D}{\mu_0 E} \left(1 + \frac{\alpha}{L_0}\right) \\
\text{s.t.} \quad & R_L \leq 1.5 \\
& R_L = \frac{2\sqrt{2\ln(2)\sigma_t^2}}{|\partial t / \partial L|} \\
& \left| \frac{\partial t}{\partial L} \right| = \frac{l_D}{\mu_0 E} \frac{\alpha}{L^2} \\
& E = V_{app} / l_C \\
& l_C = l_D + \delta l \\
& \sigma_t^2 = \frac{\sigma_x^2}{u^2} \\
& \sigma_x^2 = \sigma_{x, inj}^2 + \sigma_{x, diff}^2 + \sigma_{x, poly}^2 \\
& \sigma_{x, inj}^2 = \frac{(0.1\% l_C)^2}{12} \\
& \sigma_{x, diff}^2 = 2Dt \\
& D = \frac{D_1}{\sqrt{L+\alpha}} \\
& \sigma_{x, poly}^2 = B' \left(\frac{\alpha}{L+\alpha} \right)^2 u l_D \\
& t = \frac{l_D}{\mu_0 E} \left(1 + \frac{\alpha}{L}\right) \\
& u = \mu_0 E \left(\frac{L}{L+\alpha} \right) \\
& 0 \leq \alpha \leq \alpha_{max} \\
& l_{min} \leq l_D \leq l_{max} \\
& 0 \leq V_{app} \leq V_{max}.
\end{aligned} \tag{14}$$

The NLP (14) is written using a few assumptions and simplifications to the problem Eq. (4) developed above. First of all, the run time is known a priori to correspond to the migration time of the shortest DNA length, $t_{run} = t_{L_0}$, since the smallest DNA length, L_0 , has the least amount of charge available to electrophoreses against the drag tag and consequently migrates the slowest. The resolution constraint is also simplified as it can be shown that R_{L_0}

$R_{L_0+1} \dots R_L$. Furthermore the full width at half-maximum W_L is assumed to be equivalent to the full width of the next DNA length $W_L \approx W_{L+1}$ and the difference in migration time is calculated using a derivative $|t_{L+1} - t_L| \approx |t/L|$ which introduces negligible error for large L (Ren et al., 1999). Under these simplifications, only two DNA lengths need to be considered in the optimization problem: the length of read L which is the longest DNA length to be resolved during the separation and the shortest DNA length in the separation L_0 which sets the run time. The shortest DNA length in the separation is typically between 18 and 26 bases (Ren et al., 1999; Meagher et al., 2008; Istivan, 2012), for this work we use $L_0 = 26$. The DNA physical properties μ_0 , D_1 , the system property δl and the micelle physical property B' are taken as constant parameters.

The optimization problem Eq. (14) can be re-written using some algebraic manipulation to reveal the structure of the optimization problem shown in Eq. (15)

$$\begin{aligned}
\min \quad & t_{run} = \frac{p_1 \alpha l_D}{E} + \frac{p_2 l_D}{E} \\
\text{s.t.} \quad & p_3 s_x (L + \alpha)^2 L^2 \leq l_D^2 \alpha^2 \\
& s_x = s_{inj} + s_{diff} + s_{poly} \\
& s_{inj} = p_4 (l_D + p_5)^2 \\
& s_{diff} = p_6 \frac{l_D}{E} \frac{\sqrt{L + \alpha}}{L} \\
& s_{poly} (L + \alpha)^3 = p_7 L E l_D \alpha^2 \quad (15) \\
& V_{app} = (l_D + p_5) E \\
& 0 \leq \alpha \leq \alpha_{\max} \\
& l_{\min} \leq l_D \leq l_{\max} \\
& 0 \leq V_{app} \leq V_{\max}
\end{aligned}$$

where $s_x = \sigma_x^2$ and L is the length of read specified by the user. The parameter groupings p_i are fixed numbers given by $p_1 = \mu_0^{-1} L_0^{-1}$, $p_2 = \mu_0^{-1}$, $p_3 = 2.464$, $p_4 = 8.33 \times 10^{-8}$, $p_5 = \delta l$, $p_6 = 2D_1 \mu_0^{-1}$, $p_7 = B' \mu_0$ where we use parameter values $\mu_0 = 0.27 \text{ cm}^2/\text{kV}\cdot\text{s}$, $D_1 = 3.2 \times 10^{-6} \text{ cm}^2/\text{s}$, the parameter δl is equipment specific and is typically $\delta l = 10 \text{ cm}$, and B' is specified to range from 0 to 1.0 ms in this work to show the effect of micelle polydispersity. The optimization problem Eq. (15) is a non-convex NLP problem. Non-convex problems may have multiple local minima. The global optimization code BARON (Tawarmalani & Sahinidis, 2005) uses spatial branch and bound and symbolic reformulation to efficiently find the global minimum. For this work, we use BARON version 9.0.6 supplied in GAMS version 23.6.2.

Special care must be taken when choosing the units and scaling used for the problem. Large differences in variable values can result in poorly conditioned numerics that lead to an infeasible optimization problem. Also BARON works by making variable substitutions until functional forms that have known convex relaxations can be identified. Reformulating the problem to Eq. (15) simplifies the task for BARON which helps accelerate convergence to the global optimum.

4. Parallel capillaries

Certain capillary electrophoresis systems, such as the Applied Biosystems Prism 3100 Genetic Analyzer, are design to separate analytes with a capillary array. Capillary arrays are useful for parallel separations under the same applied voltage and capillary length with differing separation mediums (e.g. different sized micelles). Given the original optimization problem Eq. (15), it is straightforward to reformulate this problem to allow for a capillary array by copying the constraints for each capillary. The applied voltage and capillary length is the same for each capillary when a capillary array is used, but the drag tag size α in each capillary can be different. Each capillary must resolve part of the read frame and the optimization problem must therefore include a DNA length L_j which is split between two capillaries with one capillary resolving between the shortest length L_{j-1} and the split length L_j and the other capillary resolves from the split length L_j up to the next split length L_{j+1} . The specified length of read L is defined as L_N , where N is the total number of capillaries used and the specified shortest length of interest is set as L_0 .

$$\begin{aligned}
\min \quad & t_{run} = \max_i \{t_i\} \\
\text{s.t.} \quad & t_i = \frac{l_D}{\mu_0 E} \left(1 + \frac{\alpha_i}{L_{i-1}}\right), \quad i=1, \dots, N \\
& \text{(constraints from Eq. (15) for capillary } i) \\
& L_{j-1} \leq L_j \leq L_{j+1} \quad j=1, 2, \dots, N-1
\end{aligned} \tag{16}$$

The objective function $t_{run} = \max \{t_i\}$ is discontinuous and non-differentiable, however the optimization problem is reformulated to an equivalent NLP Eq. (17)

$$\begin{aligned}
\min \quad & \eta \\
\text{s.t.} \quad & t_i \leq \eta, \quad i=1, \dots, N \\
& t_i = \frac{l_D}{\mu_0 E} \left(1 + \frac{\alpha_i}{L_{i-1}}\right) \\
& \text{(constraints from Eq. (15) for capillary } i) \\
& L_{j-1} \leq L_j \leq L_{j+1}, \quad j=1, 2, \dots, N-1
\end{aligned} \tag{17}$$

which is non-convex and solved to global optimality using BARON. The split length L_j is treated as a design variable in the optimization problem. The total set of design variables are $\mathbf{z} = [\alpha_1, \dots, \alpha_N, L_1, \dots, L_{N-1}, V_{app}, l_D, \eta]^T$ for N parallel capillaries. The length of read L_N and the shortest DNA length of interest L_0 are specified by the user. Indeed this problem can be extended to an assignment problem using a mixed-integer non-linear programming (MINLP) approach (Pinto & Grossmann, 1998) to identify the optimal number of capillaries for a given separation task. This was not attempted in this current work and we restrict our view to using two parallel capillaries.

5. Results and Discussion

Figure 3 shows the optimization results for micelle ELFSE using a Beckman-Coulter P/ACE MDQ capillary electrophoresis system (0 V_{app} 30 kV, 20 cm l_D 100 cm) using polydisperse micelle drag tag bounded in size 0 α 3000 bases and assigned fractional polydispersity coefficients $B' = 0.0, 0.1, 0.6$, and 1.0 ms which are characteristic of C₁₂E₈ non-ionic surfactant mixed micelle samples previously tested (Istivan, 2012). Increasing the fractional polydispersity coefficient B' is shown to cause a significant increase on the run time for lengths of read above 150 bases.

Figure 4 shows the optimization results for two parallel capillaries with system specification consistent with an Applied Biosystems Prism 3100 Genetic Analyzer (0 V_{app} 15 kV, l_D = 36 cm, l_C = 47 cm). Using two parallel capillaries from the ABI system becomes advantageous over the single capillary Beckman system when the length of read is above 430 bases when B' is 0.1 ms, 270 bases when B' is 0.6 ms and 230 bases when B' is 1.0 ms and yields no benefit when B' is zero for the lengths of read investigated.

The optimal run conditions for the Beckman single capillary system are shown in Figure 5. When the fractional polydispersity coefficient B' is set to a low value of 0.0 or 0.1 ms, the variance generation does not inhibit the system from operating at maximum throughput by setting the capillary length to its minimum value and the applied voltage to its maximum value. The drag tag size can then be selected to attain resolution for the specified length

read. When B' is set to the value of 1.0 ms, the increased micelle polydispersity causes an increase in the run time by a factor of 6.6 for a length of read of 500 bases. Capillary length begins to increase from the base value of 20 cm when the length of read is greater than 170 bases. The applied voltage is optimal at 30 kV for all lengths of read and B' investigated. The optimal drag tag size α is similar for both $B' = 0.6$ ms and $B' = 1.0$ ms indicating that it is optimal to increase capillary length, rather than drag tag size α , to mitigate the increase in variance caused by the micelle polydispersity.

Figure 6 shows the sources of variance for each length of read and B' investigated. When the micelle drag tag is monodisperse, $B' = 0$, injection variance is the most significant source of variance for all lengths of read above 50 bases becoming nearly twice as large as the diffusion variance for large length of read. For the low B' value of 0.1 ms, injection, diffusion and polydispersity contribute to the total spatial variance on almost equal levels with injection variance contributing between 43% and 60% of the total variance for every length of read. Polydispersity variance only contributes as much as 40% to the total variance for a length of read of 500 bases. The polydispersity variance is negligible for low lengths of read contributing on 3% when the length of read 100 bases. As the polydispersity increases to correspond to B' of 1.0 ms, the polydispersity of the micelle is the most significant source of band broadening for lengths of read above 120 bases. As the polydispersity variance becomes more significant, the variance due to injection drops to between 18% and 48% with smaller lengths of read experiencing the greatest percentage of injection variance.

All computations were performed on a Desktop PC equipped with an Intel(R) Core i7 2.93 GHz quad-core processor. CPU times range from 0.22 s to 1.3 s for each length of read L and fractional polydispersity coefficient B' investigated. The CPU time increases as the resolution constraint becomes nearly infeasible as B' is set to large values. The single capillary and parallel capillary optimization problems solved in nearly the same amount of CPU time, owing to both problems being relatively small scale in terms of the number of equations and variables.

6. Conclusion

The optimization framework presented in this paper allows for rapid prototyping of DNA length based separation devices. Many DNA separation methods are modeled by algebraic equations that are well structured for global optimization. The case study for micelle end-labeled free-solution electrophoresis shows how these well structured algebraic models allow for efficient global optimization of the non-convex non-linear programming problem. DNA separation using a single capillary and two parallel capillaries were compared to show that parallel capillaries significantly reduce run time for large lengths of read using polydisperse micelles with fractional polydispersity coefficients B' above 0.6 ms. Increasing micelle polydispersity is shown to cause increasing large run times and eventually cause the NLP (15) to become infeasible. In this paper, we neglected wall adsorption and Joule heating which may need to be included in some cases. Fortunately, including the non-linear algebraic models for wall adsorption and Joule heating into the optimization problem is straightforward and our approach is easily extended to include these terms.

While the approach discussed in this paper covers a wide range of electrophoretic separation problems, there are some separation methods which are currently beyond our scope. We are essentially limited to problems in which a closed-form algebraic model exists for the migration time and the variance generation. Closed-form algebraic models are difficult to derive for complex transport phenomena. In gel electrophoresis, for instance, DNA can separate under several different regimes including sieving, Entropic Trapping, and Biased Reptation (Viovy, 2000) depending on the size of the DNA and the strength of the electric field. For optimization of these separation methods, the most fruitful approach may be to use black-box optimization techniques (Lagarias et al., 1998) with the simulations utilized for the separation problems (Deutsch, 1988; Viovy, 2000), although such an approach remains intractable for large-scale optimization problems.

Acknowledgments

Research supported by the CFD Research Corporation under Nation Institute of Health grant 2R44HG00429002 and the Center for Advanced Process Decision-making at Carnegie Mellon University.

References

- Adhya N, Tawarmalani M, Sahinidis NV. A lagrangian approach to the pooling problem. *Industrial & Engineering Chemistry Research*. 1999; 38:1956–1972.
- Adjiman C, Dallwig S, Floudas C, Neumaier A. A global optimization method, [alpha]BB, for general twice-differentiable constrained nlp-s-i. theoretical advances. *Computers & Chemical Engineering*. 1998; 22:1137–1158.
- Al-Khayyal FA, Falk JE. Jointly constrained biconvex programming. *Mathematics of Operations Research*. 1983; 8:273–286.
- Albrecht JC, Kotani A, Lin JS, Soper SA, Barron AE. Simultaneous detection of 19 K-*ras* mutations by free-solution conjugate electrophoresis of ligase detection reaction products on glass microchips. *Electrophoresis*. 2013; 00:1–8.
- Albrecht JC, Lin JS, Barron AE. A 265-base DNA sequencing read by capillary electrophoresis with no separation matrix. *Analytical Chemistry*. 2011; 80:509–515. [PubMed: 21182303]
- Butler, JM. *Forensic DNA Typing: Biology and Technology behind STR markers*. Academic Press; San Diego, CA, USA: 2001.
- Desruisseaux C, Long D, Drouin G, Slater GW. Electrophoresis of composite molecular objects. 1. relation between friction, charge and ionic strength in free solution. *Macromolecules*. 2001; 34:44–52.
- Deutsch JM. Theoretical studies of DNA during gel electrophoresis. *Science*. 1988; 240:922–924. [PubMed: 3363374]
- Dorfman KD, King SB, Olson DW, Tree JDP, Tree DR. Beyond gel electrophoresis: Microfluidic separations, fluorescence, burst analysis, and DNA stretching. *Chemical Reviews*. 2012 doi: 10.1021.
- Floudas CA, Akrotirianakis IG, Caratzoulas S, Meyer CA, Kallrath J. Global optimization in the 21st century: Advances and challenges. *Computers & Chemical Engineering*. 2005; 29:1185–1202.
- Floudas CA, Klepeis JL, Pardalos PM. Global optimization approaches in protein folding and peptide docking. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. 1999; 47:141–171.
- Fraga MF, Uriol E, Diego LB, Berdasco M, Esteller M, Caal MJ, Rodriguez R. High-performance capillary electrophoretic method for the quantification of 5-methyl 2'-deoxycytidine in genomic dna: Application to plant, animal and human cancer tissues. *Electrophoresis*. 2002; 23:1677–1681. [PubMed: 12179987]

- Fredlake CP, Hert DG, Wai Kan C, Chiesl TN, Root BE, Forster RE, Barron AE. Ultrafast DNA sequencing on a microchip by a hybrid separation mechanism that gives 600 bases in 6.5 minutes. *Proceedings of the National Academy of Sciences*. 2008; 105
- Giddings, J. *Unified Separation Science*. Wiley; 1991.
- Grosser ST, Savard JM, Schneider JW. Identification of PCR products using PNA amphiphiles in micellar electrokinetic chromatography. *Analytical Chemistry*. 2007; 79:9513–9519. [PubMed: 18020426]
- Grossman, PD.; Colburn, JC. *Capillary Electrophoresis: Theory and Practice*. Academic Press; 1992.
- Istivan, SB. PhD thesis. Carnegie Mellon University; Pittsburgh, PA 15213: 2012. Gel-free Separations of DNA via Transient Attachment to Surfactant Micelles.
- Karuppiah R, Grossmann IE. Global optimization for the synthesis of integrated water systems in chemical processes. *Computers & Chemical Engineering*. 2006; 30:650–673.
- Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*. 1998; 9:112–147.
- Lin Y, Schrage L. The global solver in the LINDO API. *Optimization Methods & Software*. 2009; 24:657–668.
- McCormick G. Computability of global solutions to factorable non-convex programs: Part I convex underestimating problems. *Mathematical programming*. 1976; 10:147–175.
- McKinnon K, Mongeau M. A generic global optimization algorithm for the chemical and phase equilibrium problem. *Journal of Global Optimization*. 1998; 12:325–351.
- Meagher R, Won J, Coyne J, Li J, Barron A. Sequencing of dna by free-solution capillary electrophoresis using a genetically engineered protein polymer drag-tag. *Analytical Chemistry*. 2008; 80:2842–2848. [PubMed: 18318549]
- Mukerjee P. Size distribution of small and large micelles. multiple equilibrium analysis. *Journal of Physical Chemistry*. 1972; 76:565–570.
- Patist A, Kanicky JR, Shukla P, Shah DO. Importance of micellar kinetics in relation to technological processes. *Journal of Colloids and Interface Science*. 2002; 245:1–15.
- Pfeiffer AJ, Mukherjee T, Hauan S. Design and optimization of compact microscale electrophoretic separation systems. *Industrial & Engineering Chemistry Research*. 2004; 43:3539–3553.
- Pinto JM, Grossmann IE. Assignment and sequencing models for the scheduling of process systems. *Annals of Operations Research*. 1998; 81:433–466.
- Quesada I, Grossmann IE. A global optimization algorithm for linear fractional and bilinear programs. *Journal of Global Optimization*. 1995; 6:39–76.
- Ren H, Karger A, Oaks F, Menchen S, Slater G, Drouin G. Separating DNA sequencing fragments without a sieving matrix. *Electrophoresis*. 1999; 20:2501–2509. [PubMed: 10499343]
- Rillaerts E, Joos PJ. Rate of demicellization from the dynamic surface tensions of micellar solutions. *Journal of Physical Chemistry*. 1982; 86:3471–3478.
- Rouse PE. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *The Journal of Chemical Physics*. 1953; 21:1272–1281.
- Sahinidis NV. *BARON 11.4.0: Global Optimization of Mixed-Integer Nonlinear Programs*. 2012 User's Manual.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977; 74:5463–5467.
- Savard JM, Grosser ST, Schneider JW. Length-dependent DNA separations using multiple end-attached peptide nucleic acid amphiphiles in micellar electrokinetic chromatography. *Electrophoresis*. 2008; 29:2779–2789. [PubMed: 18546164]
- Sherali HD, Tuncbilek CH. A global optimization algorithm for polynomial programming problems using reformulation-linearization technique. *Journal of Global Optimization*. 1992; 2:101–112.
- Shi Y. DNA sequencing and multiplex STR analysis on plastic microfluidic devices. *Electrophoresis*. 2006; 27:3703–3711. [PubMed: 16960837]
- Smith EM, Pantelides CC. A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. *Computers & Chemical Engineering*. 1999; 23:457–478.

- Stellwagen NC, Gelfi C, Righetti PG. The free solution mobility of DNA. *Biopolymers*. 1997; 42:687–703. [PubMed: 9358733]
- Tawarmalani M, Sahinidis N. Semidefinite relaxations of fractional programs via novel convexification techniques. *Journal of Global Optimization*. 2001; 20:133–154.
- Tawarmalani, M.; Sahinidis, NV. *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications*. Vol. 65. Springer; 2002.
- Tawarmalani M, Sahinidis NV. A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming*. 2005; 103:225–249.
- Van Antwerp JG, Braatz RD, Sahinidis NV. Globally optimal robust process control. *Journal of Process Control*. 1999; 9:375–383.
- Viovy JL. Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Reviews of Modern Physics*. 2000; 72:813–872.
- Wu Y, Nystrom-Lahti M, Osinga J, Looman MWG, Peltomki P, Aaltonen LA, de la Chapelle A, Hofstra RMW, Buys CHCM. MSH2 and MLH1 mutations in sporadic replication error-positive colorectal carcinoma as assessed by two-dimensional DNA electrophoresis. *Genes, Chromosomes and Cancer*. 1997; 18:269–278. [PubMed: 9087566]
- Zimm BH. Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *The Journal of Chemical Physics*. 1956; 24:269–279.

Highlights

- Optimization of DNA separation requires a model for migration time and its variance
- Micelle polydispersity is shown hinder the optimal performance of micelle ELFSE
- Parallel capillaries are advantageous for long DNA separation only

Figure 1.

Schematic of a capillary electrophoresis system. DNA is separated and detected after a distance l_D away from injection with throughput controlled by the applied electric field E which is the applied voltage V_{app} over the total capillary length l_C . The capillary is filled with either a gel to perform gel electrophoresis or micelles, characterized by their size α , to perform micelle end-labeled free-solution electrophoresis. The finish-line detector observes a Gaussian signal for each resolved DNA length. The initial few signals are sufficiently resolved to identify each profile. The last two signals are over-resolved at the expense of run time which may justify the use of a second capillary to speed up the separation.

Figure 2.

Resolution is the ratio of mean signal width $0.5 (W_{i+1} + W_i)$ to signal spacing $(t_{i+1} - t_i)$. In the case of micelle end-labeled free-solution electrophoresis the larger analytes elute first.

Figure 3.

Results from optimization problem Eq. (15) for specified lengths of read L of 30, 40, ..., 500 bases. Drag tag size α is bounded in $0 < \alpha < 3000$ bases, the applied voltage is bounded by $0 < V_{app} < 30$ kV, and the capillary length is bounded by $20 \text{ cm} < l_D < 100$ cm. The ranges on applied voltage and capillary length are consistent with a Beckman-Coulter P/ACE MDQ capillary electrophoresis system. Micelle polydispersity coefficients B' is specified as 0.0, 0.1, 0.6 and 1.0 ms. Increasing B' indicates micelles of greater polydispersity or slower surfactant and micelle trading dynamics.

Figure 4.

Results from optimization problem Eq. (17) for specified lengths of read L of 30, 40, ..., 500 bases. Drag tag size α is bounded in $0 \leq \alpha \leq 3000$ bases for each capillary, the applied voltage is bounded by $0 \leq V_{app} \leq 15$ kV, and the capillary length is specified as $l_D = 36$ cm. The range on applied voltage and specified capillary length are consistent with an Applied Biosystems Prism 3100 Genetic Analyzer system. Micelle polydispersity coefficients B' is specified as 0.0, 0.1, 0.6 and 1.0 ms.

Figure 5.

Optimal run conditions for a single capillary. Capillary length, applied voltage and drag tag size are bounded by 20 cm l_D 100 cm, 0 V_{app} 30 kV and 0 α 3000, respectively. The optimal applied voltage is 30 kV for every length of read and B' investigated.

Figure 6.

Spatial variance sources for a single capillary micelle ELFSE separation for lengths of read L of 30, 40, ..., 500 bases and B' of 0.0, 0.1, 0.6 and 1.0 ms. Polydispersity is the most significant source of band broadening for lengths of read above 120 bases and $B' \geq 0.6$ ms.