Review

Brett A. Buchholz
Annelise E. Barron

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

The use of light scattering for precise characterization of polymers for DNA sequencing by capillary electrophoresis

The ability of a polymer matrix to separate DNA by capillary electrophoresis (CE) is strongly dependent upon polymer physical properties. In particular, recent results have shown that DNA sequencing performance is very sensitive to both the average molar mass and the average coil radius of the separation matrix polymers, which are affected by both polymer structure and polymer-solvent affinity. Large polymers with high average molar mass provide the best DNA sequencing separations for CE, but are also the most challenging to characterize with accuracy. The methods most commonly used for the characterization of water-soluble polymers with application in microchannel electrophoresis have been gel permeation chromatography (GPC) and intrinsic viscosity measurements, but the limitations and potential inaccuracies of these approaches, particularly for large or novel polymers and copolymers, press the need for a more universally accurate method of polymer molar mass profiling for advanced DNA separation matrices. Here, we show that multi-angle laser light scattering (MALLS) measurements, carried out either alone or in tandem with prior on-line sample fractionation by GPC, can provide accurate molar mass and coil radius information for polymer samples that are useful for DNA sequencing by CE. Wider employment of MALLS for characterization of novel polymers designed as DNA separation matrices for microchannel electrophoresis should enable more rapid optimization of matrix properties and formulation, and assist in the development of novel classes of polymer matrices.

Keywords: Polymer characterization / Light scattering / DNA sequencing / Capillary electrophoresis
EL 4711

1 Contents

1 Introduction 4118
2 Uncross-linked polymer matrices for CE . . . 4120
2.1 DNA sequencing matrices based on linear homopolymers 4120
2.2 How polymer physical properties impact the characteristics of an entangled matrix . . 4120
3 Polymer characterization methods 4121
3.1 Viscometry 4121
3.2 Gel permeation chromatography 4121
3.3 Light scattering 4122
3.3.1 Batch multi-angle laser light scattering . . . 4122
3.3.2 Tandem GPC-MALLS 4123
4 Materials and methods 4123
4.1 Batch MALLS analyses 4123
4.2 Tandem GPC-MALLS analyses 4124
5 Analysis of LPA samples 4124
5.1 Batch MALLS analyses 4124
5.2 Tandem GPC-MALLS analysis 4125
6 Discussion 4126
7 Concluding remarks 4127
8 References 4127

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: LPA, linear polyacrylamide; **GPC**, gel permeation chromatography; **LALLS**, low-angle laser light scattering; **LS**, light scattering; **MALLS**, multi-angle laser light scattering; $\mathbf{M_n}$, number-average molar mass; $\mathbf{M_w}$, weight-average molar mass; **PDI**, polydispersity index; $\mathbf{R_g}$, root mean square coil radius

1 Introduction

In recent years, ultrahigh weight-average molar mass (M_w) linear polyacrylamide (LPA) with M_w values of 10–20 MDa has been demonstrated to have excellent properties as a separation matrix for DNA sequencing by microchannel electrophoresis [1–3]. In this application, fluid networks of highly entangled water-soluble polymers provide dynamic impediments to the electrophoretic migration of DNA frag-

≈ WILEY-VCH Verlag GmbH, 69451 Weinheim, 2001

0173-0835/01/1911–4118 $17.50+.50/0

ments. While the electrophoretic mobilities of DNA fragments in free solution are virtually independent of fragment size [4], migration of DNA chains through a dense mesh of entangled polymers provides frictional and mechanical interactions that lead to cyclic chain stretching and collapse and hence, size-dependent DNA separation [5]. Separation according to DNA chain length is, in turn, a necessary step in DNA sequencing, genotyping, and restriction mapping applications [6–8]. In particular, it is desirable for many sequencing applications to read the largest number of contiguous DNA bases possible, *i.e.*, to achieve "long reads" by capillary electrophoresis (CE), in the minimum amount of time. We consider that the present definition of a long read would be $>$ 800–1000 bases of contiguous DNA sequence read. Long-read DNA sequencing requires highly robust entangled polymer networks, as typically are formed by very long and very hydrophilic polymers, LPA being the best example [3, 9]. Entangled networks that are formed from lower molar mass polyacrylamides deliver shorter read lengths [10, 11]. The mechanistic explanation that we and others have put forward for this is that shorter polymers entangle less extensively, forming relatively weak networks that offer insufficient resistance to the migration of long DNA chains, which as a result do not separate well [1, 3, 9, 12].

The strong dependence of DNA sequencing read length on polymer molar mass which has been observed by many groups highlights the fact that until recently, the paramount importance of polymer physical properties for application in DNA separation by CE or capillary array electrophoresis (CAE) was not fully appreciated. The separation of double-stranded DNA (dsDNA) has also been shown to be sensitive to polymer molar mass, and interestingly, even to the breadth of the polymer molar mass distribution [13]. Yet polymer physical properties such as M_w, root mean square coil radius (commonly termed radius of gyration, R_g), and polydispersity index (PDI, defined as M_w divided by the number-average molar mass (M_n)) typically have not been known or determined for polymers that were employed as entangled networks (or "matrices") for DNA separation by CE. Generally, the chemical nature of the polymer has been reported in published CE work, but not the molar mass. When average molar mass was reported, researchers most often relied upon estimates that were provided by polymer manufacturers, which oftentimes are inaccurate, and generally have not had knowledge of polymer R_g or PDI. Only a few groups have determined polymer molar mass for useful CE matrices in-house, while determination of PDI has been more infrequent [14].

This history is understandable, since early (1989–1992) CE research towards development of separation technology for dsDNA fragments in entangled polymer solutions did not make clear the necessity for molar mass characterization of separation polymers. Fine, "single-base pair" (bp) resolution is not required in the analysis of dsDNA for most applications. A variety of different water-soluble polymers are able to deliver adequate resolution of dsDNA in the 25–1000 bp range. Generally, the optimal matrix concentration for each polymer sample has been determined on a case-by-case basis, and varies widely depending upon polymer chemical structure and average molar mass. Moreover, in the early days of CE research for DNA sequencing, researchers believed that it might be necessary to create a rigid, cross-linked hydrogel network within the capillary lumen, to play the well-understood role of a cross-linked polyacrylamide slab gel. Hence, the first attempts at forming intracapillary DNA sequencing matrices relied upon the use of cross-linking agents during acrylamide polymerization within the capillary [15–17]. This approach seemed to make sense, since in agarose and polyacrylamide slab gels, average pore size as determined by the ratio of monomer to cross-linker is of vital importance for optimizing the separation of DNA within a given size range [18]. It was soon realized, however, that cross-linked polyacrylamide gels had only limited usefulness when cast in a microchannel geometry, with the drawbacks to practicality usually outweighing any DNA resolution advantage gained by crosslinking.

Most of the early CE-based DNA sequencing separations made use of *in situ* polymerized matrices [19–24]; leaving out the cross-linker was a big step forward for the field. Heiger *et al.* [15] were the first to publish a study demonstrating the CE separation of DNA fragments in *in situ* polymerized matrices comprised of low- and zero-cross-linked polyacrylamide. Results showed that a reduction in the percentage of cross-linker in the *in situ* polymerization provided an increase in subsequent resolution of DNA restriction fragments. It was also observed that fluid, uncross-linked LPA networks were useful in resolving single-stranded oligonucleotides over a broad range of fragment sizes. This realization led to the investigation of a variety of uncross-linked polymer networks for DNA separation by CE, which could be replaced from the capillary under pressure. Most early studies of the separation of DNA sequencing fragments in such entangled polymer matrices did not include detailed characterization of the network polymers, as it was presupposed that pore size or mesh size rather than polymer molar mass was important. It was not clear how pore size might depend on polymer properties.

Given the experimental evidence that has been presented since that time by many groups of the impact of polymer physical properties on matrix performance [9], good methods of characterizing these polymer properties have clearly become important. Hence, we focus this review on a discussion of optimal methods for the characterization

of polymer matrices which are to be applied as DNA sequencing matrices for CE. Three major practical methods for profiling the physical properties of water-soluble polymers are discussed, and to provide some good examples we present experimental results that demonstrate the importance of achieving accurate characterization of even very high molar mass polymers that are useful for DNA sequencing. We show that multi-angle laser light scattering (MALLS), carried out either in tandem with polymer fractionation by GPC or as a stand-alone method, is the most accurate and useful technique that can be used for the characterization of polymers with application as separation matrices for genetic analysis.

2 Uncross-linked polymer matrices for CE

2.1 DNA sequencing matrices based on linear homopolymers

Early dsDNA separations in entangled (uncross-linked) polymer solutions were performed in LPA matrices that were polymerized within the capillary lumen which has an internal volume of ~ 1 µL [15, 16, 25–32]. Hence, the resultant polymers were available in only miniscule amounts and could not be characterized. In later work, M_w was estimated for LPA matrices that had been polymerized in the laboratory, and then pushed into capillaries under pressure, through determination of the intrinsic viscosity of the polymers [33] or by GPC alone with comparison to polymer size standards [34].

The first clear recognition of the importance of polymer molar mass for CE separations of DNA molecules came in 1993. In a study of the impact of polymer molar mass and concentration on dsDNA separation, it was shown that uncross-linked hydroxyethylcellulose (HEC) polymers with higher average molar mass provide much enhanced CE separation of larger dsDNA fragments ($>$ 610 bp) in comparison to low-molar-mass HEC polymers [35]. This result was interpreted in terms of the properties of the entangled polymer network (dragging of the polymers by the migrating DNA chains), rather than a "pore size". It was pointed out that pore size is a concept that is not as easily applicable to an entangled polymer network, due to the dynamic nature of the network, as it is to a rigid, porous, cross-linked hydrogel [35, 36]. Researchers began to move away from framing discussions of CE matrices mainly in terms of average pore size, and instead began to concentrate on describing the physical properties of the entangled polymer network as well [37], as they are directly influenced by the physical properties (M_w, R_g, PDI) of the polymers that form the separation matrix.

With the gearing up of the Human Genome Project in the early 1990's, the development of improved, replaceable polymer matrices and CE systems for the separation of DNA sequencing fragments became extremely important. DNA sequencing is a more demanding application than separation of dsDNA, requiring very high-resolution separations that, it was soon shown, can only be provided by highly entangled polymer solutions [9]. Experimental evidence suggests that the ability of the constituent polymer chains of the matrix to form a robust, strongly entangled network is the primary factor that controls the DNA sequencing performance of the polymer matrix [3].

Many different linear homopolymers besides LPA [1, 10, 12], including polyethylene oxide [38, 39] (PEO), poly(*N,N*-dimethylacrylamide) [5, 40–42] (pDMA), and HEC [33, 35], have been used with varying degrees of success as DNA sequencing matrices for CE. It has been found that the longest read lengths (in the shortest times) are achieved in high-M_w LPA solutions [1, 3, 12]. In a recent report, 1300 DNA bases were sequenced in 2 h by CE in a matrix composed of 2.0% w/v ultrahigh M_w LPA mixed with 0.5% w/v of a lower-M_w LPA [1]. The addition of low-M_w LPA to the polymer blend serves to elevate the total polymer concentration in the matrix without increasing its viscosity much, thereby improving the separation of short DNA fragments without making the matrix too viscous to push into a 75-µm inner diameter capillary [36]. The high-M_w LPA which forms the basis of the matrix is necessary as the primary network constituent, providing good separation of large DNA sequencing fragments and hence, long reads. The ultrahigh-molar-mass LPA that delivered this extraordinary performance was synthesized by inverse (water-in-oil) emulsion polymerization [43], and was characterized by our group using methods that we will discuss below.

2.2 How polymer physical properties impact the characteristics of an entangled matrix

An individual polymer chain adopts a random coil configuration in solution that, on average, occupies a spherical volume when the coil is in an unperturbed state. The average polymer coil radius is related to the distribution of mass within the molecule, and may be quantified as the average distance from the scattering center in the polymer to the end of the polymer chain (the mass moment). Thus, polymer coil radius depends upon the persistence length, or intrinsic chain stiffness, of each polymer as determined by both backbone and sidechain structures, as well as upon the extent to which solvent penetrates the coil [44]. It is important to recognize that on these bases, polymer coil mass moment will differ significantly for different types of water-soluble polymers that have different structures

and greater or less hydrophobicity. There is a higher degree of solvent penetration into a hydrophilic coil when compared to a more hydrophobic polymer coil, and hence the more hydrophilic polymer will adopt a more open configuration in water with a larger average coil radius.

Above a certain polymer concentration in solution, polymer coils begin to interpenetrate, overlap, and entangle. When a concentration is reached at which these entanglement interactions between polymer chains begin to influence bulk solution properties (such as solution viscosity) in a nonlinear or cooperative fashion, the solution has reached its overlap threshold, c^*, above which the polymers form an infinite entangled network [33, 35, 36]. The observed overlap threshold concentration will be directly related to the M_w and R_g of a given polymer sample [45]. At concentrations exceeding c^*, polymers will form a more densely and strongly entangled network [46]. The mechanical robustness of a matrix in providing resistance to DNA migration is determined by the number of entanglements per chain (the entanglement density); the average length of the polymer chain between entanglements is known as the "blob size" of the network [3]. Generally, the larger the polymer coil radius at a given M_w, and thus the lower the coil density, the more strongly entangled is the matrix at a given concentration in relation to the overlap threshold (c/c^*) [3]. Hence, at a given M_w, greater polymer hydrophobicity will also translate to a higher overlap threshold concentration because each individual polymer adopts a denser coil [3]. There are three practical methods for determination of the physical properties of water-soluble polymers with application in CE. These include viscometry, GPC, and light scattering (LS). The following discussion compares in some detail these three approaches to polymer characterization and their applicability to polymer classes important in CE.

3 Polymer characterization methods

3.1 Viscometry

A relatively simple and highly empirical polymer characterization method that has been widely used to estimate the average molar masses of polymer samples is based upon a determination of the intrinsic viscosity of a polymer in solution, through correlation and extrapolation of viscosity *vs.* concentration data [47]. Although modern rotational viscometry instruments can be used for these measurements, a simple Ostwald or Ubbelohde viscometer is more typically applied [48]. Polymer molar mass can be empirically related to the intrinsic viscosity through the Mark-Houwink-Sakurada relationship

$$[\eta] = KM_v^a \quad (1)$$

where $[\eta]$ is the intrinsic viscosity, K and a are empirical constants that are specific for a given polymer, solvent, and temperature, and M_v is the viscosity-average molar mass [48]. This method is restricted to the analysis of polymer samples for which Mark-Houwink-Sakurada constants are tabulated (which are relatively few), or have been measured in the same or a closely related solvent [49]. Without these tabulated or measured constants, of which a is the most important, Eq. (1) cannot be applied with accuracy. Hence, intrinsic viscosity has limited usefulness for novel polymers and copolymers that are being developed specifically for CE [47]. Due to differences in mathematical averaging of the data, M_v as measured by viscometry is typically somewhat lower than M_w. A rotational viscometer can simplify the measurements and a recent paper has suggested that intrinsic viscosity is directly related to the sieving properties of polymers useful for CE [50].

3.2 Gel permeation chromatography

GPC is the most common method of polymer characterization in use in laboratories today. In this method, a dilute polymer solution is injected into a solvent stream, which then flows under pressure through chromatography columns filled with porous gel packing typically composed of silica beads and/or a polymeric gel. As solvent flow drives the sample through the GPC column, the largest molecules pass through the column and past the detector most quickly, followed by smaller ones in order of coil size. Polymers with R_g greater than the pore radius may be completely excluded from the pores in the packing, in which case they will not be fractionated. For fractionation of a highly polydisperse sample, different columns can be used in series.

Average molar mass can be estimated by comparing the elution time of a given polymer peak to a calibration curve that has been generated with polymer size standards. This approach can be fairly accurate if the standard polymers are very similar in chemical structure and monomer composition to the analyte polymer, and are analyzed under identical solvation conditions. While this method is useful for an estimation of molar mass, it is intrinsically empirical and cannot necessarily be used quantitatively nor accurately to determine the molar mass distribution [43, 51]. An inherent assumption in this method of molar mass estimation is that the analyte polymer travels through the column at the same velocity as would a given polymer size standard of the same molar mass. This is clearly an approximation, since coil size and specific mode of migration is dependent not only on molar mass, but also on polymer persistence length and on the extent of coil solvation. Hence, GPC can be an inaccurate method of molar mass determination, especially if the

backbone or sidechain structures of the analyte polymers differ significantly from those properties of the polymer size standards.

3.3 Light scattering

Static LS, the third major method of polymer characterization, can be used to make an absolute measurement of weight-average polymer molar mass and polymer coil radius [52]. In the most common LS experimental setup, a dilute polymer solution flows through a laser photometer flow cell, where photomultiplier tube detectors are positioned along both sides of the flow cell at different angles to record intensity data as photons are deflected by interaction with the molecules in the stream. Often, the effluent from the scattering cell flows into an interferometric refractometer that measures the change in the refractive index of the solution as a function of time, hence giving a measure of polymer concentration.

Different types of LS instruments can be used for polymer analysis. Low-angle laser light scattering (LALLS), which measures the intensity of scattered light at a single, low angle of detection (6–7°) can be used to estimate the M_w of polymers, but because of the lack of data at other angles LALLS is not very accurate for polymers large enough to be useful for DNA sequencing by CE [13, 53]. MALLS measurements, on the other hand, with detection of scattered light at up to 18 different angles, can give accurate values of M_w and R_g for polymers of extremely high molar mass [54]. Moreover, if MALLS is employed in tandem with on-line sample fractionation by GPC, it is useful for the precise determination of molar mass and coil radius distributions of polymer samples [54].

Polymers that are too large in coil radius to be properly fractionated by GPC columns cannot be analyzed accurately by a tandem GPC-MALLS system. These very high molar mass samples must be analyzed by batch MALLS to allow the estimation of M_w. In this case, MALLS analysis is done without prior fractionation of the polymer sample by GPC, and hence it is not possible to determine the breadth of distribution. Generally, polymers with coil radii of gyration greater than 185 nm escape fractionation by GPC, due to a lack of the availability of columns with sufficiently large pore sizes. In our experience, this radius of gyration corresponds to a M_w of 8 MDa for LPA, which is on the lower end of the LPA size range that has proven to be the most effective for long-read DNA sequencing.

3.3.1 Batch MALLS

To determine weight-average polymer molar mass by batch MALLS, it is necessary to implement a data analysis and plotting technique developed by Zimm [55]. Zimm initially considered the scattering of light in a volume of isotropic molecules [56]. He further developed his theory to derive an expression for dilute solutions of macromolecules [55]. Commonly referred to as a Zimm plot, it is a graphical representation of the following equation:

$$\frac{R_\theta}{K^*c} = M_w P(\theta) - 2A_2 c M_w^2 P^2(\theta) \quad (2)$$

where c is the mass concentration of the solute macromolecules in the solvent (g/mL), A_2 is the second virial coefficient (mol · mL/g^2), $P(\theta)$ is a form factor relating to the dependence of scattered light intensity on angle, R_θ is the excess Rayleigh ratio (cm^{-1}), K^* is an optical constant that is equal to $4\pi^2\eta_o^2$ $(dn/dc)^2\lambda_0^{-4}N_A^{-1}$, where η_o is the refractive index of the solvent at the incident radiation (vacuum) wavelength, λ_0 is the incident radiation (vacuum) wavelength (nm), N_A is Avogadro's number (mol^{-1}), dn/dc is the differential refractive index increment (mL/g).

R_θ/K^*c data taken at a number of different polymer concentrations are extrapolated to zero angle, while the angular data for a number of different angles are extrapolated to zero concentration. Any number of angles which provide a reliable fit to the data and hence a reliable extrapolated value is acceptable. The weight-average molar mass of the polymer sample is then determined by taking the reciprocal of the common intercept of the two extrapolated curves [54]. The initial slope of the zero-angle extrapolation curve yields the second virial coefficient, which is a measure of the strength of solvent-solute interaction. The initial slope of the zero-concentration curve, on the other hand, yields the z-average root mean square radius for the polymer coil.

Different plotting formalisms of the equation may be used to correlate the data (termed the Zimm, Debye, and Berry formalisms) to provide options for fitting the LS data most closely, so as to obtain the most accurate possible estimate of what the scattering intensity would be at zero angle and zero polymer concentration [54]. The Zimm formalism is naturally more linear than the Debye or Berry formalisms and a lower-order polynomial often can be used to fit the data. Generally, the lower the order of the polynomial fit, the more reliable is the extrapolation [54]. As the molar mass of a sample is determined from the reciprocal of the extrapolated intercept at zero angle and zero concentration, a sample with a very high molar mass will have a very low reciprocal value. If the reciprocal value is close to zero, a slight amount of signal noise can yield an extrapolated intercept that is negative. On the other hand, the Debye fitting formalism utilizes curvature in the angular dependence for larger molecules [57]. To fit the angular dependence data, a high-order polynomial (3 or higher) must be used with all of the angles or a second

order polynomial can be used with only the lowest 4 or 5 "good" angles (*i.e.*, those angles with low noise from the scattering of dust). The Berry formalism has less curvature than the Debye formalism, and has been shown to give good fits to data sets for polymers with very high molar masses [58]. Data from a greater number of detectors can be used with this fitting method, to give a more accurate extrapolation of the data.

3.3.2 Tandem GPC-MALLS

The placement of a GPC system on the front end of an online MALLS detection system allows a fractionated sample to be quantitatively characterized. The effluent from the GPC system flows into a laser photometer-interferometric refractometer system. Complete fractionation of the sample permits any given thin slice of data to be considered as monodisperse, enabling the mass and size moments to be determined over the entire sample peak for different data slices. Extrapolation of fits to Eq. (2) can be utilized to determine the molar mass and mean-square radius for each slice. This allows a distribution plot of the M_w values of the sample to be generated. The polydispersity of the sample can thus be represented graphically, which provides a more complete characterization of the sample in comparison to the numerical values of M_w and R_g alone. This also allows a direct comparison of distributions of samples that may have similar M_w values but different degrees of polydispersity. It is also possible with data generated by tandem GPC-MALLS to calculate M_n, and hence to calculate the polydispersity index of the polymer sample. As M_w is never smaller than M_n for a given polymer sample, the PDI can never be less than 1.0. For polymers made by free-radical polymerization, PDI values of 1.4–1.8 are most common.

The light scattered by a given polymer is proportional to its molar mass as well as to its mass concentration, as predicted by Eg. (2), provided that the coil is scattering light as a single entity [54]. Therefore, the concentration of polymers in solution must be well below the overlap threshold concentration c^*, so that each individual coil is individually interrogated by the LS system. A large particle in the solution, such as a trace amount of dust or a sample aggregate, will tend to dominate the scattering signal. The intense scattering of light by dust is most markedly evident at low scattering angles, where one also gains $P(\theta)$ information about large polymers, hence it is very important to use good sample preparation and handling procedures to minimize dust and remove polymer aggregates. Clean, dilute, particle-free polymer samples and solvents are critical to successful LS measurements, and can be particularly difficult to obtain for aqueous polymer solutions. Below, we demonstrate the application of batch MALLS and tandem GPC-MALLS to polymers that have been shown to be highly useful for DNA sequencing by CE. The precise characterization of these and other new classes of polymers designed as DNA separation matrices for microchannel electrophoresis will enable more rapid optimization of matrix properties and formulation.

4 Materials and methods

4.1 Batch MALLS analyses

Weight-average molar masses of high molar mass polymer samples ($M_w > 8$ MDa, $R_g > 185$ nm) were determined by batch MALLS without prior GPC fractionation using a DAWN DSP Laser Photometer-Optilab DSP Interferometric Refractometer system (both, Wyatt Technology, Santa Barbara, CA, USA). The DAWN system was normalized with a protein, bovine serum albumin, which has the important virtues of being monodisperse, having a precisely known molar mass, and being small enough that it scatters light isotropically. The Optilab detector was calibrated with sodium chloride solutions of known concentration and refractive index. Hence, the entire DAWN-Optilab system is calibrated absolutely. For each batch MALLS analysis, stock solutions of the polymer solutions were prepared at concentrations of 1×10^{-5}–1×10^{-4} g/mL (accurately determined to three significant figures) in 18.0-MΩ purified H_2O from a Barnstead E-Pure system (Fisher Scientific, Glenlake, IL, USA), where the deionized water used to dissolve the polymers was first passed through 0.02 μm filters (Whatman, Maidstone, England) to remove particulates. Polymer stock solutions were mixed by slow rotation on a Roto-Torque mixer (Cole-Parmer, Vernon Hills, IL, USA) for 24 h. Aliquots of the stock solution then were diluted with prefiltered solvent into precleaned scintillation vials (Fisher Scientific), and spun on the Roto-Torque mixer. All samples were made using a high-precision balance to allow accurate calculation of concentration. A syringe pump (kdScientific, New Hope, PA, USA) was used to push the samples through 0.22 μm syringe filters into the DAWN-Optilab system, with a new filter used for each polymer concentration. For each polymer sample (*i.e.*, at each concentration), the instrument was used to measure the intensity of the scattered light as a function of angle for 16 different fixed angles. For a given sample, data are collected for about 15 min, until a plateau is established for the detector output. Pure solvent is injected first, followed by remaining polymer samples sequentially in order of increasing concentration. Typically, 6–10 different polymer concentrations were analyzed for each Zimm plot. Pure solvent was the first and last injection to set

the baseline for analysis. After the data are collected, the known concentrations are assigned to each plateau region, and Wyatt Technology ASTRA software is used to process the data and create a Zimm plot. A subset of the 16 angles is chosen for data fitting to minimize the effect of noise, with data from no fewer than 10 angles utilized for the final results.

4.2 Tandem GPC-MALLS analyses

Lower-molar-mass polymer samples ($M_w < 6$ MDa, $R_g <$ 185 nm) were analyzed by tandem GPC-MALLS. Samples were dissolved at a concentration of 0.5 mg/mL into the aqueous mobile phase of the GPC system (0.1 M NaCl, 50 mM NaH_2PO_4, and 200 ppm NaN_3). The samples were fractionated by GPC prior to on-line MALLS and refractive index detection, using a Waters 2690 Separations Module (Milford, MA, USA) with Shodex (New York, NY, USA) OHpak columns SB-806 HQ, SB-804 HQ, and SB-802.5 HQ connected in series. In this tandem GPC-MALLS mode, effluent from the GPC system flows into the DAWN -Optilab system. Tandem GPC-MALLS data were processed using ASTRA software from Wyatt Technology. 100% mass recovery was used for processing the data. Experiments have demonstrated (data not shown) that at least in the case of LPA analysis, assumption of 100% mass recovery from the GPC columns and inputting of known d*n*/d*c* values can both be used with similar results in processing the data.

5 Analysis of LPA samples

5.1 Batch MALLS analyses

To carry out method validation for the batch MALLS system, a high-molar mass linear polyacrylamide standard (American Polymer Standards Corp., Mentor, OH, USA) was analyzed. The LPA standard used has a manufacturer-specified M_w of 5.55 MDa. The standard was analyzed *via* batch methods (data not shown) and was determined to have a M_w of 5.74 MDa +/– 0.13 MDa, hence giving a difference between the two values of only 3.4%. This level of accuracy was deemed to be more than adequate for this method of characterization.

To demonstrate the capabilities and limitations of LS in the two analysis modes (*i.e.*, in tandem with GPC and batch), an LPA sample that lies at the upper limit in size of the fractionating ability of the GPC columns was analyzed by both methods. The results demonstrate that if the polymer cannot be fully fractionated by GPC, one obtains a dissimilar weight-average molar mass estimation from the two different analysis methods. The plot shown in Fig. 1a demonstrates the poor peak shape that can result from incomplete GPC fractionation of the polymer sample, as evident by the non-Gaussian shape of the distribution. The peak is sharply fronted as a result of a major fraction of the larger polymers in the sample escaping fractionation. Moreover, the apparent polydispersity index of this sample according to tandem GPC-MALLS analysis is 1.035, a value that is wholly unrealistic for the synthesis conditions used to make the polymers (free-radical polymerization). Tandem GPC-MALLS provided an M_w estimate of 8.2 MDa, while batch MALLS yields a higher value, 8.9 MDa. The difference between these two values is a consequence of the incomplete fractionation of the sample; the batch MALLS analysis is accurate, whereas that of the GPC-fractionated sample is not. Figure 1b shows the experimentally obtained R_g dis-

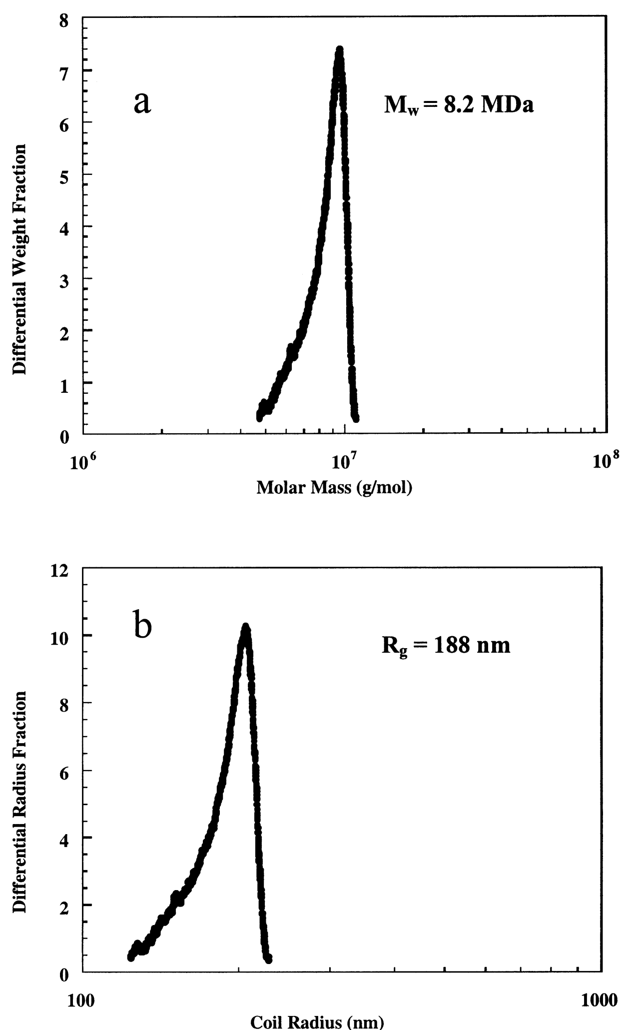

Figure 1. Tandem GPC-MALLS characterization of a polyacrylamide sample that lies at the upper limit in size of the fractionating ability of the GPC columns. (a) Molar mass distribution; (b) coil radius distribution.

tribution of the same polyacrylamide sample shown in Fig. 1a. The shape of the peak is similar to that in Fig. 1a, being asymmetrical and showing a sharp cutoff in fractionation of the larger molecules as the coil radius of the molecules approach 190 nm. Our experiments have consistently shown that for a polymer sample to be properly fractionated by GPC using Shodex OHpak columns, the average coil radius must be less than 185 nm. This result clearly makes evident the need to analyze large polymers, which are most useful for DNA analysis by CE, by batch MALLS to obtain accurate results.

The Karger research group (Northeastern University, Boston, MA, USA) [12, 43] synthesizes ultrahigh-molar-mass LPA for DNA sequencing by CE using inverse-emulsion free radical polymerization. Polymers produced by this method have been utilized in DNA sequencing matrices whose performance is far superior to previous results from matrices based on LPA made by aqueous solution-phase free-radical polymerization [1, 12]. These high-performance, ultrahigh-molar-mass polymers could not be characterized accurately by GPC, as the size-exclusion columns on the system excluded the polymers to a significant degree [43]. We have collaborated with the Karger research group to characterize these LPA polymers by batch MALLS. This high-molar mass LPA sample was analyzed with the batch MALLS system, and a Zimm plot was generated. The full analysis was done in triplicate to ensure reproducibility. As shown in Fig. 2a, the M_w of this sample was determined to be 10.4 MDa +/– 0.4 MDa, while R_g was found to be 164.1 nm +/– 4.2 nm. A DNA sequencing matrix composed of this high-molar-mass LPA at 2.0% w/v, in a mixture with 0.5% low-molar-mass LPA (50 kDa), was the first CE matrix reported to give a sequencing read of 1000 bases in under 1 h [12]. The initially published, estimated molar mass value for this polymer was 9 MDa, based on GPC [12, 43]. Batch MALLS analysis reveals a true molar mass of 10.4 MDa. The major drawback to batch MALLS, however, is that in the absence of polymer fractionation M_n and PDI cannot be determined.

It was found by the Karger group [1] that the use of an even higher molar mass LPA in a DNA sequencing matrix, again mixed at 2% high-molar-mass and 0.5% low-molar-mass LPA, provided yet another major increase in the sequencing read of the matrix of up to 1300 bases in 2 h. In Fig. 2b, a Zimm plot of data obtained by batch analysis of this second ultrahigh-molar-mass LPA sample showed the polymers to have a weight-average molar mass of 17.1 MDa +/– 0.5 MDa and an R_g of 189.0 nm +/– 2.9 nm. Importantly, the increase in read length attained with this second-generation LPA formulation could not have been predicted by matrix viscosity (both

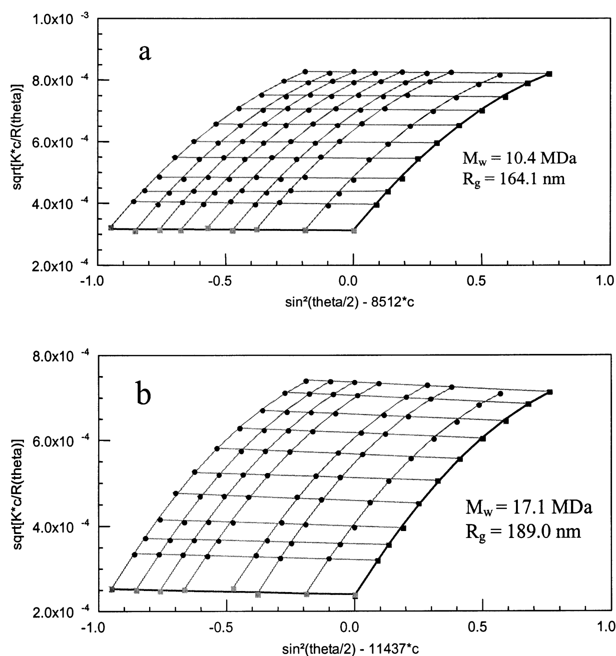

Figure 2. Zimm plot analysis (Berry formalism of 2nd degree) of high molar mass LPA provided by the Karger research group. The horizontal lines represent data for angles 35°, 43°, 52°, 60°, 69°, 80°, 90°, 100°, 111°, and 121°. The rightward-sloping lines represent data for LPA concentrations. The thick lines show the data extrapolated to zero angle and zero concentration. A stretch factor is applied to spread out the data and produce a detailed plot. (a) LPA concentrations 0.0223–0.111 g/L, stretch factor 8512; (b) LPA concentrations 0.016–0.083 g/L, stretch factor 11437.

matrices being extremely viscous) without knowledge of M_w. Moreover, this polymer sample was too large in average coil radius to allow estimation of its M_w by GPC. There most likely is a critical point where an increase in matrix polymer M_w will not increase the read length of the matrix, but that point has not been reached as of yet. Read lengths continue to be improved with each increase in the M_w of the polymer samples as the corresponding matrices are more robust. This result demonstrates the importance of polymer molar mass for DNA sequencing matrices, and hence the critical need for accurate M_w determination in the formulation of high-performance matrices for CE.

5.2 Tandem GPC-MALLS analysis

To achieve method validation for the tandem GPC-MALLS system, we analyzed a commercially available LPA molar mass standard (American Polymer Standards) to ensure that the DAWN-Optilab system had been cali-

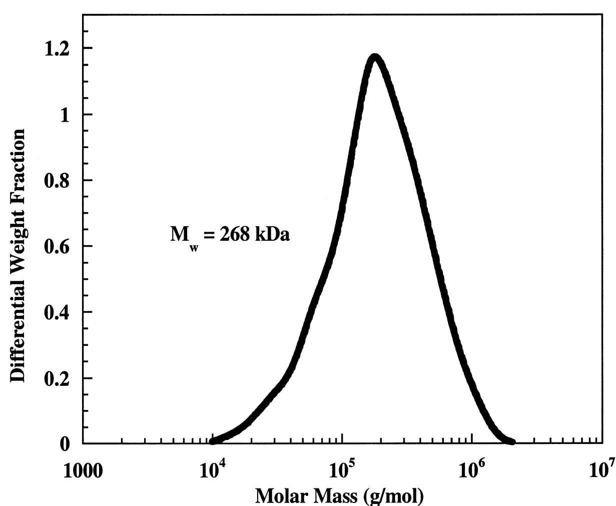

Figure 3. The results of a tandem GPC-MALLS molar mass characterization of a low-M_w LPA.

brated accurately. The LPA standard used has a manufacturer-specified M_w of 990 kDa. The calculated M_w was determined to be 993 kDa (data not shown), demonstrating an error (either ours or theirs) of 0.3%. Complete fractionation of the LPA standard was evidenced by a Gaussian shape of the overall data curve of differential weight fraction *vs*. molar mass (data not shown).

Related to the same Karger lab DNA sequencing matrix discussed above [1, 12], an LPA sample with an unknown but relatively low M_w was used as an additive to the matrix at a concentration of 0.5% (along with the high-M_w LPA at a concentration of 2.0%). The addition of the lower-molar mass LPA increases the selectivity of the matrix for smaller DNA fragments without causing a significant reduction in the resolution of the large DNA fragments [1]. The increase in total matrix concentration was also advantageous in that there was only a minimal corresponding increase in solution viscosity.

This low-molar-mass LPA sample was also provided to us by the Karger research group, and was analyzed by tandem GPC-MALLS. The data output has a relatively Gaussian distribution as seen in Fig. 3, demonstrating good fractionation of the polymers by the three GPC columns in series. The polymer was determined through analysis of the data to have a molar mass average of 268 kDa and a polydispersity of 2.1. Even though this LPA sample has a much lower molar mass than the other component of the matrix, its inclusion is necessary to improve the selectivity of the matrix for smaller DNA fragments, and in the process increase the total read length generated by the matrix on the low end of the DNA size distribution [1].

6 Discussion

We found that it required substantial experimental investigation to determine the correct analysis conditions and sample preparation methods to achieve convincingly reproducible and clearly interpretable molar mass estimates for the high-molar mass LPA samples. We were not much helped by the polymer literature, since accurate and precise characterization of high molar mass water-soluble polymers ($M_w > 8$ MDa) has not been of great technological importance prior to their use for CE. This highlights the fact that accurate measurement of M_w and the molar mass distribution of large polymers such as those useful for DNA sequencing by CE can be a nontrivial task. Intrinsic viscometry is likely to be difficult to apply with accuracy and can be very tedious depending upon the instrumentation used. In many cases, most particularly for the high-molar-mass polymers that can yield very long sequencing reads, polymer coil size is simply too large to allow good fractionation by GPC. In addition to limitations in the molar masses that can be analyzed accurately based on pore size, GPC relies on comparison to polymer size standards for molar mass estimation of an unknown sample. Difficulty in obtaining accurate results for some polymer types often arises because it is impossible to obtain appropriate size standards. Commercially available polymer standards are based upon only a select number of polymer classes, and will give less accurate results for novel or "designer" polymers that have different chemical structures.

We note that researchers are continually developing novel polymers and copolymers for application in CE analysis of DNA, which present ever greater challenges for characterization. For example, poly(*N,N*-diethylacrylamide)-*co*-poly(*N,N*-dimethylacrylamide) [59], *N*-acryloylaminoethoxyethanol [60, 61], poly(*N*-isopropylacrylamide)-*g*-poly (ethylene oxide) [62, 63], and poly(ethylene oxide)-poly (propylene oxide)-poly(ethylene oxide) triblock copolymer [64–67] are just a few novel copolymers that recently have been developed and studied for the separation of DNA fragments by CE. Clearly, optimal design of novel polymers and copolymers for high-performance bioseparations will demand their proper characterization, so that performance can be correlated with their properties. The absence of appropriate GPC size standards for these unusual copolymers presses the need for an alternative method of polymer mass determination, as each of these copolymers has a unique root mean square coil radius depending on its chemical structure. Work from several laboratories has demonstrated that the DNA sequencing capabilities of a polymer matrix are strongly dependent upon polymer physical properties, even for high molar mass polymers. Hence, accurate and precise characterization of polymer

physical properties is required for the design, optimization, and quality control of high-performance CE matrices for DNA analysis, and in our view LS is the best means by which to accomplish this.

7 Concluding remarks

Polymer physical properties are perhaps the most important determinant of the performance of a polymer matrix for DNA sequencing, and hence polymer characterization is a vital component of research and development for this technology. It has been demonstrated that a predominance of high-molar-mass polymers within a DNA sequencing matrix is essential for the generation of long DNA sequencing read lengths. The resolution of large DNA fragments is best achieved in relatively low concentrations (*e.g.*, 2–4%) of high-molar-mass polymers. Thus, a robust entangled polymer network is needed to separate large DNA fragments. The separation of smaller DNA sequencing fragments is less sensitive to polymer molar mass, and is improved by increasing the total polymer concentration of the sequencing matrix. The polydispersity of a polymer matrix is another indicator of its potential for good separation of long DNA fragments. The lower the polydispersity of the polymer sample at a given M_w, the greater the extent of polymer-polymer entanglement at a given solution concentration, and the more robust is the polymer network. To determine the PDI of a polymer sample, an accurate measurement of the molar mass distribution needs to be made, which can only be done by GPC or by tandem GPC-MALLS.

GPC is an effective method of polymer characterization for a limited subset of polymers that are useful for CE, primarily because of limitations in column packing technology and the lack of appropriate polymer standards in some cases. Therefore, a more versatile method such as tandem GPC-MALLS for polymers with coil radii of less than 185 nm, and batch MALLS for larger polymers, is needed to allow characterization of the physical properties of many classes of advanced polymers and copolymers for application in CE. We have shown here that MALLS is an effective method to characterize these novel polymers and copolymers. As the field advances, these data can be used to forecast the DNA sequencing capabilities of different polymer matrices. In turn, MALLS will expediate the optimization of the formulation of these novel matrices, not only for CE but also for DNA analysis on microfluidic devices. These methods of polymer characterization may be advantageous for other applications as well that would benefit from an accurate knowledge of the physical properties of water-soluble polymers and copolymers.

The authors acknowledge the assistance of Dr. Haihong Zhou, Dr. Lev Kotler, and Dr. Barry L. Karger of Northeastern University. Financial support for this work was provided by the National Institutes of Health NHGRI Grant #1 R01 HG01970, an NIH Presidential Early Career Award for Scientists and Engineers (1999), and the Beckman Foundation Beckman Young Investigator program (1998).

Received July 17, 2001

8 References

[1] Zhou, H., Miller, A. W., Sosic, Z., Barron, A. E., Buchholz, B., Kotler, L., Karger, B. L., *Anal. Chem.* 1999, *72*, 1045–1052.

[2] Koutny, L., Schmalzing, D., Salas-Solano, O., El-Difrawy, S., Adourian, A., Buonocore, S., Abbey, K., McEwan, P., Matsudaira, P., Ehrlich, D., *Anal. Chem.* 2000, *72*, 3388–3391.

[3] Albarghouthi, M. N., Buchholz, B. A., Doherty, E. A. S., Bogdan, F. M., Zhou, H., Barron, A. E., *Electrophoresis* 2001, *22*, 737–747.

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

[5] Heller, C., *Electrophoresis* 1999, *20*, 1962–1977.

[6] Schmalzing, D., Koutny, L., Salas-Solano, O., Adourian, A., Matsudaira, P., Ehrlich, D., *Electrophoresis* 1999, *20*, 3066–3077.

[7] Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Laviolette, J.-P., Ardlie, K., Reich, D. E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.-B., Gingeras, T., Warrington, J., Patil, N., Hudson, T. J., Lander, E. S., *Nature Genetics* 2000, *24*, 381–386.

[8] Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., Lander, E. S., *Science* 1998, *280*, 1077–1082.

[9] Albarghouthi, M. N., Barron, A. E., *Electrophoresis* 2000, *21*, 4096–4111.

[10] Ruiz-Martinez, M. C., Berka, J., Belenkii, A., Foret, F., Miller, A. W., Karger, B. L., *Anal. Chem.* 1993, *65*, 2851–2858.

[11] Grossman, P. D., *J. Chromatogr. A* 1994, *663*, 219–227.

[12] Salas-Solano, O., Carrilho, E., Kotler, L., Miller, A. W., Goetzinger, W., Sosic, Z., Karger, B. L., *Anal. Chem.* 1998, *70*, 3996–4003.

[13] Barron, A. E., Sunada, W. M., Blanch, H. W., *Electrophoresis* 1996, *17*, 744–757.

[14] Dolnik, V., Gurske, W. A., Padua, A., *Electrophoresis* 2001, *22*, 707–719.

[15] Heiger, D. N., Cohen, A. S., Karger, B. L., *J. Chromatogr.* 1990, *516*, 33–48.

[16] Cohen, A. S., Najarian, D. R., Karger, B. L., *J. Chromatogr.* 1990, *516*, 49–60.

[17] Drossman, H., Luckey, J. L., Kostichka, A. J., D'Cunha, J., Smith, L. M., *Anal. Chem.* 1990, *62*, 900–903.

[18] Righetti, P. G., *J. Biochem. Biophys. Methods* 1989, *19*, 1–20.

[19] Huang, X. C., Stuart, S. G., Bente, P. F. I., Brennan, T. M., *J. Chromatogr.* 1992, *600*, 289–295.

[20] Luckey, J. A., Drossman, H., Kostichka, A. J., Mead, D. A., D'Cunha, J., Norris, T. B., Smith, L. M., *Nucleic Acids Res.* 1990, *18*, 4417–4421.

[21] Luckey, J. A., Smith, L. M., *Anal. Chem.* 1993, *65*, 2841–2850.

[22] Manabe, T., Chen, N., Terabe, S., Yohda, M., Endo, I., *Anal. Chem.* 1994, *66*, 4243–4252.

[23] Luckey, J. A., Norris, T. B., Smith, L. M., *J. Phys. Chem.* 1993, *97*, 3067–3075.

[24] Carson, S., Cohen, A. S., Belenkii, A., Ruiz-Martinez, M. C., Berka, J., Karger, B. L., *Anal. Chem.* 1993, *65*, 3219–3226.

[25] Paulus, A., Hüsken, D., *Electrophoresis* 1993, *14*, 27–35.

[26] Yin, H. F., Lux, J. A., Schomburg, G., *J. High Resol. Chromatogr.* 1990, *13*, 624–627.

[27] Guttman, A., Nelson, R. J., Cooke, N., *J. Chromatogr.* 1992, *593*, 297–303.

[28] Cohen, A. S., Najarian, D. R., Paulus, A., Guttman, A., Smith, J. A., Karger, B. L., *Proc. Natl. Acad. Sci. USA* 1988, *85*, 9660–9663.

[29] Paulus, A., Ohms, J. J., *J. Chromatogr.* 1990, *507*, 113–123.

[30] Swerdlow, H., Wu, S., Harke, H., Dovichi, N. J., *J. Chromatogr.* 1990, *516*, 61–67.

[31] Kleemiß, M. H., Gilges, M., Schomburg, G., *Electrophoresis* 1993, *14*, 515–522.

[32] Chiari, M., Nesi, M., Fazio, M., Righetti, P. G., *Electrophoresis* 1992, *13*, 690–697.

[33] Grossman, P. D., Soane, D. S., *Biopolymers* 1991, *31*, 1221–1228.

[34] Gelfi, C., Orsi, A., Leoncini, F., Righetti, P. G., *J. Chromatogr. A* 1995, *689*, 97–105.

[35] Barron, A. E., Soane, D. S., Blanch, H. W., *J. Chromatogr.* 1993, *652*, 3–16.

[36] Bünz, A. P., Barron, A. E., Prausnitz, J. P., Blanch, H. W., *Ind. Eng. Chem. Res.* 1996, *35*, 2900–2908.

[37] Grosche, O., Bohrisch, J., Wendler, U., Jaeger, W., Engelhardt, H., *J. Chromatogr.* 2000, *894*, 105–116.

[38] Madabhushi, R. S., Vainer, M., Dolnik, V., Enad, S., Barker, D. L., Harris, D. W., Mansfield, E. S., *Electrophoresis* 1997, *18*, 104–111.

[39] Fung, E. N., Yeung, E. S., *Anal. Chem.* 1995, *67*, 1913–1919.

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

[41] Heller, C., *Electrophoresis* 1999, *20*, 1978–1986.

[42] Song, L., Liang, D., Fang, D., Chu, B., *Electrophoresis* 2001, *22*, 1987–1996.

[43] Goetzinger, W., Kotler, L., Carrilho, E., Ruiz-Martinez, M. C., Salas-Solano, O., Karger, B. L., *Electrophoresis* 1998, *19*, 242–248.

[44] Ha, B. Y., Thirumalai, D., *J. Phys. II.* 1997, *7*, 887–902.

[45] Broseta, D., Leibler, L., Lapp, A., Strazielle, C., *Europhys. Lett.* 1986, *14*, 668–676.

[46] Bueche, F., *Physical Properties of Polymers*, Wiley, New York 1962.

[47] Young, R. J., Lovell, P. A., *Introduction to Polymers*, 2nd ed., Chapman & Hall, New York 1991.

[48] Fried, J. R., *Polymer Science and Technology*, Prentice Hall PTR, Englewood Cliffs, NJ 1995.

[49] Brandrup, J., Immergut, E. H., Grulke, E. A., *Polymer Handbook*, 4 ed., Wiley, New York 1999.

[50] Dolník, V., Gurske, W. A., Padua, A., *Electrophoresis* 2001, *22*, 692–698.

[51] Li, F. C., *Chromatogr. Sci. Series* 1995, *69*, 249–277.

[52] Wyatt, P. J., *LC-GC* 1997, *15*, 160–168.

[53] Wyatt, P. J., Jackson, C., Wyatt, G. K., *Am. Lab.* 1988, *20*, 86–91.

[54] Wyatt, P. J., *Anal. Chim. Acta* 1993, *272*, 1–40.

[55] Zimm, B. H., *J. Chem. Phys.* 1948, *16*, 1093–1099.

[56] Zimm, B. H., *J. Chem. Phys.* 1945, *13*, 141–145.

[57] Debye, P., *J. Phys. Coll. Chem.* 1947, *51*, 18–32.

[58] Berry, G. C., *J. Chem. Phys.* 1966, *44*, 4550–4564.

[59] Buchholz, B. A., Doherty, E. A. S., Albarghouthi, M. N., Bogdan, F. M., Zahn, J. M., Barron, A. E., *Anal. Chem.* 2001, *73*, 157–164.

[60] Chiari, M., Micheletti, C., Nesi, M., Fazio, M., Righetti, P. G., *Electrophoresis* 1994, *15*, 177–186.

[61] Chiari, M., Nesi, M., Righetti, P. G., *Electrophoresis* 1994, *15*, 616–622.

[62] Liang, D., Zhou, S., Song, L., Zaitsev, V. S., Chu, B., *Macromolecules* 1999, *32*, 6326–6332.

[63] Liang, D., Song, L., Zhou, S., Zaitsev, V. S., Chu, B., *Electrophoresis* 1999, *20*, 2856–2863.

[64] Wu, C., Liu, T., Chu, B., *Macromolecules* 1997, *30*, 4574–4583.

[65] Wu, C., Liu, T., Chu, B., *Electrophoresis* 1998, *19*, 231–241.

[66] Liang, D., Chu, B., *Electrophoresis* 1998, *19*, 2447–2453.

[67] Song, L., Fang, D., Kobos, R. K., Pace, S. J., Chu, B., *Electrophoresis* 1999, *20*, 2847–2855.