Resolution of Discrete and Continuous Molecular Size Distributions by Means of Diffusion-Ordered 2D NMR Spectroscopy

Kevin F. Morris and Charles S. Johnson, Jr.*

Contribution from the Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received November 19, 1992

Abstract: Diffusion-ordered 2D NMR (DOSY) experiments have been developed for the analysis of mixtures. These experiments yield conventional chemical shift spectra in one dimension and spectra of diffusion coefficients (or approximate molecular radii) in the other. At the heart of this method are the analysis and display of data obtained with pulsed field gradient (PFG) NMR techniques. At each point on the chemical shift axis, a set of intensities is obtained as a function of K^2 ($K = \gamma g \delta$, where γ is the magnetogyric ratio and g and δ are the gradient pulse amplitude and duration, respectively). These diffusion-dependent data sets are inverted by means of the computer programs SPLMOD, when discrete diffusion coefficients are expected, and CONTIN, when continuous distributions are present. Since the inversion is ill-conditioned, it is necessary to introduce additional information to limit the range of solutions. In addition to assumed prior knowledge of the decay kernels and the nonnegativity of amplitudes and decay constants, a set of criteria has been constructed for the discrete analysis case that takes into account physical limits on diffusion coefficients, experimentally accessible values, and artifacts resulting from variations in effective decay kernels resulting from instrumental nonlinearities. Examples of analyses of simulated data and experimental data for mixtures are displayed. Also, 2D spectra generated by CONTIN are displayed for polydisperse polymer samples. These results demonstrate "at a glance" qualitative analysis of molecular size distributions in mixtures and, with proper care, quantitative determination of diffusion coefficients.

Introduction

The extreme selectivity of high-resolution NMR is based on short-range interactions that are primarily electronic in nature. Long-range interactions associated with the overall size or shape of molecules or molecular aggregates often have small or even negligible effects on observed spectra. For example, sodium dodecyl sulfate (SDS) exhibits essentially the same ^1H spectrum as a monomer or as a component of a micelle, and the spectral patterns of amino acid residues in polypeptides depend on the local environment rather than on structures nanometers away. However, sizes and shapes are related to friction factors for reorientational and translational motion and, thus, influence nuclear relaxation rates. At least for macromolecular and supramolecular systems, reorientational motion primarily affects T_2 and $T_{1\rho}$ values through the modulation of local magnetic dipole–dipole interactions. Studies of these parameters are valuable sources of information about molecular dynamics, but the extraction of size information is complicated by local or segmental motion and varying strengths of magnetic interactions for nuclei at different sites. Translational motion can also contribute to nuclear relaxation rates through intermolecular dipole–dipole interactions, but it is most effectively studied by means of pulsed field gradient (PFG) NMR, where signal attenuation depends on diffusion coefficients and gradient strengths. In contrast to the situation with reorientational motion, the relationship between translational diffusion rates and particle sizes is relatively straightforward and can provide the basis for size resolution in NMR.[1]

We have recently demonstrated a diffusion-ordered 2D NMR (DOSY) experiment that displays conventional chemical shift spectra in one dimension and diffusion spectra in the other dimension.[2] In this experiment, the selectivity of high-resolution NMR compensates for the limited resolution in the diffusion dimension to provide a global view of translational dynamics in mixtures. There are two essential ingredients for this experiment. First, procedures and technology must exist for encoding information about the translational motion for a wide range of particle sizes into NMR data sets. This part has been provided through continued development of the PFG-NMR method;[3,4] the most important improvements for the DOSY application are the stimulated echo (STE) modification for samples with short T_2's,[5,6] introduction of the Fourier transform (FT) to provide chemical shift resolution,[7] eddy current alleviation through active shielding,[8,9] and the longitudinal eddy current delay (LED) pulse sequence.[10] The LED sequence used in this work is illustrated in Figure 1. Thus, we have a high-resolution experiment with signal amplitudes that are limited by T_1's rather than T_2's and with tolerance for gradient pulses of several hundred gauss/centimeter in a narrow bore magnet. With this experiment the accessible particle radii span 5 decades, and the threshold for useful DOSY experiments has been reached.

The second ingredient, the heart of DOSY, is the transformation and display of data for the diffusion dimension. The FT-PFG-NMR experiment provides a 2D data set of the form

$$I(K,\nu) = \sum_n A_n(\nu) \exp[-D_n(\Delta - \delta/3)K^2] \qquad (1)$$

where $K = \gamma g \delta$ is the area of the gradient pulse in cm^{-1}, γ is the gyromagnetic ratio, g and δ are the amplitude and duration of the gradient pulses, respectively, and D_n is the tracer diffusion coefficient of the nth species. Here $A_n(\nu)$ is the 1D NMR spectrum of the nth diffusing species in the limit that $g = 0$. The important

* Author to whom correspondence should be addressed.

(1) Stilbs, P. *Prog. Nucl. Magn. Spectrosc.* **1987**, *19*, 1.
(2) Morris, K. F.; Johnson, C. S., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 3139.
(3) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1964**, *42*, 282.
(4) Stejskal, E. O. *J. Chem. Phys.* **1965**, *43*, 3597.
(5) Woesner, D. E. *J. Chem. Phys.* **1961**, *34*, 2057.
(6) Tanner, J. E. *J. Chem. Phys.* **1970**, *52*, 2523.
(7) James, T. L.; McDonald, G. G. *J. Magn. Reson.* **1973**, *11*, 58.
(8) Mansfield, P.; Chapman, B. *J. Magn. Reson.* **1986**, *66*, 573.
(9) Gibbs, S. J.; Morris, K. F.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *94*, 165.
(10) Gibbs, S. J.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *93*, 393.

0002-7863/93/1515-4291$04.00/0

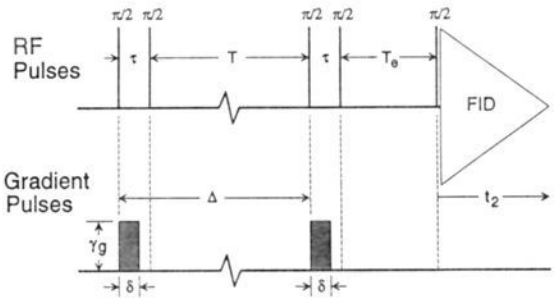

Figure 1. LED pulse sequence. Gradient prepulses, homospoil pulses during T_e, and phase cycling of the radio frequency pulses are not shown.

point is that each frequency ν in the NMR spectrum is associated with a 1D data set (column) that is described by a sum of exponential components with K^2 as the independent variable. Thus, at a given frequency, the data set can be described by

$$G(s) = \int_0^\infty g(\lambda) \exp(-\lambda s)\, d\lambda \qquad (2)$$

where $\lambda = D(\Delta - \delta/3)$ and $s = K^2$. Equation 2 emphasizes the facts that $g(\lambda)$ is the "spectrum" of diffusion coefficients and that the PFG-NMR data set $G(s)$ is the Laplace transform of $g(\lambda)$. Of course, when discrete decay constants λ_i are present, $g(\lambda)$ consists of a sum of δ functions, $\delta(\lambda - \lambda_i)$, with appropriate coefficients. In principle, the spectrum $g(\lambda)$ can be obtained by inverse Laplace transformation (ILT) of the data set $G(s)$; however, this is an ill-conditioned problem and may be completely intractable. The problem is reminiscent of the analysis of FIDs to obtain frequency spectra, but in fact is quite different. FIDs consist of damped sinusoidal components, and a Fourier transformation (FT) suffices to yield a unique spectrum of resonance lines in the absence of truncation errors.

The exponential functions in eq 2 are strongly nonorthogonal, and attempts to transform $G(s)$ directly either yield limited information or fail completely. It is instructive to consider the analysis of Gardner et al. for discrete sums of exponential components.[11,12] In the notation of eq 2, s is replaced with $\exp(x)$ and the function $\exp(x)\, G[\exp(x)]$ is constructed. It can be shown that this function is the convolution of the desired frequency spectrum with the shape function $\exp(x)\, \exp[-\exp(x)]$. The problem is thus recast into a deconvolution to remove the broadening that obscures detail in the spectrum. The difficulties are obvious in any deconvolution scheme since noise becomes unacceptable as the line width is decreased. A better procedure is to use such transforms as the starting point for nonlinear least-squares analyses. This is in fact the strategy used in the automatic analysis program DISCRETE.[13,14]

The initial demonstration of DOSY was based on diffusion analysis by means of DISCRETE. This was quite successful for species with isolated resonance lines, i.e., single exponential decays, and for species at the same chemical shift but having widely separated diffusion coefficients. However, the construction of each 2D spectrum required independent analyses of hundreds of data sets and suffered from fluctuations in decay constants and, in some cases, even in the assignment of different numbers of exponential components for the same diffusing species in adjacent data columns. It was also clear with some samples that discrete components were not appropriate.

In this article, we are concerned with limits on the resolution of discrete components and the possibility of characterizing continuous distributions of diffusion coefficients, e.g., polydisperse samples. Provencher and co-workers have developed two data inversion programs that are of particular interest in this connection. The program SPLMOD is restricted to discrete components,[15] but can handle a number of data sets simultaneously, while the program CONTIN determines continuous distributions of decay rates by means of a constrained regularization method.[16,17] Here we explore the effectiveness of these programs in analyzing PFG-NMR data sets. The ill-posed nature of the analysis problem demands that absolute prior knowledge (APK) including physical limitations be introduced to limit the range of solutions.[16] To investigate this problem, we have analyzed both simulated multicomponent data sets and experimental PFG-NMR data sets for mixtures. From this experience we have constructed a set of criteria including APK and guidelines for the rejection of solutions on the basis of reported errors and conservative limits on the resolution of decay rates that can be justified by our knowledge of the decay kernels. These criteria serve to define the "standard" DOSY experiment.

In this report, we describe the criteria and procedures in detail as an aid to understanding the nature of DOSY and to facilitate the discussion and comparison of future modifications and enhancements. Improvements in technology and additional experience with the analysis programs will undoubtedly lead to modifications of the criteria. Our basic conclusion from this work is the following: Pitfalls abound, but with objective rules and automated procedures to eliminate user bias, DOSY can be a powerful tool for the analysis of dynamics in mixtures. Similar procedures are valid for relaxation spectroscopies based on other types of nuclear relaxation, e.g., transverse, rotating frame, or longitudinal.[18]

Experimental Section

Materials. The nonionic surfactant poly(ethylene glycol) dodecyl ether (Brij-30) was purchased from Fluka Chemical Corp., and Sigma Chemical Company supplied tetraethylammonium chloride, D-(+)-glucose, and methyl cellulose (viscosity of 2% solution at 25 °C approximately 1500 cp). Sodium dodecyl sulfate (98%), adenosine 5′-triphosphate, disodium salt hydrate, and 1,3,5-triisopropylbenzene (97%) were provided by Aldrich Chemical Company, Inc.

Sample Preparation. Chemicals were used without further purification, and samples were prepared gravimetrically in deuterium oxide (99.9% D) purchased from Cambridge Isotope Laboratories. The oil-in-water microemulsion sample was prepared by adding 0.150 mL of 1,3,5-triisopropylbenzene, 0.165 mL of Brij-30, and 0.075 g of sodium dodecyl sulfate to 20.0 mL of D_2O and thoroughly mixing the components. The sample was then transferred to a separatory funnel and allowed to sit overnight at room temperature. Finally, the bottom layer was drained off and diluted by a factor of 6, and the resulting mixture was allowed to sit for approximately 1 week before use.[19]

PFG-NMR Spectrometer. All NMR experiments were performed on a Bruker AC-250 spectrometer with custom-built 10-mm probes supplied by Cryomagnet Systems, Inc. The probes were equipped with actively shielded gradient coils, and calibration of the coils with diffusion measurements on HOD in D_2O gave coil constants of 20.8 and 15.6 G cm^{-1} A^{-1} for the two gradient coils used in these experiments. The computer-controlled gradient drivers have been described elsewhere.[20-22]

Data Acquisition and Conversion. In a typical DOSY experiment, 30–50 free induction decays, each associated with a different value of K, were collected on an ASPECT-3000 computer by means of the LED pulse sequence. The K^2 values were closely spaced at small K values, i.e., at the beginning of the data set, and the spacings increased with each FID in the series. This spacing insured that small K values were sufficiently

(11) Gardner, D. C.; Gardner, J. C.; Laush, G.; Meinke, W. W. *J. Chem. Phys.* **1959**, *31*, 978.
(12) Swingler, D. N. *IEEE Trans. Biomed. Eng.* **1977**, *BME-24*, 408.
(13) Provencher, S. W. *J. Chem. Phys.* **1976**, *64*, 2772.
(14) Provencher, S. W. *Biophys. J.* **1976**, *16*, 27.
(15) Provencher, S. W.; Vogel, R. H. In *Numerical Treatment of Inverse Problems in Differential and Integral Equations*; Deuflhard, P., Hairer, E., Eds.; Birkhäuser: Boston, 1983; p 304.
(16) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 213.
(17) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 229.
(18) Bulliman, B. T.; Kuchel, P. W. *J. Magn. Reson.* **1985**, *62*, 556.
(19) He, Q. Two dimensional Electrophoretic NMR Spectroscopy. Ph.D. Thesis, University of North Carolina, Chapel Hill, NC, 1990.
(20) Saarinen, T. R.; Woodward, W. S. *Rev. Sci. Instrum.* **1988**, *59*, 761.
(21) Johnson, C. S., Jr.; He, Q. In *Advances in Magnetic Resonance*; Warren, W. S., Ed.; Academic Press: San Diego, 1989; Vol. 13, p 131.
(22) Boerner, R. M.; Woodward, W. S. Manuscript in preparation.

close together to detect the fastest diffusing component and that the K values at the high end were large enough that the signal of the slowest component would be attenuated by at least an order of magnitude. After data collection, the FIDs were transferred to a Silicon Graphics 4D-25 or INDIGO (MIPS R3000) workstation via Ethernet by means of the transfer program BRUKNET. The FIDs were then processed with the NMR software package FELIX 1.1 (Hare Research, Inc.). This included apodization, Fourier transformation, phasing, and polynomial base-line corrections. The resulting set of NMR spectra with peak intensities decaying exponentially with K^2 was then read into the program selected for data inversion.

Data Inversion and Display

The fitting of experimental data to sums of exponential functions by means of nonlinear least squares or any other procedure does not insure unique solutions. Instead we must seek the best fit that satisfies previously established criteria. We must first decide whether discrete components are appropriate for the system under study. This is often obvious from a knowledge of the composition of the sample, e.g., N monodisperse nonaggregating components are completely described by N pairs of amplitudes and decay constants. When the choice is not obvious because of lack of knowledge of composition or state of aggregation, an analysis scheme must be chosen that can handle continuous distributions. In the event that the analysis reveals narrow distributions for some components, discrete single component analyses can then be applied in appropriate chemical shift ranges to improve the resolution of the diffusion coefficients.

In all cases, we impose prior knowledge (APK) that the amplitudes and decay constants for the exponential components are nonnegative. The decay constants are always positive, but the signs of the amplitudes may depend on the type of PFG-NMR experiment, e.g., J-modulation can produce inverted spin echoes. Since the nonnegativity of amplitudes is an extremely important criterion for selecting solutions, it is necessary to use a pulse sequence such as LED where transverse evolution periods are too small to generate J-modulation effects. Furthermore, we establish upper and lower limits for possible diffusion rates based on the Stokes–Einstein relation and an additional lower limit based on the range of our PFG-NMR experiment. We assume that accurate measurement of D requires signal attenuation by a factor of 10. In addition, our experience with DOSY data sets suggested a second level of criteria. These have to do with the level of confidence that can be placed in certain combinations of results that arise in discrete analyses. In the following we consider discrete and continuous analyses separately, so that situations unique to each type of analysis can be treated in detail.

Resolution of Discrete Components. In all analyses, a signal-to-noise (S/N) threshold is established to select data columns in the 2D set that merit analysis. When a small number of discretely diffusing components is known to be present, the diffusion dimension is inverted by fitting each selected column in the 2D data set to a sum of exponentials. In previous work we used the program DISCRETE, and in this work we have implemented the more efficient program SPLMOD to perform the inversion.[15,23] SPLMOD offers all the same possibilities as DISCRETE and in addition is not restricted to exponential components and permits the parallel processing of N_d data sets. Thus, SPLMOD can be used to analyze data sets of the form

$$y_{kn} \approx \sum_{j=1}^{N_\lambda} \alpha_{jn} f(\lambda_j, t_{kn}) + \sum_{\mu=1}^{N_\xi} \xi_{\mu n} g_\mu(t_{kn}) \qquad (3)$$

where $k = 1, ..., N_{y(n)}$ and $n = 1, ..., N_D$. Here y_{kn} and t_{kn} are the experimental data in the nth data set, $y(n)$ is the number of data points in the nth set, N_ξ and the functional forms of $f(\lambda_j,t)$ and $g_\mu(t)$ are known, and α_{jn}, λ_j, $\xi_{\mu n}$, and N_λ are to be determined.[15]

(23) Vogel, R. H. *SPLMOD Users Manual (Ver. 3)*, Data Analysis Group, EMBL: Heidelberg, Germany, 1988; EMBL-DA09.

The program, therefore, handles sums of exponentials and convoluted exponentials as well as other functions and permits base-line corrections.[15] Here we have taken $\xi_{\mu n}$ to be zero and $f(\lambda_j,t) = \exp(-\lambda_j t)$, so that sums of exponentials with no base line are fit to the data. The symbols in the first term on the rhs of eq 3 agree with Vogel's notation.[23] The decay constant λ was introduced in eq 2, and the data set index n in eq 3 is associated with discrete values of the frequency ν in eq 1.

Special problems exist in carrying out the DOSY analysis because of the danger that user interaction with the analysis program will bias the results and because a large number of analyses must be performed (typically at least 200 per data set). Fortunately, the SPLMOD program has features that allow these problems to be overcome. First, no initial values (guesses) of starting values of λ_j or α_{jn} are required or even permitted by SPLMOD, and only the maximum number (NNL) of components N_λ that will be searched is required from the user. Thus, unbiased analyses are performed. Second, a large number of data sets can be fit in parallel with the restriction that each data set have the same set of decay constants but different amplitudes. SPLMOD has the ability to approximate the model functions with B-splines, thus permitting reliable analyses of data while decreasing the computation time by more than an order of magnitude. However, in the DOSY application, the major advantage of carrying out analyses on all of the points in a peak or multiplet in parallel is that column-to-column variations in the calculated decay constants that were present with independent column-by-column analysis with DISCRETE are eliminated.

The first step in the DOSY analysis is to decide which columns in the 2D data set are to be analyzed simultaneously. This is done by an input program written in-house that first prompts the user for a threshold value and then divides the spectrum into a number of regions containing the columns to be analyzed. The program sequentially scans the points in the spectrum collected with the lowest K value and chooses the regions starting whenever a point above the threshold is found and ending that region when a point below the threshold is found. The user can then delete regions, divide them further, enter a higher or lower threshold and reiterate, or proceed with the regions chosen. Provisions also exist for selecting the region boundaries manually. At present, we analyze up to 100 data sets simultaneously with SPLMOD. Accordingly, the input program has been designed to divide automatically any region with more than 100 points into smaller regions with equal numbers of points. The number of data sets that can be processed in parallel is limited only by computation time and the computer memory required.

The input program then allows decisions to be made about the nature of the SPLMOD fit. Typically an exponential decay kernel is chosen for $f(\lambda_j,t)$ in eq 3, and λ_j and α_{jk} are restricted to have nonnegative values. The maximum number of discrete components (NNL) must also be input at this stage. Although SPLMOD in principle handles up to 10 components, DOSY analyses are performed with a maximum of 3 components. This is a conservative choice based on the following: (i) the number of data points in each set is 50 or fewer; (ii) S/N ratios are quite small in some spectral regions; and (iii) our experience with errors and artifacts in fits with three or more components suggests caution. The input program finally writes the parameters defining the fit and the raw data into a form that can be read by SPLMOD.

The CPU time required to perform the SPLMOD analysis, of course, increases with the number of points in each data set, the number of columns analyzed, and the number of components fit. For the analyses performed here with less than 50 points per column, less than 500 columns analyzed in each 2D data set, and a maximum of 2 or 3 components to be fit in each column, 5 min or less CPU time was required on the INDIGO workstation for each 2D data set. The output from the SPLMOD analysis is a set of decay constants for each region, a set of amplitudes for

each column analyzed, and standard deviations for each of these parameters. When NNL is specified, SPLMOD determines the best NNL, NNL − 1, ..., 1 component fits and reports which fit corresponds to the best solution, i.e., N_λ is determined. Additional information such as correlation coefficients, residuals, and plots of data and fits can also be displayed if desired.

When the parameters corresponding to the SPLMOD best solution are used directly, DOSY analyses of simulated data based on exponential kernels and Gaussian random noise usually give completely satisfactory 2D displays. However, the SPLMOD-based DOSY analyses of experimental PFG-NMR data sets often generate artifacts. The experimental data may be distorted by instrumental nonlinearities that invalidate the assumption of exponential kernels. For example, phasing and base-line corrections are determined for the spectrum with the lowest K value and are not changed for the other spectra. Also, the current pulses required to produce large K values may introduce signal distortions through eddy currents, sample heating, etc. Thus, a K-dependent deviation in signal amplitudes is possible. In the absence of exactly known decay kernels, we have established objective criteria for rejecting physically and mathematically unreasonable fits. A third program, which acts as a rejection filter, implements these criteria, and then generates the final 2D spectrum. When an N_λ component fit is rejected for any reason, the best $N_\lambda - 1$ component fit in that region is taken as the best solution, and this iteration continues until either the criteria are met or all fits in that region have been rejected. The additional criteria are the following: (1) diffusion coefficients must be feasible and experimentally accessible; (2) standard errors in diffusion coefficients must be less than 30%; and (3) pairs of diffusion coefficients must differ by a factor of more than 2.

Here we face the problem that a multicomponent fit to experimental data including noise can usually be improved by allowing more components. The last criterion was included because two component fits with closely spaced diffusion coefficients were often found for samples containing only one discrete component, e.g., the HOD peak in aqueous solution. Rejection of fits in which the diffusion coefficients are not a factor of 2 apart can, of course, lose real peaks, but this resolution is beyond the capabilities of the experiment and the rule is necessary to generate DOSY spectra consistently without artifacts. The guiding principle is that it is better to lose information than to introduce information not demanded by the data. Fortunately, DOSY displays contain information from several spectral regions that can confirm or call into question the analysis of a given region. This type of information can provide the basis for additional rejection criteria.

Once a solution that meets the above criteria has been found in each region, the final 2D spectrum is generated with normalized Gaussians having center positions and intensities equal to the diffusion coefficients and amplitudes determined by SPLMOD, respectively. The standard deviations for the Gaussians are set equal to the percent deviations in the decay rates reported by SPLMOD plus 4%. Inclusion of the extra line width is necessary because the analysis of different NMR multiplets corresponding to the same diffusing species can give slightly different diffusion coefficients but usually within the 4% of their average value. The line-width threshold permits smooth projections to be obtained and reflects errors in the measured diffusion coefficients more realistically.

In order to illustrate the features of SPLMOD, we first consider a successful inversion of simulated data. The data were synthesized for two isolated Lorentzian peaks and three overlapping Lorentzian peaks with a S/N of 20 for the most intense peak and half-widths at half-maximum (HWHM) of 5.0 Hz as shown in Figure 2. The diffusion coefficients are as follows: 5.00×10^{-6}, 5.00×10^{-8}, 1.00×10^{-7}, and 5.00×10^{-7} cm^2 s^{-1} for the peaks at 750, 520, 500, and 480 Hz, respectively, and a two-component decay with diffusion coefficients of 5.00×10^{-8} and 5.00×10^{-6} cm^2 s^{-1} for the peak at 250 Hz. The simulated data set contained 50 spectra with equally spaced K values ranging from 556 to 2.78×10^4 cm^{-1}, and $\Delta = 0.1$ s where the decay constants are given by $\lambda_j = D_j(\Delta - \delta/3)$. This choice of K value spacings is, in fact, very inefficient as it gives only a few points in the range of the fastest decay curve and an excess of points for the slower decays. Twenty points with optimum spacings should give at least as good accuracy in the analysis. Exponential spacing of K^2 values ensures adequate sampling of all decay curves. A convenient expression is $K_n = K_0 2^{n/m}$ where K_0 is the smallest required value and m is an integer that determines the density of sampling.

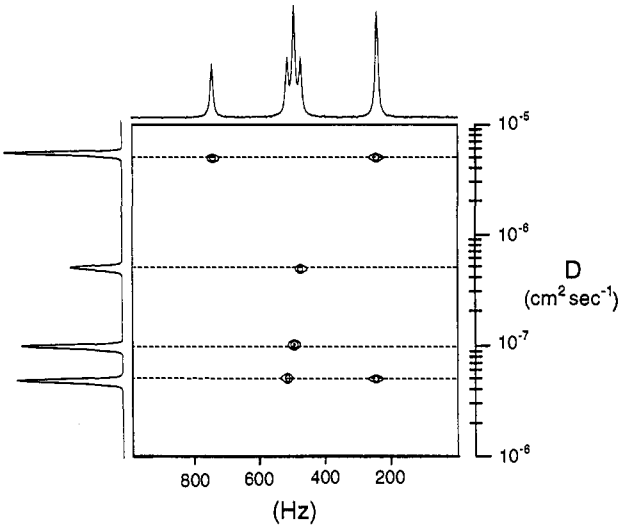

Figure 2. DOSY display of the SPLMOD analysis of simulated data. All diffusion peaks are correctly positioned.

The SPLMOD analyses were performed by dividing the data set into three regions: one centered at 750 Hz, one corresponding to the overlap region centered at 500 Hz, and the third centered at 250 Hz. The fits in each region were performed with a maximum of three components. The spectrum synthesized with the lowest K is shown as a projection onto the frequency axis, and the diffusion spectrum of the mixture was obtained by projecting the 2D spectrum onto the diffusion axis. The diffusion coefficients determined from the maxima of the peaks in the diffusion spectrum are 4.95×10^{-8}, 9.92×10^{-7}, 4.97×10^{-7}, and 4.99×10^{-6} cm^2 s^{-1}, in good agreement with values used to synthesize the data.

Figure 3 illustrates a type of artifact, found in the DOSY analysis of another synthesized data set, that was not rejected by the listed criteria. This simulation contains three overlapping Lorentzian peaks with S/N of 20 and HWHM = 10.0 Hz where the center peak at 515 Hz is the sum of two components with distinguishable decay rates. The diffusion coefficients were 5.00×10^{-7} cm^2 s^{-1} for the peak at 555 Hz, 1.00×10^{-7} and 2.00×10^{-8} cm^2 s^{-1} for the peak at 515 Hz, and 5.00×10^{-8} cm^2 s^{-1} for the peak at 475 Hz. A total of 50 spectra were analyzed with the same K values as the previous simulation, and the 85 frequency points defining the three overlapping peaks were analyzed simultaneously and fit to a maximum of four components. Examination of the SPLMOD-generated 2D spectrum shows a three-component decay with diffusion coefficients of 1.76×10^{-8}, 1.15×10^{-7}, and 4.98×10^{-8} cm^2 s^{-1} for the center peak instead of the correct two-component decay. The spurious "cross-talk" peak at 4.98×10^{-8} cm^2 s^{-1} corresponds to the decay rate for the overlapping peak at 475 Hz. Also, the calculated diffusion coefficients of 1.76×10^{-8} and 1.15×10^{-7} cm^2 s^{-1} for the center peak at 515 Hz show that the derived diffusion coefficients for this peak deviate from the input values because of the presence of the spurious third peak.

Disturbingly, this three-component fit even met the analysis criteria established for experimental data and was not rejected.

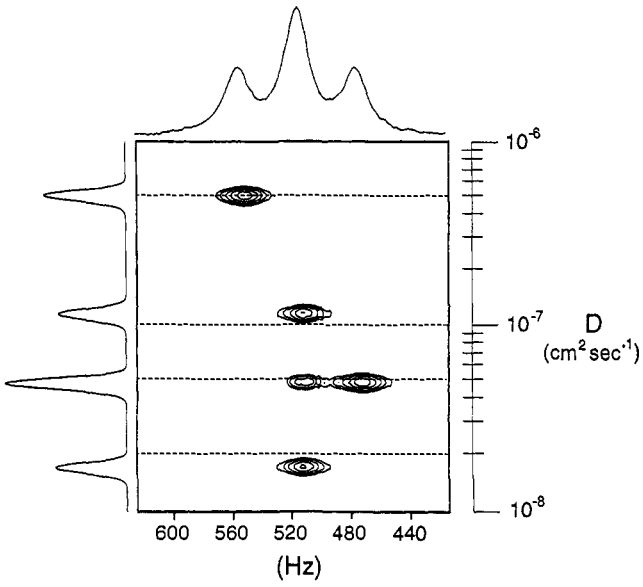

Figure 3. DOSY display of the SPLMOD analysis of simulated data. A cross-talk artifact is illustrated. The correct (input) diffusion coefficients are indicated by dotted lines.

This artifact reveals a special vulnerability of fits with more than two components and possibly suggests additional guidelines. Studies on this point are not complete, but experience shows that discrete fits with three components are suspect. Given the limited range of accessible diffusion coefficients, analysis of more than three components is probably impossible. Simulations with the same frequencies and diffusion coefficients, but with reduced line widths in the chemical shift dimension, have also been analyzed as a single block with SPLMOD. The spurious third peak at 515 Hz remains, albeit with reduced intensity, even when overlap between components of the multiplet vanishes. The problem is associated with the number and magnitudes of the decay constants in one analysis block and is exacerbated by spectral overlap. It turns out that SPLMOD is vulnerable to this type of artifact because any decay constant required by part of a block will be used anywhere in the block to improve the fit. Of course, our standard input program automatically separates blocks of data when the signal drops below a threshold, so that the artifact at 515 Hz vanishes. Just as coincidences of decay constants with peaks in the same analysis block suggest cross-talk, peaks that show no coincidences with peaks found in other blocks suggest spurious results as well. When SPLMOD reports three or more diffusion coefficients at the same chemical shift that differ by factors of less than 10, errors are likely to be significant and peaks with small amplitudes are likely to be artifacts. Fortunately, this amount of overlap is unlikely even with the 2–4 Hz spectral line widths found in a nonspinning sample. When extensive overlap is encountered, it is appropriate to either restrict the analysis to spectral regions where fewer diffusion coefficients are found or to perform a different type of NMR experiment, e.g., ^{13}C rather than ^{1}H NMR, so that overlapping peaks can be avoided. For these situations, other analysis strategies are being investigated. For example, a set of decay constants can be determined from regions that are well fit by one or two exponential components. Then, the data blocks for regions with multiple overlaps can be fit with the restriction that the kernels contain only the decay constants previously established.

Now we turn to the analysis of real data. Figure 4a shows the diffusion-ordered spectrum of a mixture containing 40.0 mM tetraethylammonium chloride (TEA) and 20.0 mM sodium dodecyl sulfate (SDS). This experiment was performed at 22.0 °C by collecting 40 NMR spectra with K values ranging from 167 to 7.79×10^3 cm^{-1}, δ from 1 to 2 ms, Δ = 100.0 ms, T_e = 50.0 ms, and τ = 2.50 ms. Here each spectral region was fit to a maximum of two components (NNL = 2), and to illustrate the presence and elimination of artifact peaks, the DOSY display was generated directly from the SPLMOD best fits without the benefit of the postanalysis filter. The CH_3 peak for TEA and CH_2 peaks for the surfactant at 1.11 ppm overlap in the 1D NMR spectrum, and the two-component fit in this region gives real diffusion coefficients at 4.55×10^{-6} and 8.32×10^{-8} cm^2 s^{-1}. However, in every other region a single component with noise is fit to two exponential components, and the decay constants of the components repel one another in the sense that one falls at a lower value and one at a higher value than the true single exponential decay constant. Projection of the DOSY spectrum onto the diffusion axis shows that a number of artifacts are present, and physical interpretation of the spectrum in the presence of these artifacts is impossible.

However, when the SPLMOD solution set is subjected to the rejection criteria described above, the DOSY display in Figure 4b is generated. All artifacts are eliminated, and projection of the DOSY peaks onto the diffusion axis clearly shows that three diffusing components are present in the mixture. Peaks in the diffusion spectrum for HOD, TEA, and SDS give diffusion coefficients of 1.74×10^{-5}, 4.54×10^{-6}, and 8.84×10^{-7} cm^2 s^{-1}, respectively. This DOSY analysis is typical of what we have been able obtain for simple mixtures of monodisperse components.

Figure 5 illustrates the ability of DOSY to separate complete NMR spectra of components in mixtures, thus providing an implementation of size-resolved NMR.[1] In Figure 5a we show the DOSY display for a mixture containing 25.0 mM SDS, 50.0 mM adenosine 5′-triphosphate (ATP), and 50.0 mM D-(+)-glucose. The experiment was performed at 22 °C by collecting 40 spectra with K values ranging from 139 to 8.35×10^3 cm^{-1}, δ from 1 to 2 ms, Δ = 100.0 ms, T_e = 50.0 ms, and τ = 2.50 ms. In the SPLMOD analysis, each region was again fit to two components. Projection onto the diffusion axis shows that four components are present in the mixture with diffusion coefficients of 1.64×10^{-5}, 4.55×10^{-6}, 2.55×10^{-6}, and 8.32×10^{-7} cm^2 s^{-1} for HOD, glucose, ATP, and SDS, respectively. There is one region centered at 4.44 ppm where the glucose and ATP peaks overlap, and although the average diffusion coefficients for these two species are not quite a factor of 2 apart, the fit in this region did satisfy that criterion and, therefore, the two-component fit in this region was not rejected. This illustrates the repulsion effect in which pairs of diffusion coefficients obtained by means of data inversion have larger separations than the true individual values. Figure 5b shows the projections onto the chemical shift axis for each component in the mixture. Comparison of the projection for each component with the 1D NMR spectrum for the mixture shows that the NMR spectrum of each component of the mixture is successfully resolved on the basis of molecular size. The individual spectra were obtained by integration over the diffusion peaks in a narrow range of diffusion values so that small shifts in center positions would not affect the retrieved 1D NMR spectrum.

The DOSY analysis can also be used to detect molecular aggregates. For example, in supramolecular systems, micelles, microemulsion droplets, and bilayer assemblies are found in different size ranges. This is illustrated by the DOSY display shown in Figure 6 of an oil-in-water microemulsion containing 1,3,5-triisopropylbenzene (TIPB), SDS, and Brij-30. In this experiment, 45 spectra were collected at 22 °C in a K range from 278 to 1.50×10^3 cm^{-1}, δ from 1 to 3 ms, Δ = 100.0 ms, T_e = 100 ms, and τ = 3.60 ms. Three distinct diffusing species are identified with diffusion coefficients of 1.80×10^{-5}, 3.00×10^{-6}, and 4.44×10^{-7} cm^2 s^{-1}. The Brij-30 surfactant peak at 3.48 ppm, the TIPB aromatic peak at 7.48 ppm, and the methyl peak at 0.80 ppm are all shown to diffuse at 4.44×10^{-7} cm^2 s^{-1}. The peak at 1.07 ppm corresponds to the CH_2 peaks in both Brij-30 and SDS and shows two distinct components at 4.44×10^{-7} and

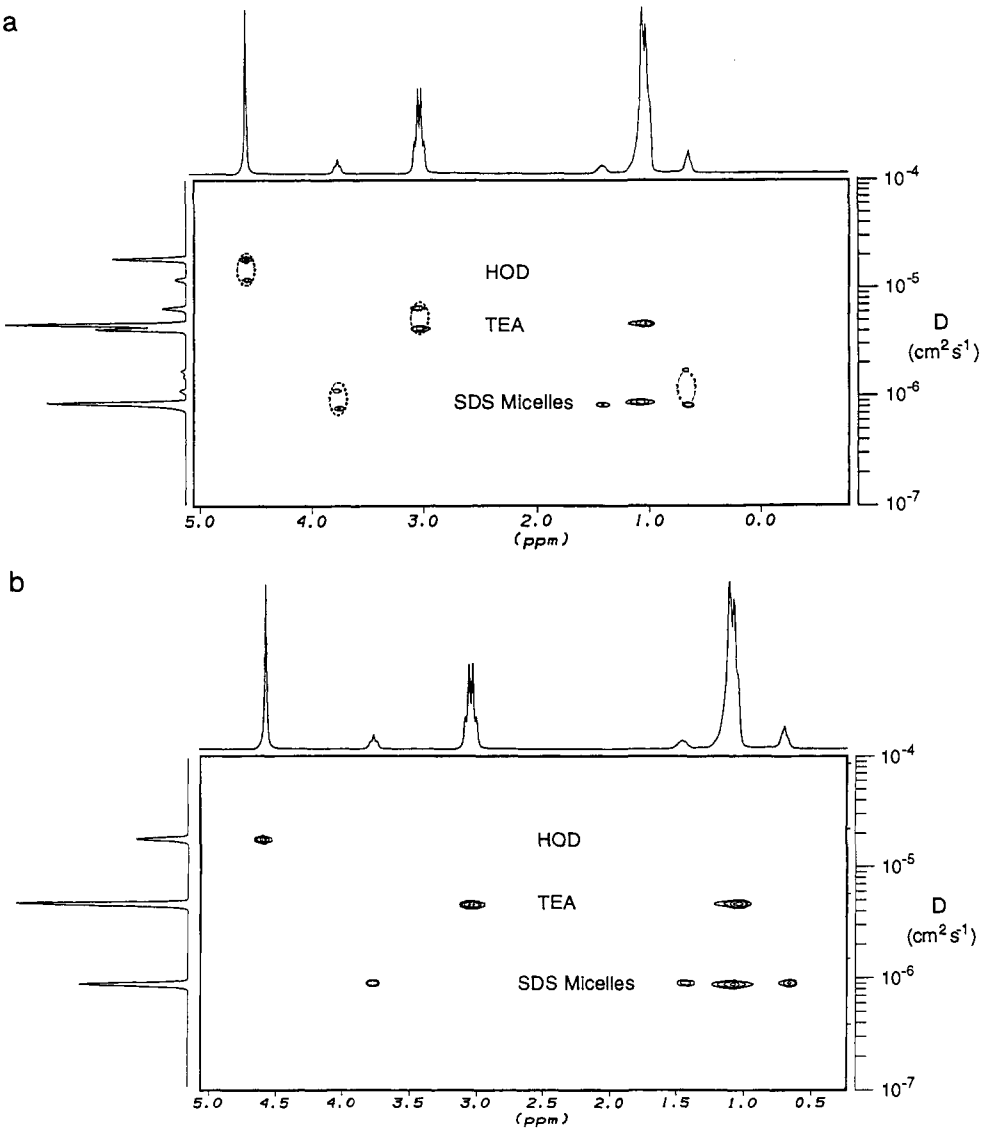

Figure 4. DOSY displays for a mixture containing D_2O, TEA, and SDS micelles: (a) 2D spectrum generated directly from SPLMOD best fit parameters; (b) 2D spectrum generated from SPLMOD parameters after processing with rejection criteria (see text).

3.00×10^{-6} cm^2 s^{-1}. This spectrum suggests that TIPB and Brij-30 are incorporated in the microemulsion droplet and the SDS partitions between the oil droplet and free solution. This is consistent with 2D electrophoretic NMR experiments performed on the same microemulsion.[24] A hydrodynamic radius of 40.4 Å was estimated for the oil dropet by means of the Stokes–Einstein equation by using a value of 1.2027 cP for the viscosity of D_2O at 22 °C without correction for the volume fraction of droplets.[25]

The data columns for TIPB in the microemulsion droplets give no indication of a distribution of particle sizes. This agrees with previous reports that oil is exchanged between droplets in collisions.[26,27] We conclude that peaks in the diffusion spectrum are exchange narrowed and appear at the average diffusion coefficient for the droplets. It can be shown that rapid exchange between sites with different diffusion coefficients yields a Gaussian line shape in the diffusion dimension with a width that depends on the number of exchanges occurring during the diffusion time Δ.[28]

Resolution of Continuous Distributions. For many samples of interest the data cannot be represented by a discrete sum of exponentials, but require instead a continuous distribution of diffusion coefficients. Examples of polydisperse systems include polymers and phospholipid vesicles.[29] To perform the required type of analysis, we have implemented the constrained regularization program CONTIN that fits experimental data, y_k and t_k, to functions of the form[30]

$$y_k \approx \int_a^b f(\lambda,t_k)g(\lambda)\,d\lambda + \sum_{i=1}^{N_L}\beta_i L_i(t_k) \qquad (4)$$

Here N_L and the forms of $f(\lambda,t_k)$ and $L_i(t_k)$ are known, and $g(\lambda)$ and β_i are to be determined by the analysis.[16,17,30] The determination of the distribution $g(\lambda)$ for a given set of y_k and t_k with noise is an ill-posed problem that permits an infinite number of solutions. The program CONTIN approximates solutions to this problem by means of linear least-squares augmented with APK and the principle of parsimony. Thus, of those solutions that are consistent with *a priori* information, the simplest distribution is selected that is consistent with the data. Often smoothness is an important criterion for selecting the simplest solution. Again, the guiding principle is that it is better to lose details than to add more than are required by the data.

(24) He, Q.; Johnson, C. S., Jr. *J. Magn. Reson.* **1989**, *85*, 181.
(25) Kellomaki, A. *Finn. Chem. Lett.* **1975**, 51.
(26) Oldfield, C. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2607.
(27) Clark, S.; Fletcher, P. D. I.; Xilin, Y. *Langmuir* **1990**, *6*, 1301.
(28) Johnson, C. S., Jr. *J. Magn. Reson.*, in press.
(29) Hinton, D. P. Ph.D. Thesis, University of North Carolina, Chapel Hill, NC, 1992.
(30) Provencher, S. W. *CONTIN Users Manual* (*Ver. 2*), Data Analysis Group, EMBL: Heidelberg, Germany, 1984; EMBL-DA07.

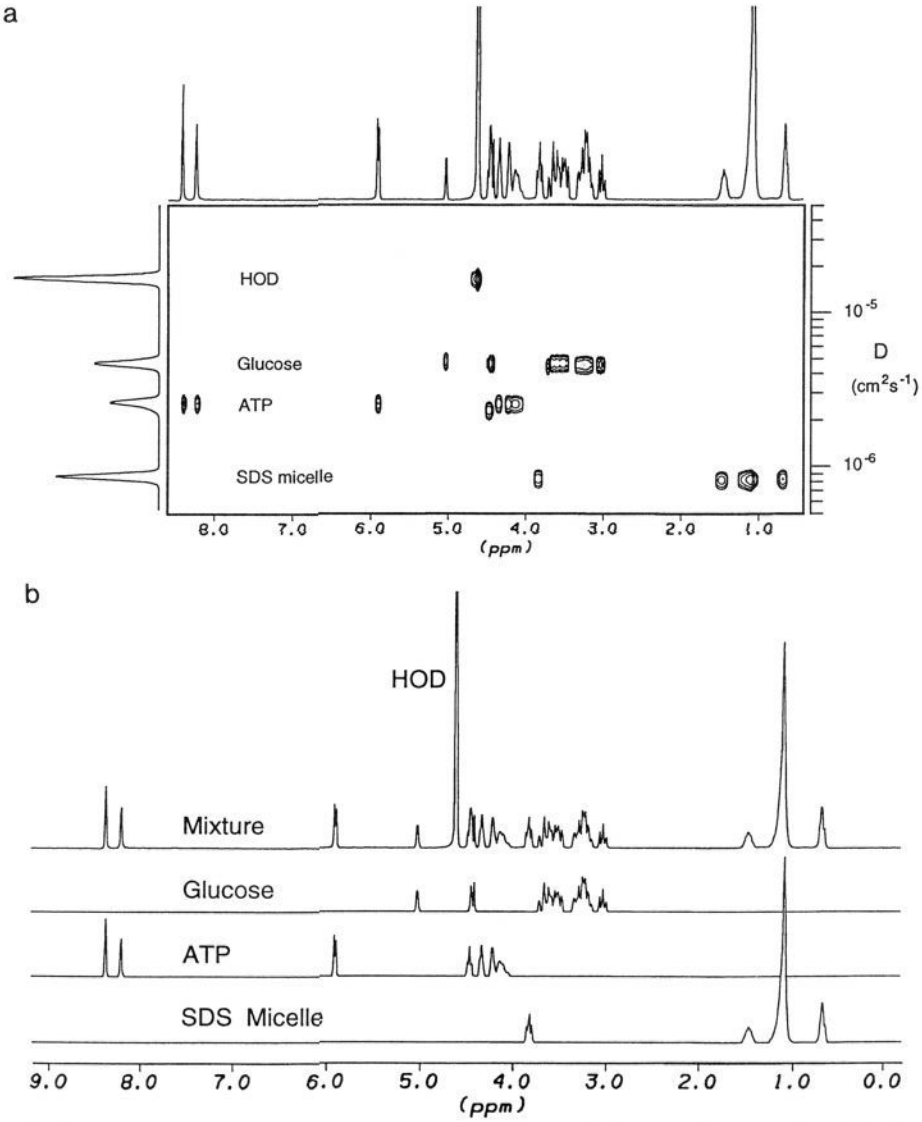

Figure 5. Illustration of diffusion-ordered NMR: (a) DOSY display for a mixture containing D_2O, glucose, ATP, and SDS micelles; (b) individual spectra of components in the mixture obtained by projection of intensities in the various diffusion ranges onto the chemical shift axis.

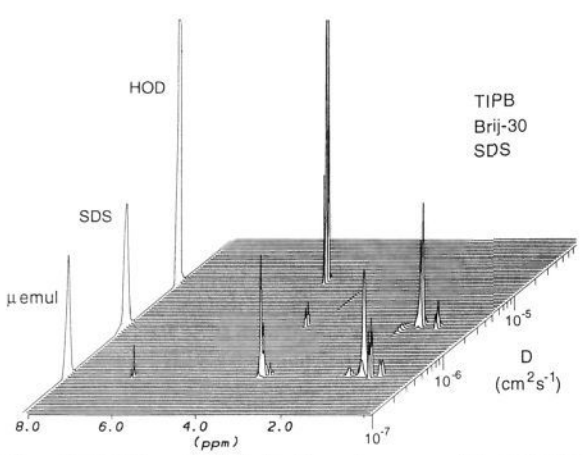

Figure 6. DOSY stack plot display for a mixture containing D_2O, TIPB, and SDS. Microemulsion droplets are evident at 4.44×10^{-7} cm^2 s^{-1}.

Therefore, inversion of DOSY spectra with CONTIN relaxes the requirement that the data be represented by sums of discrete components and allows the resolution of narrow and broad distributions of diffusion coefficients in the same mixture. Other analysis algorithms may equal or even exceed the capabilities of CONTIN in special cases, but this program has important practical advantages. It has been freely distributed and supported for a decade, and independent evaluations have been published.[31-33] CONTIN has become the standard method for the inversion of data representing relaxation in polydisperse systems.

CONTIN offers great flexibility through numerous control variables and user arrays. For the DOSY application, a kernel of the form $f(\lambda,t_k) = \exp(-\lambda t_k)$ was chosen, and the distribution of diffusion coefficients, $g(\lambda)$, was constrained to have only positive values. Unweighted fits were always performed, and $L_i(t_k)$ was set equal to zero, i.e., the base line was assumed to be zero. Also, the points calculated in each diffusion distribution were exponentially spaced, and all distributions contained 50 points. The maximum and minimum diffusion coefficients in the distribution were specified for each analysis, with a typical maximum of 1.00×10^{-4} cm^2 s^{-1} and a minimum chosen to be an order of magnitude lower then the slowest diffusion coefficient expected. The CONTIN analyses were typically performed with the regularizer set so that the smoothest solution consistent with the data was selected as the solution.

(31) Stock, R. S.; Ray, W. H. *J. Polym. Sci., Polym. Phys. Ed.* **1985**, *23*, 1393.
(32) Jakes, J. *Czech. J. Phys.* **1988**, *B38*, 1305.
(33) Kroeker, R. M.; Henkelman, R. M. *J. Magn. Reson.* **1986**, *69*, 218.

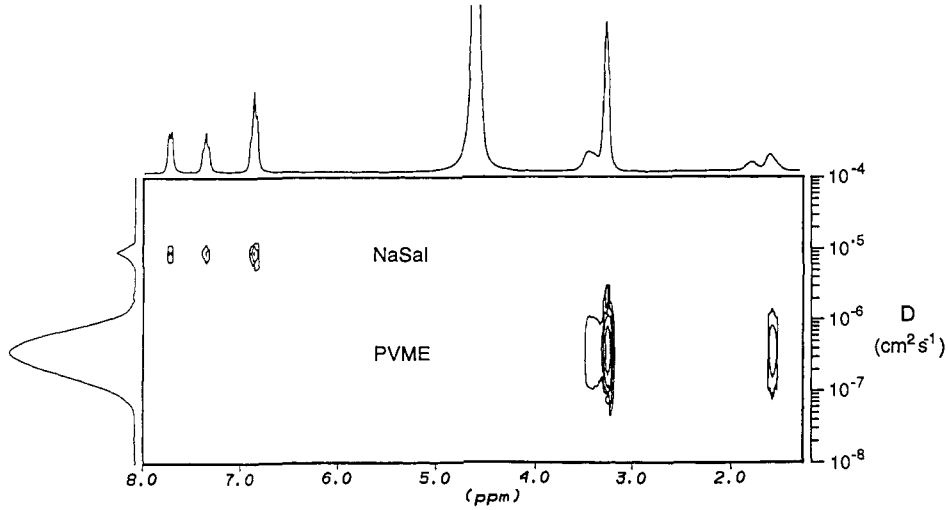

Figure 7. DOSY display processed by CONTIN for a mixture containing D_2O, sodium salicylate, and poly(vinyl methyl ether).

The structure of CONTIN makes the parallel processing of large numbers of data sets impractical, and a separate CONTIN analysis had to be performed at every point above a selected threshold. Also, there is no parameter related to the number of components in the fit other than the range of diffusion coefficients described above. Therefore, implementation of CONTIN for PFG-NMR data sets required much less in-house programming than was required for SPLMOD, and the input program simply permitted the threshold to be chosen. The loss of parallel processing and the extensive nature of the CONTIN analysis led to a great increase in processing time. The analysis required approximately 1 h on an INDIGO workstation to obtain 50 point diffusion coefficient distributions from 500 columns in the 2D data set. A polynomial interpolation routine was then used to generate 512 data points from each 50-point distribution. The CONTIN analysis is marginal for the processing of discrete components in the presence of continuous distributions. Regularization gives some width for an isolated discrete component, but even so there may be only a few points. The polynomial interpolation with a small number of points often leads to triangular or other oddly shaped peaks in the diffusion dimension.

Figure 7 shows the CONTIN-derived DOSY spectrum for a mixture containing 15.0 mM sodium salicylate and 5.0 mg mL^{-1} of the polymer poly(vinyl methyl ether) kindly provided by Prof. T. Wong.[34] The average molecular weight of the polymer is approximately 30 000 g mol^{-1}. The DOSY experiment was performed at 30 °C by collecting 39 spectra with a K range from 417 to 1.08 × 10^3 cm^{-1}, δ from 1 to 2 ms, Δ = 100.0 ms, T_e = 100.0 ms, and τ = 2.50 ms. The salicylate peaks at 7.71, 7.34, and 6.85 ppm all appear at 8.51 × 10^{-6} cm^2 s^{-1} in the diffusion spectrum, and projection onto the diffusion axis shows a narrow distribution. A much broader distribution of diffusion coefficients is obtained for the polymer peaks at 3.26 and 1.58 ppm. This distribution is centered at 3.52 × 10^{-7} cm^2 s^{-1} and has a full-width at half-maximum (FWHM) of 5.12 × 10^{-7} cm^2 s^{-1}. The HOD peak (4.6 ppm) was not included when the DOSY spectrum was generated from the CONTIN distributions.

A second example of DOSY analysis of a mixture containing monodisperse and polydisperse components is shown in Figure 8. This sample contained 75.0 mM glucose and 1.00 wt % methyl cellulose and required a wider range of diffusion coefficients than previous illustrations. The experiment was performed at 22 °C by collecting 26 NMR spectra with a K range from 278 to 2.50 × 10^4 cm^{-1}, δ from 1 to 5 ms, Δ = 350.0 ms, T_e = 50.0 ms, and τ = 5.80 ms. In this spectrum, the peaks for monodisperse glucose overlap peaks for the polydisperse polymer, and DOSY allows one to measure simultaneously the distribution of diffusion coefficients for the polymer and to identify the peaks in the 1D NMR spectrum that correspond to glucose and methyl cellulose. The diffusion coefficient determined for the glucose peak is 3.92 × 10^{-6} cm^2 s^{-1}, and the center of the methyl cellulose distribution is at 1.49 × 10^{-8} cm^2 s^{-1}. The methyl cellulose distribution has a full-width at half-maximum of 8.10 × 10^{-8} cm^2 s^{-1}, showing that this distribution is considerably broader than the one obtained for the poly(vinyl methyl ether) polymer in the last example. Once again, the HOD peak was not included when the DOSY display was constructed from the CONTIN distributions.

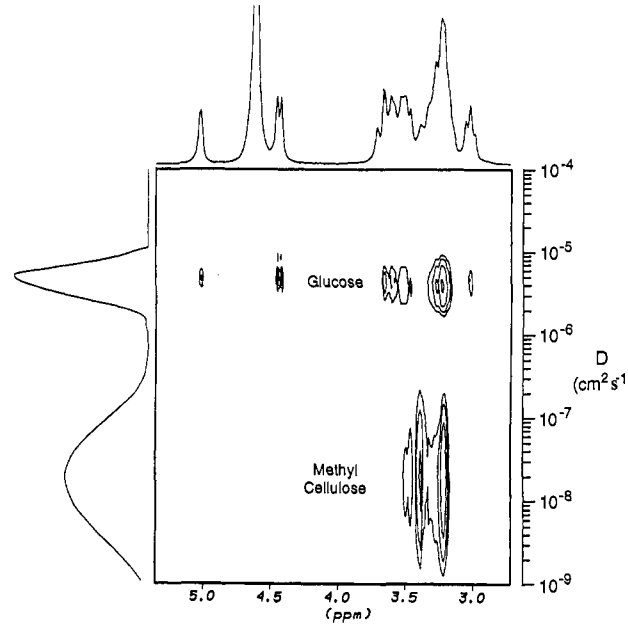

Figure 8. DOSY display processed by CONTIN for a mixture containing D_2O, glucose, and methyl cellulose.

When the same data set was fit with DISCRETE, the glucose peak gave a diffusion coefficient of 4.01 × 10^{-6} cm^2 s^{-1}, but the methyl cellulose peak was centered at 9.47 × 10^{-9} cm^2 s^{-1} and gave no indication of polydispersity. The comparison of the results obtained with DISCRETE and CONTIN shows that the analysis of data sets containing wide distributions of diffusion coefficients with programs that assume discrete components leads to results that are physically unreasonable and to misinterpretation of the data.

It is reasonable to ask how the distributions of diffusion coefficients and particle radii agree with those obtained by other methods. First, we note that PFG-NMR determines tracer diffusion coefficients while dynamic light scattering (DLS)

(34) Wong, T. C.; Lui, C.; Poon, C. D.; Kwoh, D. *Langmuir* **1992**, *8*, 460.

typically determines mutual diffusion coefficients. Also, there is the question of weighting. The DOSY intensities are weighted by the number of protons, and this is essentially mass weighting for hydrocarbons. In contrast, DLS gives mass-squared weighting of diffusion coefficients in addition to a form factor that is dependent on the scattering angle. Previous tests of CONTIN in connection with dynamic light scattering indicate that it is robust and is generally reliable in analyzing experimental data, but has a tendency to oversmooth the distribution.[31,32] The most extensive test of CONTIN-based DOSY thus far involves a comparison of size distributions for unilamellar phospholipid vesicles obtained with electron microscopy, dynamic light scattering, and DOSY. The results turn out to be quite consistent when the appropriate weighting factors are included.[29]

Conclusions

DOSY analyses have been demonstrated for both discrete and continuous distributions of diffusion coefficients. The current implementations of DOSY depend on the inversion programs SPLMOD, for discrete distributions, and CONTIN, for continuous distributions. DOSY is still under development, but in its present form it can provide useful information about mixtures. However, all users should be aware of the ill-conditioned nature of the required data inversions and the possibility of artifacts. The analysis of mixtures containing monodisperse components and the display of discrete diffusion coefficients require great care. SPLMOD, with assumed exponential kernels, permits the parallel analysis of large numbers of data sets and yields fitting parameters with very small reported errors. However, when the experimental decays consist of components that deviate from pure exponentials, this analysis can only respond by introducing additional exponential components into the fit. The resulting diffusion spectra include artifacts and show unreasonably small error-based line widths. We have succeeded in producing accurate DOSY solutions and 2D displays through SPLMOD analyses by introducing criteria for rejecting commonly encountered artifacts. We expect that these procedures will change as NMR technology improves and more experience is obtained with the analysis methods.

The DOSY treatment of continuous distributions is more satisfactory. The CONTIN-based DOSY analysis automatically yields distributions of decay constants from PFG-NMR data for all spectral regions having signals above a predetermined threshold. Polymer solutions and supramolecular systems are suitable for this type of analysis, and information is provided that is difficult to obtain by other means. Examples include the resolution of various molecular species entrapped inside vesicles and in the bilayer and the interaction of polymers with surfactant micelles.

We close by noting that the analysis programs, SPLMOD and CONTIN, are readily available, and the required computing resources are consistent with those normally available for NMR data analysis. Also, the gradient probes and drivers are now commercially available, primarily because of their use in imaging and gradient-assisted 2D NMR. Thus, the implementation of DOSY is within easy reach of most NMR laboratories, and it is reasonable to assume that DOSY-based analytical capabilities will be widespread within the next few years.

Acknowledgment. This work was supported in part under National Science Foundation Grant CHE-8921144 and North Carolina Biotechnology Grant 9212-ARG-0901.