JOURNAL OF MAGNETIC RESONANCE **130,** 125–130 (1998)
ARTICLE NO. MN971292

# Analysis of DOSY and GPC-NMR Experiments on Polymers by Multivariate Curve Resolution

Leon C. M. Van Gorkom and Thomas M. Hancewicz

*Unilever Research US, 45 River Road, Edgewater, New Jersey 07020*

Received July 7, 1997; revised September 30, 1997

**Multivariate curve resolution (MCR) was successfully applied to the analysis of DOSY experiments on polymer mixtures and GPC-NMR experiments on industrial copolymer samples. MCR generates pure factors of spectral and concentration profiles using, successively, principal factor analysis, Varimax rotation, and alternating least-squares optimization. The method described is robust and can be directly applied to DOSY and GPC-NMR data and one obtains $^1$H NMR spectra of the individual compounds with their corresponding diffusion or elution profiles, respectively.** © 1998 Academic Press

*Key Words:* **GPC-NMR; diffusion-ordered spectroscopy; principal factor analysis; multivariate curve resolution; polymers.**

## INTRODUCTION

The application of diffusion-ordered spectroscopy to determine the composition of mixtures is based on differences in the diffusion coefficients of the individual compounds. Several methods have been developed to process the diffusion dimension of DOSY data. In cases of high spectral resolution and no spectral overlap, the data can be processed by fitting each frequency channel to a single Stejskal–Tanner exponential (*1*). Johnson and co-workers have developed other approaches for the evaluation of multicomponent echo decay (*2*). Multicomponent decays may be observed in cases of spectral overlap in mixtures or in cases of polydispersity for single compounds. The easiest approach, SPLMOD, assumes a discrete diffusion coefficient for each of the components (*3*). This procedure fits the data to a multiexponential curve and requires a good signal-to-noise ratio. This analysis works reliably for two components if the diffusion coefficients differ by at least a factor of 2. The other analysis, namely CONTIN, involves an inverse Laplace transformation (ILT) that assumes a continuous distribution of sizes, i.e., diffusion coefficients (*4*). Yet again, this analysis requires a good signal-to-noise ratio and was developed to determine size distributions in polydisperse samples.

In the aforementioned data analysis methods, the diffusion domain is analyzed with either a single- or a multiexponential fit or ILT at each frequency channel. The analysis is performed independently of other frequency channels. Thus these methods do not use intrinsic information in the DOSY data that the entire spectral bandshape of each component is attenuated similarly. This information has been incorporated into the global least-squares analysis (CORE) of component-resolved FT-PFGSE NMR spectroscopy (*5*). In this analysis, for $n$ components, $n$ global self-diffusion coefficients are sufficient for modeling the DOSY experiment completely. This analysis is less dependent on the signal-to-noise ratio, but requires the presence of nonoverlapping parts of component bandshapes, or parts that overlap differentially and to a varying extent with other component bandshapes throughout the spectrum. However, the method still relies on a multiexponential fit of each of the frequency channels which is optimized using a direct search minimization routine. A major drawback of this analysis is that one must know the number of spectral components prior to the analysis.

In this paper, we describe a quick analysis of component-resolved FT-PFGSE spectroscopy by means of modern statistical methods to extract the individual components from complex mixtures. This analysis is similar to the NIPALS analysis described by Schulze and Stilbs (*6*). For the latter, however, two data sets are needed in which the spectral components are attenuated differently prior to the actual diffusion experiment resulting in spectral distortions. Analtek and Windig modified the previous method and expressed it in terms of the generalized rank annihilation method (GRAM) (*7, 8*). This method uses one regular DOSY data set that is split into two submatrices and assumes that all spectral components are present in both submatrices. The method requires a priori understanding of the intensity profiles of the spectral components or the spectra of the pure components and, therefore, cannot be used to analyze regular kinetic or chromatographic data. In contrast, MCR analysis is a more generalized method and no assumptions are made with regard to the intensity profiles and only a single data matrix is required for the computation.



1090-7807/98 $25.00

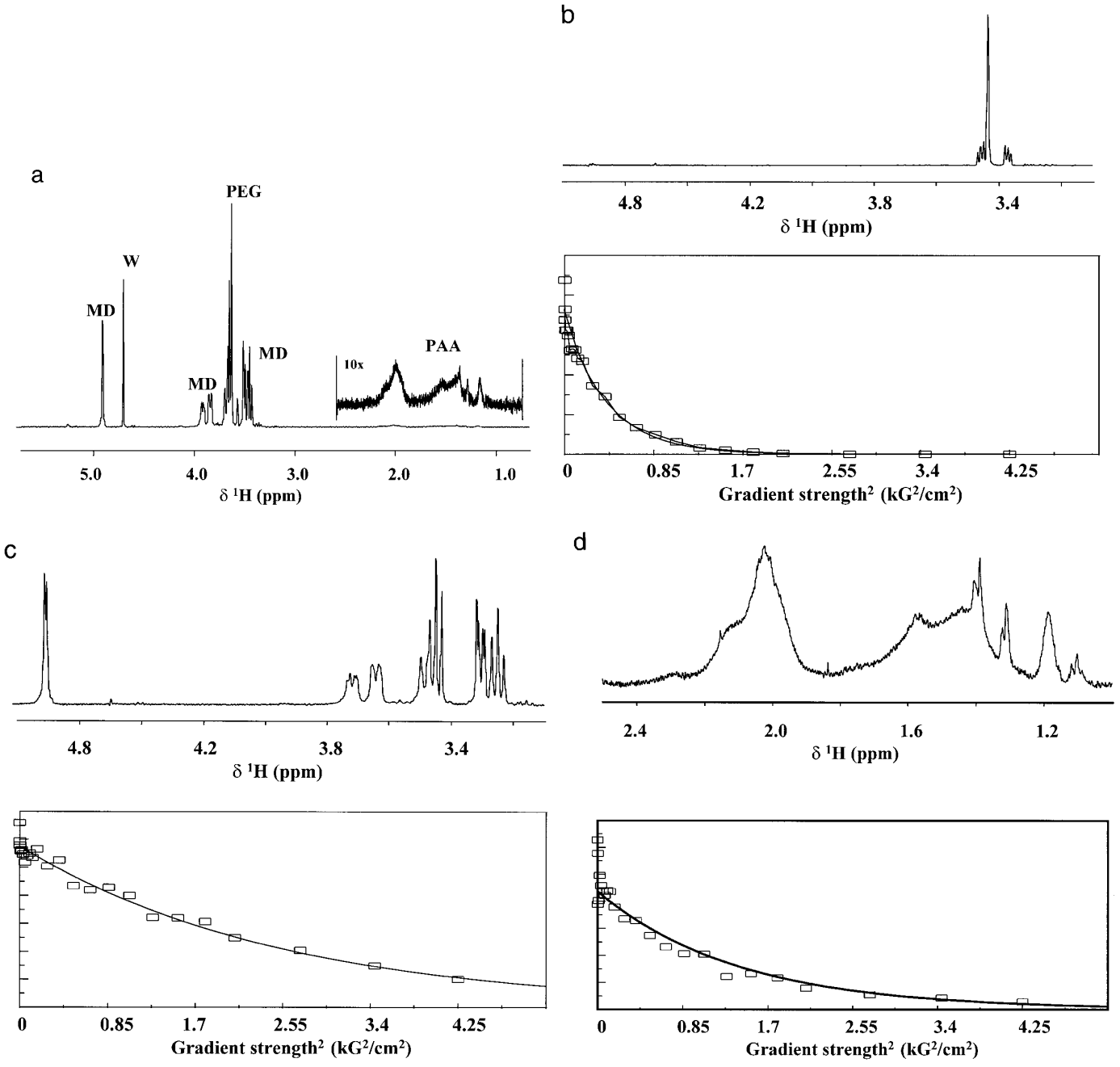


**FIG. 1.** (a) Mean spectrum of DOSY data set, comprising 24 proton NMR spectra recorded with varying gradient strengths (0–65 G/cm). Sample consisted of three polymers in water (W): polyethyleneglycol (PEG-425), polyacrylic acid (PAA, 2 kDa), and maltodextran (MD, 18 kDa). (b) Proton NMR spectrum of 1st factor, polyethyleneglycol (PEG-425) (top), and spin-echo decay profile of 1st factor (bottom). (c) Proton NMR spectrum of 2nd factor, maltodextran (MD, 18 kDa) (top), and spin-echo decay profile of 2nd factor (bottom). (d) Proton NMR spectrum of 3rd factor, polyacrylic acid (PAA, 2 kDa) (top), and spin-echo decay profile of 3rd factor (bottom). (e) Proton NMR spectrum of 4th factor, water (top), and spin-echo decay profile of 4th factor (bottom).

The MCR analysis described in this paper is applied directly to one DOSY data set obtained with the LED pulse sequence using bipolar gradient pulses (*9*). The result of MCR analysis is a set of individual spectra with their echo decay profiles. These pure echo decay profiles for each component can be used to determine the size distribution in the sample by means of multiexponential fitting routines or inverse Laplace transformations. The requirements for this

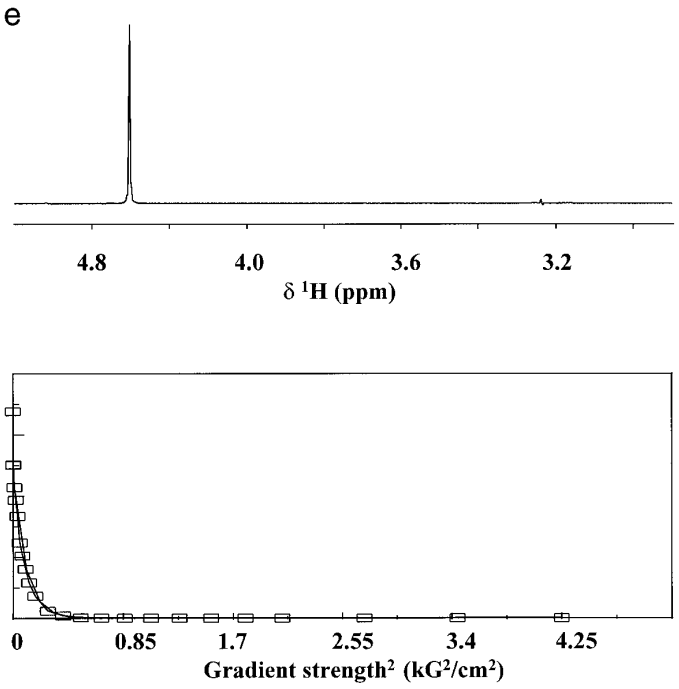


**FIG. 1**—*Continued*

analysis are similar to those for CORE analysis, but the advantage is that MCR analysis does not require any a priori knowledge. In cases of completely identical bandshapes, MCR analysis will result in a single spectral component with a multicomponent decay, which may be further analyzed for polydispersity in polymer samples.

Another important tool in the analysis of polymer samples is separation based on their hydrodynamic volume. Gel permeation chromatography has long been established as a tool for determining the apparent molecular weight and molecular weight distribution of polymers based on their retention volumes. GPC or any other liquid chromatography method can be interfaced with NMR and may provide complementary information to the DOSY data. GPC-NMR allows the correlation of MW distribution with chemical information and provides important information on the composition of polymer blends or copolymers. GPC analysis of polymer blends or copolymer samples has insufficient chromatographic resolution, however, and $^1$H NMR spectroscopy of polymers results in broad, often severely overlapping peaks. In this paper, we demonstrate that MCR analysis can be applied to the analysis of GPC-$^1$H NMR data of polymer samples to extract the pertinent information.

## THEORY

Principal factor analysis (PFA) (*10*) generates abstract eigenvectors and eigenvalues and an additional step is required to convert the abstract vectors into real chemical factors. This means generating real or ''pure'' spectra of the individual species so that proper identifications can be made, and generating real or ''pure'' concentration information (decay profiles for DOSY and elution profiles for GPC-NMR) for each species. This is accomplished by using a Varimax rotation (VMAX) (*10–12*) followed by alternating least-squares optimization (ALS) (*13*). This method is similar in concept to the iterative target transformation factor analysis described by Vandeginste (*14*). The theories of PFA, VMAX, and ALS have been described thoroughly in the literature (*15–19*) and only a conceptual treatment will be given below.

Principal factor analysis assumes a bilinear data structure and attempts to separate the original ($m \times n$) data matrix, $D$, into two submatrices, $C$ and $S$:

$$D = C \cdot S^{\mathrm{T}}. \qquad [1]$$

$C$ is an ($m \times q$) matrix of the coefficients related to the real concentration profiles and $S$ is an ($n \times q$) matrix of vectors related to the real spectra. For the sake of clarity, in PFA the reference to ''spectra''-like vectors will be called factors ($S$) and reference to ''concentration''-like vectors will be called scores ($C$). The solution of the above relationship (Eq. [1]) is not unique and any number of solutions will satisfy this equation. The abstract solutions of $C$ and $S$ are composed of linear combinations of the real concentration profiles and the real spectral profiles of the pure chemical species present in the original data set.

The objective in MCR is to produce pure factors and scores. The number of possible solutions to Eq. [1] must be restricted by chemical or physical meaning so that the concentration and spectral profiles reflect the character of the true components in the data set. The most common way to achieve this is through a process called constrained alternating least-squares optimization. ALS forces the spectral and concentration profiles to conform to a priori knowledge about the data structure. In the case of DOSY data, the spectral intensities and the diffusion decay profiles must both be nonnegative. These are reasonable constraints to place on the expected data structure since in both instances, negative values are not allowed if the solution to Eq. [1] is to be real.

Usually, ALS takes abstract factors and scores as input at the start of the process and performs an iterative least-squares vector rotation of the data. In order to facilitate the optimization process, one can apply an orthogonal rotation of abstract vectors in a self-modeling way, taking advantage of the basic underlying structure of the data. Varimax is such a rotation and is applied to abstract factors in order to segregate the more significant factors from the less signifi-

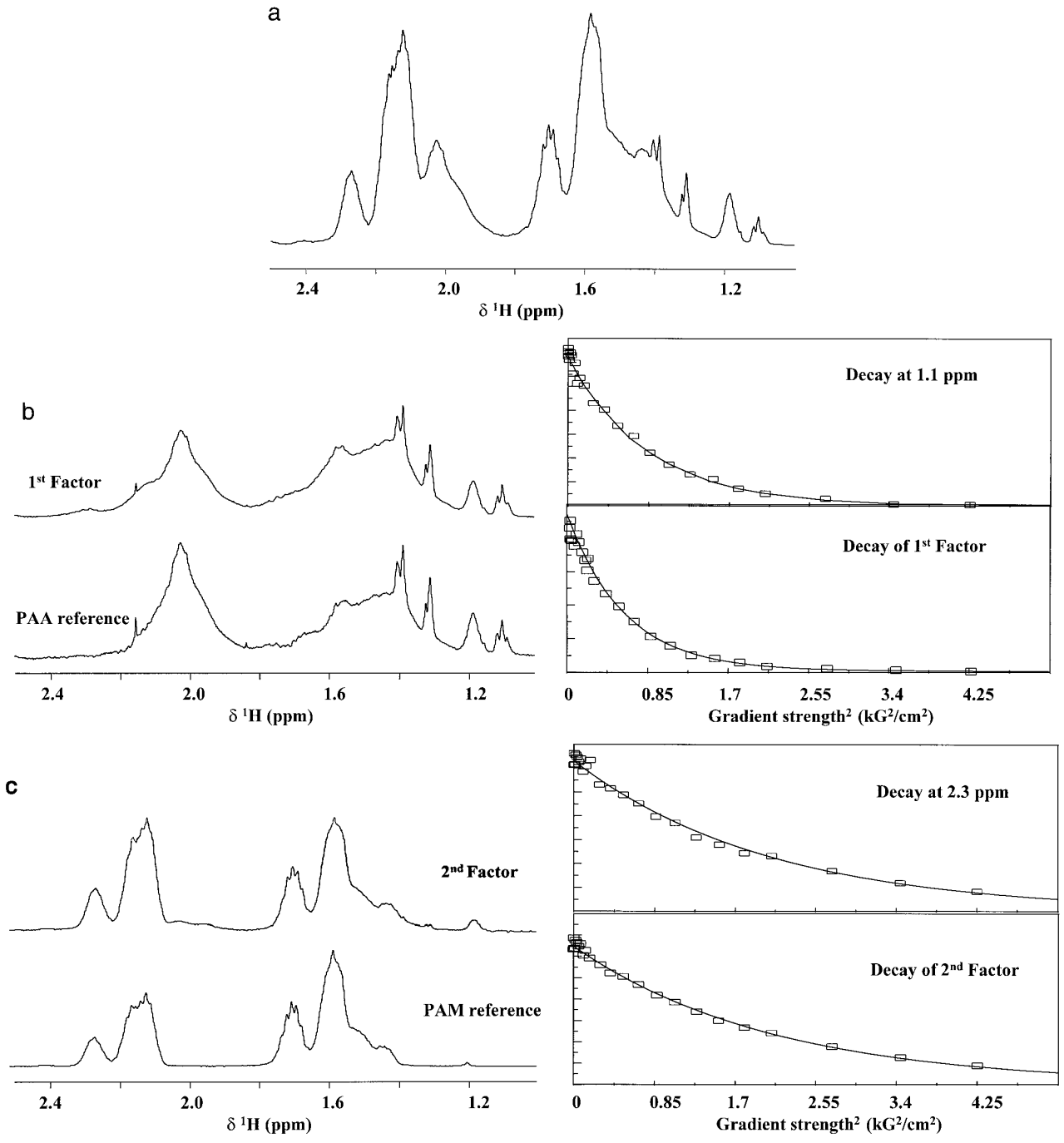


**FIG. 2.** (a) Average spectrum of the DOSY data set for a mixture of polyacrylic acid (PAA, 2 kDa) and polyacrylamide (PAM, 7 kDa). (b) Proton NMR spectrum of 1st factor, polyacrylic acid (2 kDa), obtained from MCR analysis of DOSY data (top) and reference spectrum of PAA (bottom). Comparison of spin-echo decay curve of ''resolved'' resonance directly obtained from DOSY data and the spin-echo decay of the 1st factor obtained from MCR analysis. (c) Proton NMR spectrum of 2nd factor, polyacrylamide (7 kDa), obtained from MCR analysis of DOSY data (top) and reference spectrum of PAM (bottom). Comparison of spin-echo decay curve of ''resolved'' resonance directly obtained from DOSY data and the spin-echo decay of the 2nd factor obtained from MCR analysis.

cant ones. The factors are rotated such that the total variance of the squared factors is maximized which has the effect of giving more importance to either very large or very small factors. This rotation helps to segregate the factor space, making interpretation easier.

The ALS process is initiated after either the Varimax rota-

tion step or, if the Varimax rotation is not performed, immediately after the abstract factor analysis step. Constraints are imposed sequentially on the scores and factors followed by a least-squares optimization of data before proceeding to the next step. Starting with the score matrix ($C$), all negative values are set equal to zero, and then new factors are generated in a least-squares manner from the data matrix using a generalized inverse equation.

$$C = D \cdot S \cdot (S^{\mathrm{T}} \cdot S)^{-1}. \quad [2]$$

Subsequently, the factor matrix ($S$) is constrained and new scores are generated in a similar manner:

$$S = D^{\mathrm{T}} \cdot C \cdot (C^{\mathrm{T}} \cdot C)^{-1}. \quad [3]$$

The entire process is done iteratively and continues until the solution converges according to a predefined criterion which is, typically, a sum-squared residual calculation on the reconstructed data set.

MCR analysis is applicable to any data set where the spectral features change with time and are recorded with sufficient time resolution such that each component that needs to be resolved has a unique maximum or variance with time. MCR analysis can be applied to chromatographic data with seemingly coeluting peaks or kinetic experiments in which the concentration of chemical species changes with time. Diffusion-ordered NMR spectroscopy provides data with this type of information which can be analyzed by MCR.

## RESULTS

Figure 1a shows the mean spectrum of a DOSY data set comprising 24 experiments with gradient strengths varying from 0.6 to 65 G/cm. The sample consists of a mixture of three polymers in water: PEG, MW ~425; PAA, MW ~2 kDa; and maltodextran, MW ~18 kDa. MCR analysis of the DOSY data results in four significant factors, including water. The spectroscopic information and spin-echo decay of each individual factor are shown in Figs. 1b through 1e. This experiment clearly demonstrates that MCR analysis works well in cases of low signal-to-noise ratio (see polyacrylic acid (PAA)) and overlapping signals (PEG and maltodextran). The small $S/N$ ratio for the PAA results in a less than perfect exponential decay. MCR analysis does not rely, however, on the exponential character of the decay but only on the variance of the intensity with gradient strength. This would imply that MCR analysis of DOSY data obtained with less than perfect field gradients (i.e., room temperature shims) would be possible.

Figure 2a shows the average spectrum of the DOSY data set for a mixture of polyacrylic acid (2 kDa) and polyacrylamide (PAM, 7 kDa). In the region between 1.0 and 2.5 ppm, severe overlap of the resonances between PAA and PAM exists. The MCR analysis, however, has no problems in extracting the ''pure'' spectral profiles of PAA and PAM as shown in Figs. 2b and 2c. The reference spectra of both polymers are shown for comparison. The spin-echo decays of the pure components have improved over the decay profiles of the ''resolved'' peaks of each of the polymers (Figs. 2b and 2c). This is due to the better signal-to-noise since the whole spectrum is used in the determination of the intensity (score) of each spectral component in the diffusion-attenuated spectra. For mixtures of polydisperse samples, it is better to extract the pure profiles before analyzing the spin-echo decays with multiexponential fits or inverse Laplace transformations.

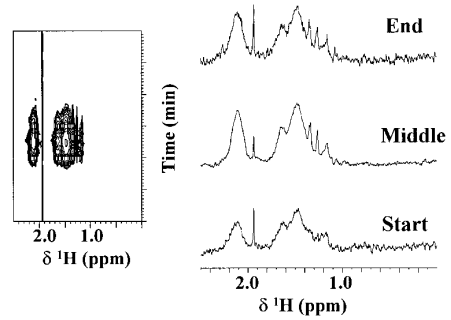


**FIG. 3.** Two-dimensional representation of GPC-NMR data of Narlex DC-1 polymer (left). The *x* axis represents the spectroscopic information and the *y* axis represents the chromatographic dimension. Proton NMR spectra are shown at the leading, middle, and trailing edge of the eluting peak (right).

MCR analysis can also be applied to GPC-NMR data of polymers where the separation is insufficient to resolve individual components in mixtures. Figure 3 shows a two-dimensional representation of the GPC-NMR experiment on a commercial copolymer, Narlex DC-1, which consists of acrylic acid (AA) and laurylmethylacrylic acid (LMA). The GPC elution profile is broad, indicating a wide distribution in the molecular weight. Examination of the proton NMR spectra throughout the elution peak indicates the presence of more than one spectral component. One needs a more thorough analysis, however, to correlate the chemical information with molecular weight information which is achieved by MCR.

The application of MCR analysis to this data set reveals the existence of two significant factors with severely overlapping elution profiles as shown in Fig. 4. MCR analysis of GPC-NMR data is more precise since each component in a mixture will have its unique elution maximum. The results are consistent with the fact that this polymer is a mixture of two forms, i.e., the extended and collapsed form of the AA/LMA copolymer. Depending on the 2-propanol to water

ratio used in the polymerization reaction, different forms of the polymer are obtained. Narlex DC-1 is made under experimental conditions that fall between the two extremes that would yield mainly collapsed or extended polymer forms. The results are also consistent with the fact that the collapsed form of the Narlex DC-1 polymer tends to have a higher molecular weight.

## CONCLUSIONS

This paper demonstrates that MCR analysis of DOSY data generates pure spectra of the individual components for identification. The pure spin-echo diffusion decays that are obtained for the individual components may be used to determine the diffusion coefficient/distribution. The method is reliable and robust and does not rely on the perfect exponential character of the DOSY data and may allow the analysis of DOSY data obtained using gradient pulses created by room temperature shims. In general, MCR analysis can be performed on a data set consisting of 16 to 24 gradient-attenuated spin echoes. For small differences in diffusion coefficients of overlapping peaks, however, the analysis may be more successful if larger data sets are acquired. MCR analysis can also be applied to GPC-NMR data which enables the correlation of molecular weight and chemical information of severely overlapping peaks in both the elution and the spectroscopic domain.

## EXPERIMENTAL

All the data were acquired on a Bruker DMX-500 NMR spectrometer equipped with a 5-mm triple-resonance probe with triple-axis gradients. The DOSY data, recorded as a series of 1D experiments with varying gradient strengths (0.6–65 G/cm), were Fourier transformed and phased prior to MCR analysis.

The GPC-NMR experiment was performed using a 4-mm LC-NMR probe on a Bruker DMX-500 NMR spectrometer. The separation was obtained using three size exclusion columns (Phenomonex, Lichrosorb, 3000, 4000, and 5000A). The mobile phase was 200 mM $Na_2SO_4$ in $D_2O$ with 10% (v/v) methanol. The flow rate was 0.7 ml/min and the NMR data were acquired continuously and time-averaged over 30-s intervals.

MCR data analyses were performed using software (PFA, Varimax, and ALS) written for GRAMS/386 (Galactic Industries) in the native basic array programming language.

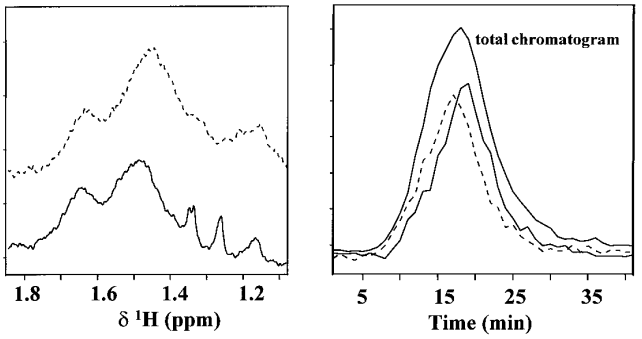


**FIG. 4.** Proton NMR spectra of the two factors obtained from MCR analysis of GPC-NMR data of Narlex DC-1 (left). The factors are identified as the extended (solid) and collapsed (dashed) form of Narlex DC-1. Elution profiles of extended (solid) and collapsed (dashed) form of Narlex DC-1 copolymer.

The operator input is limited to selecting the number of significant factors from the PFA before the Varimax rotation is performed.

## REFERENCES

1. H. Barjat, G. A. Morris, S. Smart, A. G. Swanson, and S. Williams, *J. Magn. Reson. B* **108,** 170 (1995).
2. K. F. Morris and C. S. Johnson, Jr., *J. Am. Chem. Soc.* **114,** 3139 (1992).
3. S. W. Provencher, *J. Chem. Phys.* **64,** 2772 (1976).
4. S. W. Provencher, *Comput. Phys. Commun.* **27,** 213 (1982).
5. P. Stilbs, K. Paulsen, and P. C. Griffiths, *J. Phys. Chem.* **100,** 8180 (1996).
6. D. Schulze and P. Stilbs, *J. Magn. Reson. A* **105,** 54 (1993).
7. B. Analtek and W. Windig, *J. Am. Chem. Soc.* **118,** 10331 (1996).
8. B. Analtek and W. Windig, *Chemometr. Intell. Lab. Syst.* **37,** 241 (1997).
9. D. Wu, A. Chen, and C. S. Johnson, Jr., *J. Magn. Reson. A* **115,** 260 (1995).
10. E. R. Malinowski, ''Factor Analysis in Chemistry,'' Wiley–Interscience, New York (1991).
11. H. F. Kaiser, *Psychometrika* **23,** 187 (1958).
12. M. A. Sharaf, D. L. Illman, and B. R. Kowalski, ''Chemometrics,'' Wiley–Interscience, New York (1986).
13. R. Tauler, *Trends Anal. Chem.* **12,** 319 (1993).
14. B. G. M. Vandeginste, *Anal. Chem.* **57,** 971 (1985).
15. E. R. Malinowksi, *Anal. Chem.* **49,** 606 (1977).
16. H. Gampp, *Talanta* **32,** 1133 (1985).
17. H. Gampp, *Talanta* **33,** 943 (1986).
18. B. G. M. Vandeginste, *Anal. Chim. Acta* **173,** 253 (1985).
19. F. Cuesta, *Anal. Chim. Acta* **314,** 131 (1995).