*Journal of Chromatography A*, 659 (1994) 247–253
Elsevier Science B.V., Amsterdam

CHROM. 25 668

# Ensemble averaging and digital filtering in chromatography and electrophoresis

Michael F. Regehr, Sandeep K. Paliwal and Fred E. Regnier*

*Department of Chemistry, Purdue University, West Lafayette, IN 47907 (USA)*

ABSTRACT

Ensemble averaging and digital filtering were implemented for signal-to-noise ratio improvement in the separation techniques of size-exclusion chromatography, immunoaffinity chromatography, capillary zone electrophoresis and capillary ion analysis. Results of ensemble averaging were always greater than statistically predicted. Techniques included five to nine replicate separations and yielded signal-to-noise improvement factors of 2.5 to 9.3. Running-average and time constant (RC)-convolution digital filters yielded increases in the signal-to-noise ratio ranging from zero to twelve. This paper will discuss and illustrate the usage of ensemble averaging and digital filtering in liquid-phase separation techniques.

## INTRODUCTION

Efforts to lower detection limits are an ongoing trend in analytical chemistry. Sensitivity enhancement in liquid chromatography has been achieved by a variety of methods including the use of derivatizing agents, miniaturization, advances in electronics, optimization of absorbance detectors, and the introduction of laser fluorescence and electrochemical detection. Recent studies suggest that computational methods associated with data processing may be another method of increasing sensitivity.

Acquisition, storage and manipulation of digital data are becoming more accessible to the chromatography laboratory due to the increasing use of computers to automate chromatography systems. Recent trends in digital data handling can be found in the articles from "The Data File" column of Ouchi [1–5]. Digital data can be stored, displayed and analyzed regardless of the method used to perform a separation.

Data-handling software allows for easier display and reporting of chromatograms with the additional possibility of extensive computational analysis. The use of ensemble averaging and digital filtering to improve signal-to-noise ($S/N$) ratios in chromatographic and electrophoretic techniques will be examined here. Ensemble averaging is predicted to improve $S/N$ ratios as the square root of the number of chromatograms summed [6]. The problem with ensemble averaging in chromatography had been that considerable time was required to carry out the multiple runs needed for averaging. Recent advances in liquid chromatography media allow separations of biological macromolecules to be achieved an order of magnitude faster [7]. Digital filtering improvements are dependent on the nature of the noise and the filter applied to the data.

Ensemble averaging was applied to four different separation techniques with the goal of lower-

* Corresponding author.

0021-9673/94/$07.00 
*SSDI* 0021-9673(93)E1116-H

ing detection limits: size-exclusion chromatography for lysozyme, immunoaffinity chromatography for bovine IgG, capillary zone electrophoresis for $\alpha$-amylase and capillary ion analysis for benzyltriethylammonium chloride.

## EXPERIMENTAL

### *Materials*

Potassium phosphate, magnesium chloride, sodium chloride and sodium sulfate were purchased from Mallinckrodt (Paris, KY, USA). Lysozyme, bovine IgG, $\alpha$-amylase and Tris (TRIZMA) base were purchased from Sigma (St. Louis, MO, USA). Acetic acid was purchased from J.T. Baker (Phillipsburg, NJ, USA). Mesityl oxide and benzyltriethylammonium chloride were purchased from Aldrich (Milwaukee, WI, USA).

### *Size-exclusion chromatography (SEC)*

Liquid chromatography was performed on the BioCAD HPLC system, a gift from PerSeptive Biosystems (Cambridge, MA, USA). SEC was achieved with an $300 \times 8$ mm column packed with TSK G3000SW packing material. Columns were packed according to manufacturers instructions on an HPLC column packer (Shandon Southern Instruments, Sewickley, PA, USA). The buffer contained 100 m*M* potassium phosphate and 100 m*M* sodium sulfate at pH 7.0. Flow-rate was 0.5 ml/min. The analyte was lysozyme (1 mg/ml). Absorbance was monitored at 254 nm. Chromatograms were transferred to Quattro Pro (Borland, Scotts Valley, CA, USA) and referenced by the point of injection.

### *Immunoaffinity chromatography (IAC)*

IAC was performed on the BioCAD HPLC system with a $30 \times 4.6$ mm POROS A/M column (a gift from PerSeptive Biosystems). Loading buffer for the affinity column was the same as that for the SEC separation. Desorption buffer contained 0.3 *M* magnesium chloride and 2% acetic acid. Flow-rate was 3.0 ml/min. The bovine IgG analyte (1 mg/ml) was detected at 254 nm. Chromatograms were transferred to Quattro Pro and referenced by the point of injection.

### *Capillary zone electrophoresis (CZE)*

CZE experiments were performed on the ISCO Model 3850 capillary electrophoresis system (ISCO, Lincoln, NE, USA). Data collection was achieved with a 486/33 personal computer (IBM compatible) running DAQWare to drive a PC-LPM-16 multifunction input/output board (National Instrument Corporation, Austin, TX, USA). The fused-silica capillary (Polymicro Technologies Inc, Phoenix, AZ, USA) was 45 cm total length (25 cm separation length) $\times$ 75 $\mu$m I.D. Applied voltage was 178 V/cm. Running buffer contained 50 m*M* TRIZMA base and 120 m*M* sodium chloride at pH 7.0. The analytes were mesityl oxide (1 mg/ml) and $\alpha$-amylase (45 $\mu$g/ml) in TRIZMA–NaCl buffer. Detection was at 214 nm. Hydrostatic injections were made by 10 cm vertical displacement. Injection volumes were estimated hydrostatically. The instrument was set to 0.8 s rise times (10–90% response) and 0.005 AUFS output. Electropherograms were transferred to Quattro Pro and referenced on the mesityl oxide peak.

### *Capillary ion analysis (CIA)*

CIA was performed on the ISCO Model 3850 capillary electrophoresis system utilizing the same data collection system and instrument settings noted above. The fused-silica capillary was 40 cm total length (20 cm separation length) $\times$ 103 $\mu$m I.D. The applied voltage was 300 V/cm. Running buffer contained 25 m*M* TRIZMA base at pH 7.0. The benzyltriethylammonium chloride analyte (30 n*M* in distilled water) was detected at 210 nm. Electrokinetic injections were performed at 12 kV for 20 s. Electropherograms were transferred to Quattro Pro and referenced to the benzyltriethylammonium migration peak.

## RESULTS AND DISCUSSION

### *Ensemble averaging*

Ensemble averaging has proven to be useful to increase $S/N$ ratios in gas chromatography [6,8]. It involves repeated separations and the numerical addition of chromatograms. The advantage of this approach is that peak heights increase proportional to the number of chromatograms sum-

med ($n$), while amplitude of the baseline noise will only increase as the square-root of $n$ [6]. Ensemble averaging four chromatograms should double the $S/N$ ratio. Applications to liquid chromatography and electrophoresis have been hindered by the generally longer separation times required. The advent of perfusion chromatography and high-performance capillary electrophoresis have reduced separation times to a point where the time invested in ensemble averaging can be worth the effort.

*Digital filtering*

Digital filtering was performed within a spreadsheet by two algorithms: two-column and convolution. In the two-column method, mathematical equations indicating the desired derivation were placed alongside the column of data. The convolution algorithm required the establishment of a functional worksheet.

The running-average filter, the simplest application of the two-column algorithm, derives each filtered datapoint by averaging all values between a specified distance before and after the point being evaluated. The cells of the filtered-data column either contain an algebraic equation quoting each necessary cell of the original column, or utilize a spreadsheet function. Each datapoint is given equal weight in this evaluation. For example, a five-point running-average filter sums points $x-2$ through $x+2$ and normalizes with division by five.

A Savitsky–Golay filter [9] finds the filtered-data value from the least-squares fit to the portion of data under consideration. A weighted contribution scheme can be derived mathematically based on the dimensions of the data segment evaluated. For example, a five-point Savitsky–Golay filter is weighted ($-3$,12,17,12, $-3$) with a normalization constant of 35 [10].

The convolution algorithm utilizes spreadsheet functions or macros to approximate the convolution integration shown below.

$$h(x) = \int f(u)g(x-u)\,du$$

In the algorithm developed for our study, the Quattro Pro "sumproduct" function was used, multiplying each point of the convolving function (g) with the corresponding point of the data (f), and summing the products. The algorithm must be written for a specific length of data set and convolving function. Running-average and Savitsky–Golay filters can be run in a convolution algorithm by using a data set containing the appropriate weighing factors in place of the convolving function.

*Ensemble averaging applications*

Capillary electrophoresis has been widely applied to the separation of charged molecules. Narrow, fused-silica capillaries provide rapid heat dissipation and suppression of induced convection [11,12]. This allows application of larger axial field strengths. The result is a rapid, high-resolution separation. Application of ensemble averaging requires that migration times are reproducible over multiple runs [6]. Variation would broaden the ensemble-averaged peak and give $S/N$ gains lower than expected. Two examples of compensation are illustrated in the electrophoretic applications.

CZE of $\alpha$-amylase was performed using a marker to identify the position of the electropherogram in the collected data set. No further time scaling was done to compensate for variation of electroosmotic flow. Limits of detection were challenged by high current and a modest detector time constant. Ensemble averaging was performed with five electropherograms of 0.3-$\mu$l injections. A $S/N$ enhancement of 2.5 was found when compared to a single injection (Fig. 1).

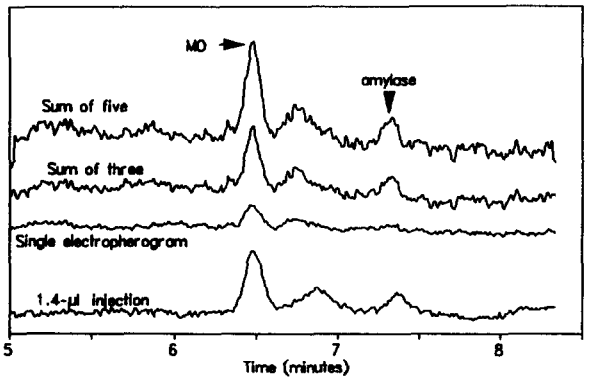


Fig. 1. Ensemble averaging of CZE electropherograms.

The ensemble average is now comparable to a single 1.4-$\mu$l injection.

CIA electropherograms showed a 7% R.S.D. variation in migration time of the analyte peak over nine sequential runs. This could be blamed on variations in interaction between the negatively charged silica surface and the positively charged analyte, the bane of bare silica in separations [13]. Ensemble averaging yielded meaningless peak heights until each data set was aligned such that analyte peaks coincided precisely. This was facilitated by a spreadsheet cut-and-paste option. A selected subset of the data (*i.e.* 20 s before and after the peak maximum) can be cut from the original data and pasted into a new area before averaging. A $S/N$ enhancement of 3.3 compared to a single injection was achieved (Fig. 2).

SEC inherently leads to dilution of the sample component as it is being separated. This limits the use of this technique to moderately concentrated samples. Concentration has been known in a variety of cases [14] to induce dimer formation and skew the final results. Also, process stream samples can be run under overloaded condition, but resolution is lost. Therefore, there is a definite need to get better $S/N$ ratios. Ensemble averaging nine chromatograms of 5-$\mu$l injections resulted in a $S/N$ enhancement of 3.5 (Fig. 3). The result was baseline corrected by digital subtraction of a straight line approximating the baseline slope. The same correction would be possible with a median filter [15] if the peak were not so broad relative to the entire data set. Removal of pump noise will be addressed later in the digital filtering applications.

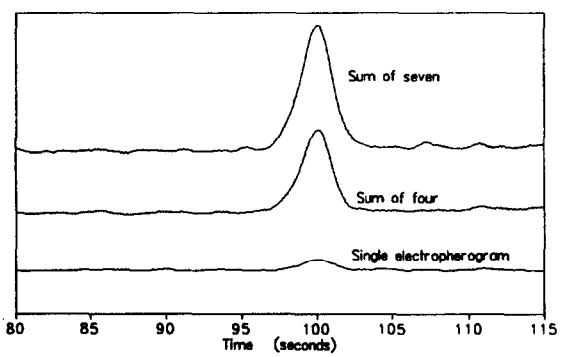


Fig. 2. Ensemble averaging of CIA electropherograms.

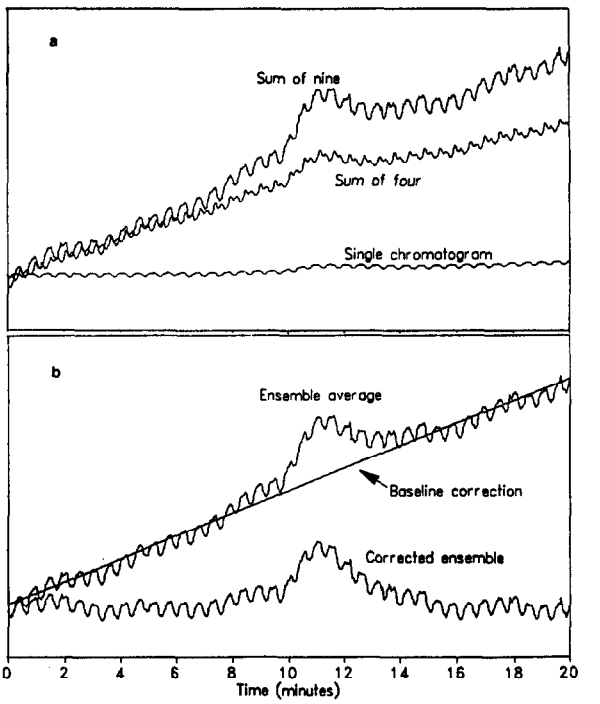


Fig. 3. (a) Ensemble averaging of SEC chromatograms. (b) Baseline subtraction of SEC ensemble average.

With the advent of perfusion chromatography [7,16,17], it is now routinely possible to do very-high-speed separations. This form of chromatography utilizes the throughpores of the particles, thus permitting the use of very high flow-rates and diminishing mass transfer. Column back pressure remains significantly low under these conditions. Perfusion chromatography has major implications in process monitoring. Recently, production and purity of anti-fibronectin antibody was analyzed on-line from CRL-1606 hybridomas [18]. A similar immunoaffinity procedure was performed here with low levels of bovine IgG. Elution peaks did not appear in chromatograms of 2-$\mu$l injections. For such an experiment, ensemble averaging would be the technique of choice to enhance peak height (Fig. 4a). Background subtraction compensated for baseline drift (Fig. 4b). Ensemble averaging seven chromatograms of 2-$\mu$l injections resulted

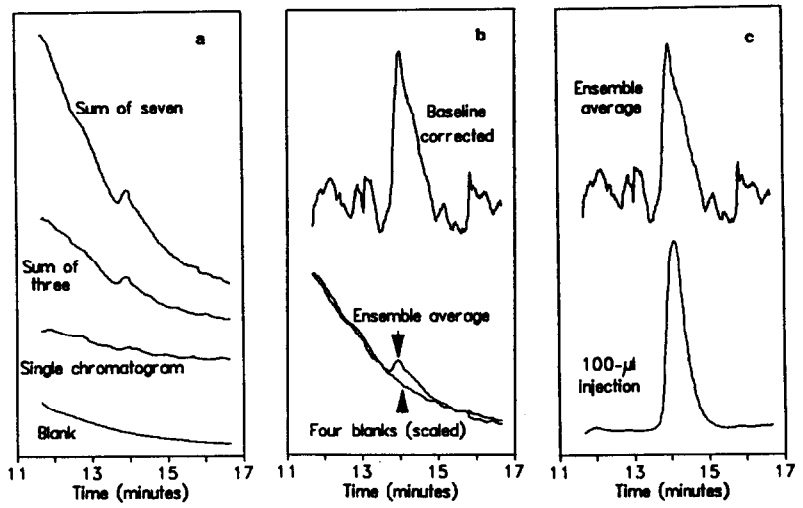


Fig. 4. (a) Ensemble averaging of IAC chromatograms. (b) Blank subtraction for baseline correction of IAC ensemble average. (c) Comparison of IAC ensemble averaging to larger injection.

in a $S/N$ enhancement of 9.3. Comparison of the ensemble average to a 100-µl injection demonstrates the $S/N$ enhancement and that the elution time is not distorted (Fig. 4c). In all four techniques, the experimental enhancement from ensemble averaging was greater than the $\sqrt{n}$ factor expected.

*Digital filtering applications*

Digital filtering of chromatograms and electropherograms is not to be considered a perversion of empirical data. Manipulation of the time constant (RC or $\tau$) of absorbance detectors has been an acceptable method for smoothing baselines. This is an analog filtering technique which is often ignored. Use of an RC waveform in a convolution algorithm mimics the function of the time constant with a digital filter.

Electropherograms were taken with the same CZE capillary and buffer used previously, but analyte and voltage were not applied (Fig. 5). The result was white noise (noise of equal intensity at all frequencies) due to limitations of the apparatus. As Hieftje [19] has given the standard deviation of the samples to be equivalent to the noise, $N$, it was used to make comparisons. The standard deviation of each 2-min data sample (600 data points) was determined for various time constants.

Increased detector time constants resulted in decreased noise levels. The same noise reduction was possible by digital filtering. If the recording of $\tau = 0.05$ s was considered "unfiltered" data, we could mimic the effects of the detector settings. Reversed RC-decay waveforms were produced as exp (time/$\tau$). Convolution of the "unfiltered" data set with the decay functions produced a pattern of noise reduction similar to the analog filter. Variations can be attributed to the assumptions that a 0.05-s time constant was non-filtering, time constants of the detector were accurately reported, and the white noise was

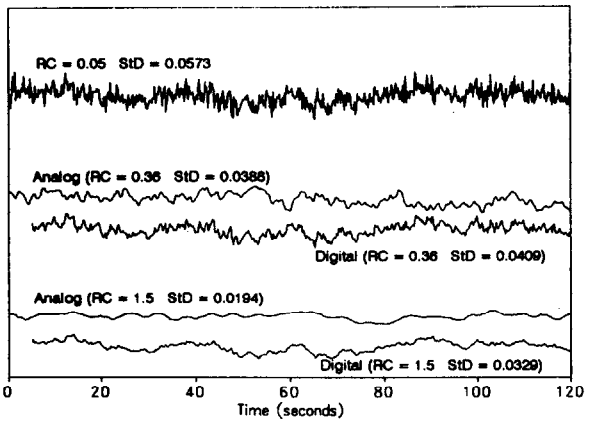


Fig. 5. Comparison of analog and digital baseline smoothing based on the time constant of the filters and the standard deviation of the data.

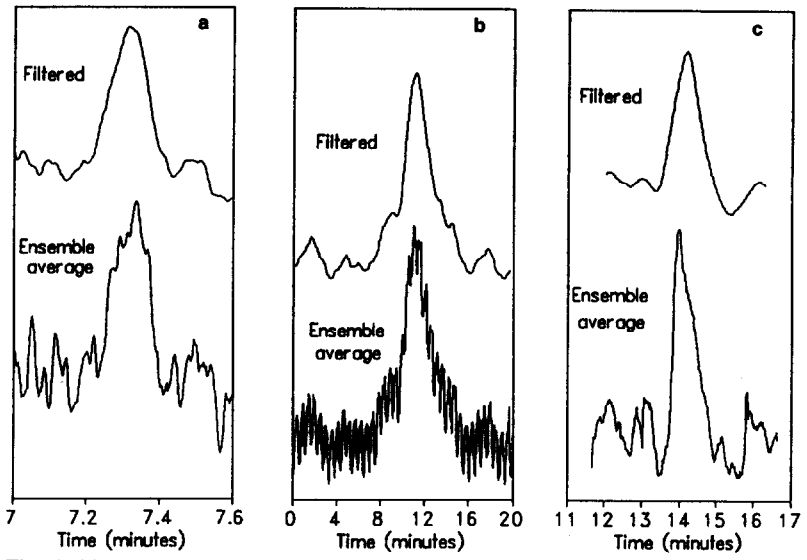


Fig. 6. (a) RC filtering of CZE ensemble average. (b) Running-average filtering of SEC ensemble average. (c) Running-average filtering of IAC ensemble average.

statistically invariant throughout the data collection process.

For electrophoretic experiments, the RC-convolution filter was applied because it mimicked the analog filtering normally conducted by the detector. This showed greater improvement than running-average or Savitsky–Golay filters, and the noise spectrum was assumed to be white. A 3-s $\tau$ was used to generate a 26-point (5.2-s) waveform resulting in a $S/N$ increase of eight for the CZE ensemble average (Fig. 6a). The same filter gave no measurable $S/N$ improvement when applied to the ensemble average of the CIA experiment.

For SEC and IAC, the major noise source was pressure variation throughout the cycle of the pump piston. This was proven by monitoring the column back pressure. The apparent frequency of the pump noise was found from either the absorbance or pressure recordings. Frequency may be aliased[a] dependent on rate of data acquisition. The running-average filter selectively removed the pump noise. The SEC ensemble average was filtered with a 33-s running-average, giving a $S/N$ increase of twelve (Fig. 6b). The IAC ensemble average was filtered with a 45-s running average, giving a $S/N$ increase of 4.5 (Fig. 6c).

The selectivity of this method can be explained in two ways. Most simply, the width of the running average was chosen so as to extend one wavelength [(noise frequency)$^{-1}$] and would reflect a net contribution of zero from the pump noise. The selectivity can also be explained by the frequency-domain behavior of the filter. In the time domain, the running-average filter is a square pulse with a maximum of one. The Fourier transform (conversion between the time and frequency domain) is the sinc function $\{=[\sin(\pi\nu)]/(\pi\nu)\}$ (Fig. 7). Since convolving two functions or data sets is the same as multiplying their Fourier transforms, the pulse will eliminate all frequencies $x\nu_0$ or $x/\tau_0$ [10] where $\tau_0$ is the pulse width. The first crossing was used for frequency cancellation because higher crossings require even greater pulse widths for the same noise frequency.

## CONCLUSIONS

This paper demonstrates that techniques are now applicable for ensemble averaging and digital filtering chromatographic and electrophoretic data. In each experiment, the ensemble average

[a] The term aliasing is from the Nyquist Sampling Theorem and describes how frequencies greater than half the sampling frequency ($f_N$) will yield an erroneous frequency between 0 and $f_N$.

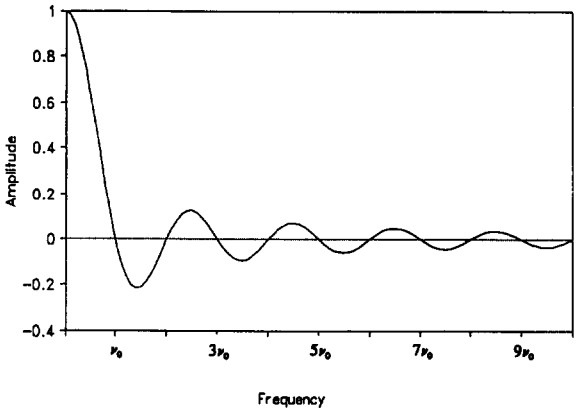


Fig. 7. A running-average filter is a sinc function in the frequency domain where the filter width is defined as $\tau_0$ or $(\nu_0)^{-1}$.

increased the $S/N$ ratio by a factor at least as large as the square root of the number of chromatograms summed. Digital filtering can take many forms, from baseline subtraction to filter convolution algorithms. $S/N$ gains ranged from none to twelve, and were dependent on the type and degree of noise present.

## ACKNOWLEDGEMENTS

This work was supported by NIH grant number GM-25431. M.F.R. acknowledges Fellowship support from the Department of Education and DuPont.

## REFERENCES

1 G. Ouchi, *LC·GC*, 9 (1991) 474.
2 G. Ouchi, *LC·GC*, 9 (1991) 628.
3 G. Ouchi, *LC·GC*, 9 (1991) 767.
4 G. Ouchi, *LC·GC*, 10 (1992) 106.
5 G. Ouchi, *LC·GC*, 10 (1992) 436.
6 L. Ghaoui and L.D. Rothman, *J. High Resolut. Chromatogr.*, 15 (1992) 36.
7 N.B. Afeyan, N.F. Gordon, I. Mazsaroff, L. Varady, S.P. Fulton, Y.B. Yang and F.E. Regnier, *J. Chromatogr.*, 519 (1990) 1.
8 L.J. Lorenz, R. Culp and L.B. Rogers, *Anal. Chem.*, 42 (1970) 979.
9 H.H. Madden, *Anal. Chem.*, 50 (1978) 1383.
10 F.E. Lytle, *Signal-to-Noise Enhancement in Analytical Chemistry*, Purdue University, West Lafayette, IN, 1992, in preparation.
11 J.W. Jorgenson and K.D. Lukacs, *Anal. Chem.*, 53 (1981) 1298.
12 S. Hjertén, *J. Chromatogr.*, 270 (1983) 1.
13 J.K. Towns and F.E. Regnier, *Anal. Chem.*, 64 (1992) 2473.
14 T.K. Nadler, S.K. Paliwal and F.E. Regnier, *J. Chromatogr.*, submitted for publication.
15 A.W. Moore, Jr. and J.W. Jorgenson, *Anal. Chem.*, 65 (1993) 188.
16 F.E. Regnier, *Nature*, 350 (1991) 634.
17 S.P. Fulton, N.B. Afeyan, N.F. Gordon and F.E. Regnier, *J. Chromatogr.*, 547 (1991) 452.
18 S.K. Paliwal, T.K. Nadler and F.E. Regnier, *Anal. Chem.*, in press.
19 G.M. Hieftje, *Anal. Chem.*, 44 (1972) 81A.