# Generalized Correlation NMR Spectroscopy

**ARTICLE** *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MARCH 2002

Impact Factor: 12.11 · DOI: 10.1021/ja011819v · Source: PubMed



**2 AUTHORS:**

Charles Eads
Procter & Gamble
**24** PUBLICATIONS **374** CITATIONS
SEE PROFILE

Isao Noda
University of Delaware
**346** PUBLICATIONS **9,037** CITATIONS
SEE PROFILE

J|A|C|S
ARTICLES
Published on Web 01/17/2002

# Generalized Correlation NMR Spectroscopy

Charles D. Eads*,† and Isao Noda‡

*Contribution from the Miami Valley Laboratories, The Procter & Gamble Company, P.O. Box 538707, Cincinnati, Ohio 45253, and The Procter & Gamble Company, 8611 Beckett Road, West Chester, Ohio 45069*

Received July 26, 2001. Revised Manuscript Received October 19, 2001

***Abstract:*** The use of generalized correlation analysis (Noda, I. *Appl. Spectrosc.* **1993**, *47*, 1329−1336) for processing two-dimensional arrays of NMR data is described. This analysis produces complex two-dimensional spectra whose cross-peak intensities are related to correlations in the responses of pairs of signals to systematically incremented perturbations. The technique extends and generalizes the applicability of two-dimensional NMR by allowing model-independent analysis of nonperiodic signals as well as model-dependent analysis of such signals. When applied to diffusion-ordered NMR data, the processing scheme produces two-dimensional output spectra having two frequency axes. Relative diffusion coefficients are encoded in the signs and intensities of the cross-peaks. Key properties of the resulting spectra are model-independent, so the approach provides an alternative to traditional DOSY processing and offers advantages for data sets that do not provide pure exponential or Gaussian response curves. When data do conform well to a known response function, the technique provides a method for extracting descriptors in a two-dimensional plot having one axis corresponding to the descriptor and the other axis corresponding to the usual chemical shift scale. Finally, the technique may be used to identify differences in the response functions of closely related samples, generating a one-dimensional spectrum with signals at frequencies whose response functions differ between two samples.

## Introduction

Nearly all implementations of two-dimensional NMR spectroscopy correlate nuclear resonance frequencies in two different experimental time domains using a standard protocol consisting of a preparation period, an evolution period, a mixing period, and a detection period.[1,2] Data are converted from the time domain to the frequency domain by using two-dimensional Fourier transformation or a similar technique. In general, the correlations among frequencies present during the evolution and detection periods are established by a coherence transfer process that occurs during the mixing period. The correlations are encoded as different modulation frequencies for signals in the two time domains. In recent years, two-dimensional NMR spectroscopy has been extended to include correlations that are not encoded as frequency modulations.[3] Based on the work of Johnson, mapping diffusion coefficients along a second axis has become especially widespread. In these experiments, correlations are encoded in the decay constants describing different signals during the evolution period. Signals having closely similar decay constants are correlated in the sense that they are likely to originate from the same molecule. Spectra in which one axis corresponds to nuclear resonance frequencies and the other corresponds to diffusion coefficients are referred to as DOSY (Diffusion Ordered SpectroscopY) spectra.

Procedures for creating two-dimensional spectra in which the correlations arise from processes other than coherence transfer or in which the correlations are not encoded as modulation frequencies are well established in the field of optical spectroscopy.[4,5] These procedures were developed because the standard two-dimensional spectroscopy protocol used for NMR is difficult to apply in optical spectroscopy. One difficulty is that excited-state lifetimes in IR, Raman, and UV spectroscopies are too short to allow easy applications of multiple pulse sequences for manipulation of coherences. Femtosecond laser pulses[6−8] have been used recently to circumvent this limitation.

To achieve a more general approach to 2D spectroscopy, a conceptual departure from the conventional 2D NMR approach was proposed. It was noted that most elements of the radio frequency pulse sequences used in 2D NMR can be regarded as systematically varied perturbations to the spin systems, while the final pulse and the acquisition period serve to probe the response of the system to the perturbation. The method for generating 2D correlation spectra without requiring coherence transfer follows from the observation that the perturbation does not necessarily need to be produced by the same type of process used to probe the response. So long as the response of the system

* Corresponding author. E-mail: eads.cd@pg.com.
† The Procter & Gamble Company, Cincinnati.
‡ The Procter & Gamble Company, West Chester.

(1) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229.
(2) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1987
(3) Morris, K. F.; Johnson, C. S., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 3139.
(4) Noda, I.; Dowrey, A. E.; Marcott, C.; Story, G. M.; Ozaki, Y. *Appl. Spectrosc.* **2000**, *54*, 236A.
(5) Noda, I. *Appl. Spectrosc.* **1993**, *47*, 1329.
(6) Zhao, W.; Wright, J. C. *Phys. Rev. Lett.* **1999**, *83*, 1950.
(7) Tominaga, K.; Yoshihara, Y. *Phys. Rev. Lett.* **1995**, *74*, 3061.
(8) Tanimura, Y.; Mukamel, S. *J. Chem. Phys.* **1993**, *99*, 9496.

to the perturbation occurs on a time scale that can be probed spectroscopically, it is possible to generate correlation spectra based on the response curves at various frequencies. For example, the first two-dimensional IR spectra were generated by using a periodic mechanical deformation of a polymer film monitored by IR spectroscopy.[9]

The key to establishing the correlations is to define a computational process whereby pairs of response curves can be compared. A particularly useful and general process for correlating response curves is to compute the overlap integral of the complex Fourier transforms of response curve pairs. The real and imaginary parts of these integrals provide specific information on the relative behavior of the response curves. Some aspects of this information are independent of the functional form of the response curves. Arranging these pairwise Fourier-space overlap integral values in a two-dimensional matrix creates the two-dimensional correlation spectrum. This approach is known as generalized correlation analysis.[5] In applying generalized correlation analysis, it is not necessary to obtain the response curve pairs from within a single data set. Sources of data can include response curves from different samples probed by the same technique, from models for predicted responses, or even from unrelated spectroscopic techniques.

Even though the development of generalized correlation spectroscopy was initially motivated by 2D NMR spectroscopy, it has evolved into a somewhat different but similarly powerful spectroscopic tool.[10] Interestingly, however, the generalized correlation scheme has rarely been applied to the analysis of NMR data. Nonetheless, there are many applications of NMR spectroscopy for which it is valuable to establish correlations between the behavior of signals at different frequencies, in different samples, using other spectroscopic probes, or among samples and models. The purpose of this report is to enumerate and demonstrate a few applications of generalized correlation analysis to NMR data to show the potential utility of this technique. Benefits of using generalized correlation analysis for NMR data include the following: (1) It is possible to calculate and interpret correlations among NMR response functions at different frequencies without reference to a model and without knowing the mathematical form of the response functions. (2) It is possible to compare the experimental spectral response to a model response if the response function is known or postulated. (3) It is possible to identify differences in the responses of closely related samples.

Model-free data analysis with generalized correlation methods is illustrated by presenting a new processing strategy for DOSY data. The resulting spectrum has two frequency axes, and information on the relative diffusion coefficients is contained in the intensities and signs of the cross-peaks. No curve fitting is used to generate the spectrum. Generalized correlation analysis makes no assumptions about the form of the data whereas conventional curve fitting methods require that the response curves are well-described by exponential decay functions. For many data sets this approach circumvents difficulties that occur during the use of existing strategies for processing DOSY data.[11]

The ability of generalized correlation to compare model and experimental response curves is illustrated by using the same diffusion data and a model data set to produce a traditional DOSY spectrum having chemical shift along one axis and diffusion coefficient along the other axis. The ability to evaluate differences in the behavior of complicated samples is illustrated by presenting a new processing strategy for identifying those signals in a mixture whose response curves are changed by addition of another component. The output of this approach is a one-dimensional spectrum containing signals at those chemical shifts whose response curves are altered by addition of a new component.

## Methods for Generalized Correlation Spectroscopy

Generalized correlation spectroscopy is thoroughly described in the literature.[5] The present description is limited to those aspects and modifications pertinent to analysis of NMR data. The core operation in generalized correlation analysis is to compare two response curves, $y(\omega_m,p)$ and $y(\omega_n,p)$, where $\omega_m$ and $\omega_n$ refer to specific frequencies, and $p$ corresponds to the independent variable describing the progress of the response curve. In the formal treatment of generalized correlation analysis these curves are adjusted by subtraction of a reference function. In the NMR applications described in this report this operation is not performed, though there may be situations for which it is necessary. The Fourier transforms of these curves are calculated by using the following expressions:

$$Y(\omega_m,q) = \int_0^\infty y(\omega_m,p)e^{-ipq}\,dp = \mathrm{R}(\omega_m,q) - i\mathrm{I}(\omega_m,q) \quad (1)$$

$$Y(\omega_n,q)^* = \int_0^\infty y(\omega_n,p)e^{+ipq}\,dp = \mathrm{R}(\omega_n,q) + i\mathrm{I}(\omega_n,q) \quad (2)$$

where the asterisk represents the complex conjugate, $q$ is the variable conjugate to $p$ with respect to Fourier transformation, and the symbols R and I denote the real and imaginary parts of the Fourier transforms, respectively.

The value of the complex generalized correlation matrix $S$ at coordinates $(\omega_m,\omega_n)$ is given by

$$S(\omega_m,\omega_n) = \int_0^\infty Y(\omega_m,q)Y(\omega_n,q)^*\,dq = \Phi(\omega_m,\omega_n) + i\Psi(\omega_m,\omega_n) \quad (3)$$

where

$$\Phi(\omega_m,\omega_n) = \int_0^\infty [\mathrm{R}(\omega_m,q)\mathrm{R}(\omega_n,q) + \mathrm{I}(\omega_m,q)\mathrm{I}(\omega_n,q)]\,dq \quad (4)$$

$$\Psi(\omega_m,\omega_n) = \int_0^\infty [\mathrm{R}(\omega_m,q)\mathrm{I}(\omega_n,q) - \mathrm{I}(\omega_m,q)\mathrm{R}(\omega_n,q)]\,dq \quad (5)$$

Here, $\Phi$ is the real part of the matrix and $\Psi$ is the imaginary part. The real part is symmetric about the diagonal, while the imaginary part is antisymmetric. Thus, the generalized correlation matrix is Hermitian. The real and imaginary parts are also called the *synchronous* and *asynchronous* spectra, respectively. This nomenclature traces back to time series analysis which provided the foundation for generalized correlation analysis. Extension of these expressions to discretely sampled data is straightforward.[12,13]

Data analyzed by generalized two-dimensional correlation is often from a single set of measurements. It is also possible to calculate a generalized correlation matrix in which response curves $y(\omega_m,p)$ and $y(\omega_n,p)$ come from different data sets. There are no restrictions on the nature or origins of the data sets, so NMR data may even be correlated with data from another spectroscopic technique. The process of

(9) Noda, I. *Bull. Am. Phys. Soc.* **1986**, *31*, 520.
(10) Ozaki, Y.; Noda, I., Eds. *Two-Dimensional Correlation Spectroscopy*; American Institute of Physics Conference Proceedings No. 503, Mellville, New York, 2000.
(11) Johnson, C. S., Jr. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203.
(12) Noda, I. *Appl. Spectrosc.* **1990**, *44*, 550.
(13) Noda, I. *J. Am. Chem. Soc.* **1986**, *111*, 8116.

generating spectra from pairs of disparate data sets is known as hetero-spectral correlation.[12] Due care should be exercised in the use of this terminology in the context of NMR data because of the established use of the term hetero-nuclear correlation.

Taking the idea of hetero-spectral correlation a step further, it is not even required for the response curves to be experimentally measured. Thus, one set of response curves may originate in a data set and the other may arise from a model. In this case, the generating function for the model data set can be viewed as a kernel for an integral transform. If a kernel or external data set is used in the correlation analysis, then the axis in the correlation spectrum corresponding to the reference data set will not necessarily be a frequency, but will rather describe the parameter that is varied in the kernel or in the external data set.

A general result of central importance follows directly from the definitions above. If the response curves at positions $\omega_m$ and $\omega_n$ are proportional, i.e., the same to within a real (not complex) scaling factor, the value of the asynchronous part must be zero. If the response curves are different or have a complex scaling factor corresponding to a phase difference, then the asynchronous spectrum at coordinates $(\omega_m,\omega_n)$ is not necessarily zero. It follows from this result that *a model-independent and response curve-independent interpretation is possible for generalized correlation spectra*. This is a key benefit of generalized correlation analysis.

## Examples

**Analysis of Diffusion Data for Mixtures.** Diffusion-ordered NMR spectroscopy (DOSY)[3,11] is a technique for separating signals from different molecules within a mixture based on their differing diffusion coefficients. DOSY has become a standard tool in the analysis of mixtures by NMR. The DOSY experiment requires collection of a series of spectra using a pulse sequence for measuring diffusion coefficients. A parameter such as gradient strength is systematically incremented for each member of the series, leading to a decay of the signal intensity according to a known function of the diffusion coefficient and the systematically varied parameter. After collection the data set must be processed and displayed. A DOSY plot consists of a contour plot having chemical shift on one axis and diffusion coefficient on the other. Calculation of the diffusion coefficient or spectrum of diffusion coefficients at each frequency is usually achieved by using any of a number of algorithms for fitting or deconvoluting the exponential decay curves.[11]

This section demonstrates an alternate approach to evaluating and displaying diffusion-ordered data based on generalized correlation analysis. Starting with the same data set that might be used to calculate a DOSY spectrum, generalized correlation analysis gives a traditional two-dimensional correlation plot, $S(\omega_1,\omega_2)$, having both axes corresponding to frequencies. Qualitative information on relative diffusion coefficients is contained in the signs and intensities of the cross-peaks. In particular, a cross-peak at frequency coordinates $(\omega_m,\omega_n)$ will have a positive, negative, or zero intensity if the diffusion coefficient characterizing the signal at $\omega_m$ is larger than, smaller than, or equal to the diffusion coefficient characterizing the signal at $\omega_n$, respectively.

To illustrate the properties for an idealized case, consider a simple example in which the spectra consist of signals which decay exponentially as a function of the external parameter $p$.

$$y(\omega_m,p) = a_m e^{-b_m p} \tag{6}$$

Response curves of this form arise in diffusion, relaxation, kinetics, and similar NMR experiments. For most implementations of DOSY, $p$ corresponds to the square of the gradient strength multiplied by some other constant and known parameters, $b_m$ corresponds to the diffusion coefficient, and $a_m$ is the signal intensity.[15] It follows directly from eqs 1−5 that the real and imaginary intensities at coordinates $(\omega_m,\omega_n)$ in the generalized correlation spectrum are given by

$$\Phi(\omega_m,\omega_n) = \frac{\pi a_m a_n}{(b_m + b_n)} \tag{7}$$

$$\Psi(\omega_m,\omega_n) = \frac{a_m a_n \ln\left(\frac{b_m}{b_n}\right)}{(b_m + b_n)} \tag{8}$$

The asynchronous component $\Psi$ is proportional to the log of the ratio of the decay constants. This can be negative, positive, or zero depending on the relative values of the decay constants. Hence, for response curves described by eq 6, the signs and intensities of the cross-peaks in the two-dimensional asynchronous spectrum allow one to sort the decay constants for the various molecules. The ratio of the asynchronous and synchronous intensities is a quantitative measure of the log of the ratio of diffusion coefficients.

For many experiments, the signal response curves will not comply with eq 6. For example, the pulsed field gradient for many commercial probes is spatially inhomogeneous, and the resulting response curves are strongly nonexponential.[14] The results of generalized correlation processing of such a data set will still give informative results because matched response functions will give zero intensity and unmatched response functions will give nonzero intensity in the asynchronous spectrum. Therefore this approach may be particularly useful for analysis of DOSY data generated by using many commercially available NMR probes.

In applying pulse field gradient experiments for measuring diffusion coefficients, one has the option of generating either exponential or Gaussian signal response functions by choosing the manner in which the gradient strength is varied.[11] There appears to be certain advantages to the Gaussian form. For example, it is possible to cover a wider range of diffusion coefficients in a single experiment since the gradient strengths do not bunch up at the higher values. A simple quantitative treatment of Gaussian response curves leading to expressions analogous to eqs 7 and 8 is not possible because no simple closed-form analytical expression is available for the Fourier sine transform of the Gaussian function. Nonetheless, it is possible to show numerically that the asynchronous spectrum will have positive, negative, or zero cross-peak intensity depending on if the decay constant of one response curve is greater than, less than, or equal to the decay constant of the response curve to which it is being compared. Therefore, regardless of the functional form of the response functions obtained from diffusion-based experiments, the generalized correlation approach gives spectra that distinguish and rank order signals having different diffusion coefficients.

To demonstrate the use of generalized correlation analysis for evaluation of diffusion data on mixtures, a solution

(14) Hurd, R. E.; Deese, A.; O'Neil-Johnson, M.; Sukumar, S.; van Zijl, P. C. M. *J. Magn. Reson. A* **1996**, *119*, 285.
(15) Stejskal, E. O.; Tanner, J. E. *Chem. Phys.* **1965**, *42*, 288.

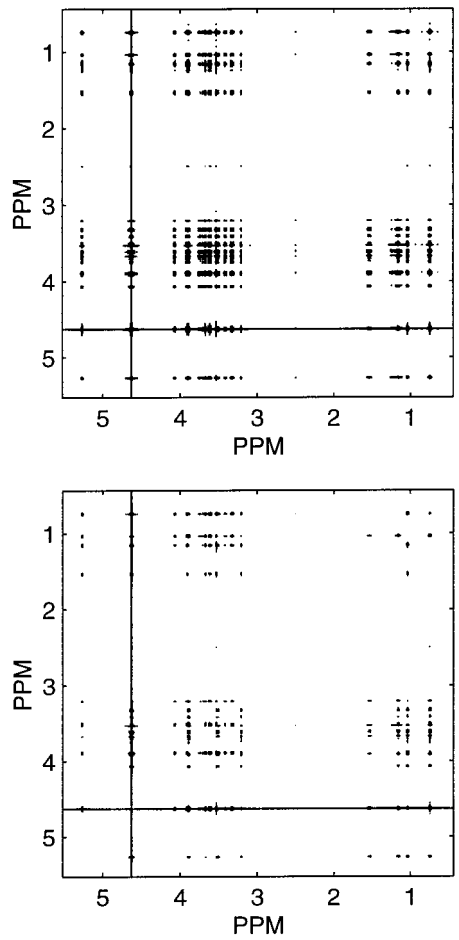


***Figure 1.*** Synchronous (top) and asynchronous (bottom) generalized correlation spectra of a mixture comprised of 32 mM each of sodium dodecyl sulfate, sucrose, ethanol, and methanol in deuterium oxide. Data were acquired as described in the text. Each one-dimensional spectrum was Fourier transformed after application of a half sine wave apodization function having a width of 8196 points. Complex frequency-domain spectra were converted to their absolute values. Complex Fourier transformation along the diffusion dimension was then performed to create a doubly transformed matrix. The complex generalized correlation spectrum was calculated as the matrix product of this matrix with the complex conjugate of its own transpose.

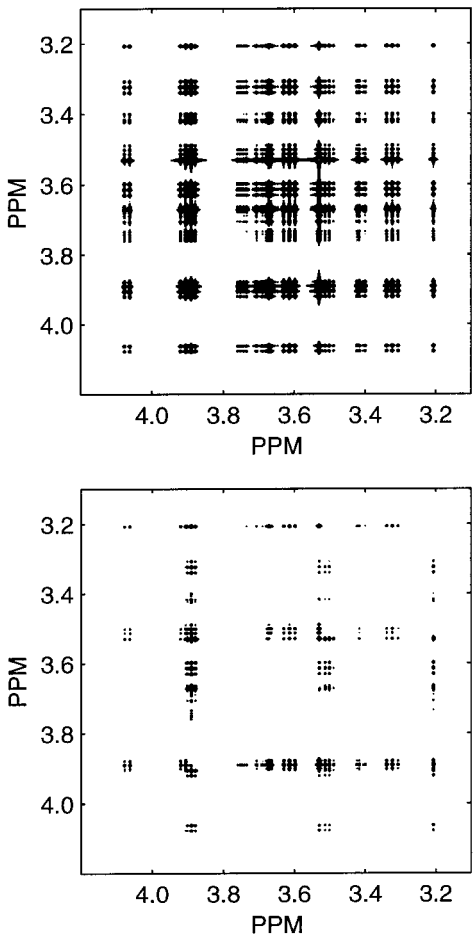


***Figure 2.*** Expanded view from 3.1 to 4.2 ppm of the generalized correlation spectrum shown in Figure 1. The synchronous spectrum (top) shows cross-peaks among signals from all molecules. The asynchronous spectrum (bottom) shows cross-peaks only among signals from molecules having different diffusion coefficients. Peak patterns near the diagonal are indicative of nearly overlapped signals from different molecules. Signs of cross-peaks are not discernible from this contour plot.

comprised of 32 mM sodium dodecyl sulfate (SDS), 32 mM sucrose, and 32 mM ethanol and 32 mM methanol in $D_2O$ was prepared. Data were acquired by using the LED pulse sequence[16] modified with bipolar gradients and convection compensation,[17] from which diffusion coefficients and DOSY spectra can be calculated based on the signal response functions. The experiment was carried out with an array of 32 linearly spaced gradient strengths leading to Gaussian response curves. The NMR probe was of the "ultra linear" design provided by the spectrometer vendor, giving a gradient strength of 60 G/cm at the highest setting. Data were processed with MATLAB (The Mathworks, Inc.) with use of MatNMR[18] to facilitate NMR-specific processing tasks.

Figure 1 shows the full synchronous and asynchronous spectra calculated from this data set. The synchronous spectrum shows cross-peaks among all signals regardless of their molecule of origin. Any horizontal or vertical one-dimensional slice through any cross-peak constitutes a scaled copy of the one-dimensional spectrum. As expected, the asynchronous spectrum shows cross-peaks only among signals from molecules having different diffusion coefficients. The horizontal and vertical "stripes" at the chemical shift of HOD arise from the large number of exchangeable hydrogen atoms provided by the alcohol and sugar components along with the residual HOD in the $D_2O$ stock.

Figure 2 shows expansions of the spectra in the chemical shift range from 3.1 to 4.2 ppm. Within this narrow range of chemical shifts, signals from all components except HOD can be observed. Common features of diffusion-resolved experiments processed using generalized correlation analysis are apparent in these plots. The diagonal peak apparent in the synchronous spectrum near 4.1 ppm corresponds to a sucrose resonance and is missing from the asynchronous spectrum. This is expected, because by definition a diagonal peak correlates a signal with itself and identical or proportional response curves give zero intensity in the asynchronous spectrum. However, there is a pattern of signals visible near the 3.9 ppm diagonal position in the asynchronous spectrum. Both SDS and sucrose have signals at this position. Thus, the presence of diagonal peak patterns in the asynchronous spectrum is indicative of near

(16) Gibbs, S. J.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *93*, 395.
(17) Jerschow, A.; Müller, N. *J. Magn. Reson.* **1997**, *125*, 372.
(18) MatNMR is an NMR processing package written for MATLAB by J. van Beek, distributed (http://www.nmr.ethz.ch/) under the GNU public license.

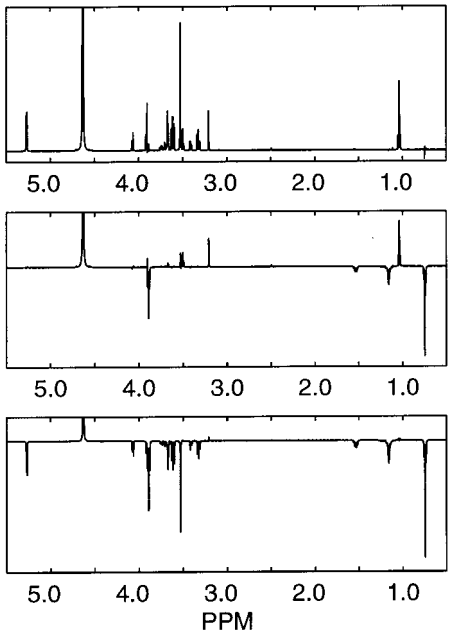


***Figure 3.*** Selected one-dimensional slices of the asynchronous generalized correlation spectrum shown in Figure 1. Slices through SDS (top), sucrose (middle), and ethanol (bottom) show signals only at chemical shifts corresponding to different molecules. Positive peaks indicate higher diffusion coefficients, and negative peaks indicate lower diffusion coefficients.

overlap of signals having different response curves. A similar situation is observed near 3.5 ppm, where signals from ethanol and sucrose are partially overlapped. All the remaining signals in the asynchronous spectrum correlate signals from molecules having different diffusion coefficients.

The signs of cross-peaks in the asynchronous spectrum, which are not apparent from the contour plot, give information on the relative diffusion coefficients of the components. As an illustration, Figure 3 shows one-dimensional horizontal slices through the asynchronous spectrum at chemical shifts corresponding to SDS, sucrose, and ethanol chemical shifts. In each case, signals from molecules having faster or slower diffusion coefficients have positive or negative signs, respectively. Signals arising from atoms in the same molecule have zero intensity in these slices. A mix of positive and negative character can appear at locations for which there is overlap of signals from different molecules, such as near 3.9 ppm in the slice through a sucrose resonance in Figure 3.

On the basis of these features, a simple procedure may be devised for interpreting a generalized correlation spectrum generated from diffusion data sets. First, discount all cross-peaks in the synchronous spectrum that have a corresponding peak in the asynchronous spectrum. The ladders of remaining peaks correspond to spectra of individual molecules or groups of molecules with indistinguishable diffusion coefficients. One can then rank the diffusion coefficients of the molecules based on the signs of cross-peaks among them in the asynchronous spectrum. Simple computer algorithms can be envisioned that use synchronous and asynchronous peak coordinates as input and that produce lists of chemical shifts for each molecule ranked by diffusion coefficient as output.

Whether done manually or with computer automation, the analysis as described so far is deterministic, nearly independent of any model, and does not require a least-squares fitting

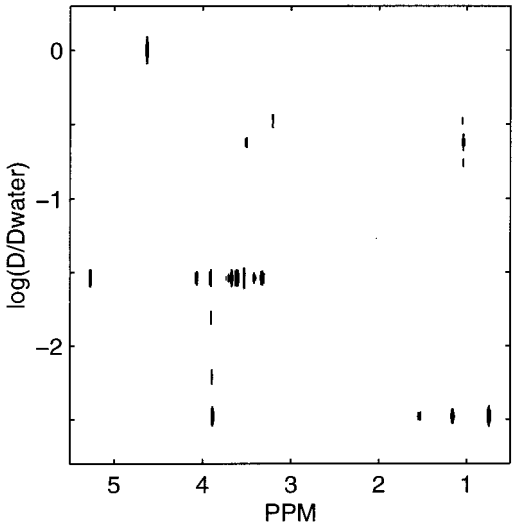


***Figure 4.*** DOSY plot constructed by using generalized correlation analysis.

algorithm. It relies mostly on the observation that proportional response curves give zero peak intensity in the asynchronous spectrum. The assertion that the signs of cross-peaks give a qualitative measure of relative diffusion coefficients is only weakly model dependent, and therefore it is highly robust. Analyzing a generalized correlation plot is about as easy as analyzing traditional 2D spectra such as COSY and TOCSY. The result of such a procedure is a separation of spectra from different components, which is a main goal of DOSY. For data acquired with a high gradient linearity, it is possible to further analyze the correlation spectra to calculate quantitative ratios of diffusion coefficients using cross-peak intensities. It is also possible to analyze the data by comparison to a model, as described in the following section.

**Comparison to Model Data.** One way to determine quantitative parameters describing response curves with use of the generalized correlation approach is to introduce a model data set. The procedure again takes advantage of the fact that when an experimental peak having a particular response curve is cross-correlated with a model having the same (or proportional) response curve, the intensity in the asynchronous spectrum will be zero. Assuming that the analytical expression for the response curve is known, a model data set can be constructed that contains response curves incremented systematically and that span the range of parameters expected in the sample. In the generalized hetero-spectral correlation plots generated with this model data set, one axis will correspond to the chemical shift in the experimental data set, and the other axis will correspond to the index for the model response curves. Locations of zero crossings along the model parameter axis in the asynchronous hetero-spectral correlation plot indicate the index of the model response set having the same response parameter (i.e., diffusion coefficient) as the signal.

Figure 4 shows a traditional DOSY plot calculated from the same data set used to generate Figures 1−3 following this procedure. The model data set consisted of a series of 1000 Gaussian response curves generated with logarithmically spaced diffusion coefficients. Since the asynchronous hetero-spectral correlation slices cross zero at the index of the model response curve having the same decay constant as the experimental signal, the diffusion-resolved spectrum was created by examining the

***Table 1.*** Natural Logarithms of Diffusion Coefficients Compared to Water

| compound | curve fit | zero crossing | smoothed DOSY |
|---|---|---|---|
| methanol | −0.473 | −0.484 | −0.479 |
| ethanol | −0.629 | −0.641 | −0.622 |
| sucrose | −1.544 | −1.549 | −1.540 |
| SDS | −2.467 | −2.462 | −2.480 |

value of the asynchronous spectrum at every point. A preliminary output spectrum was created such that if the asynchronous hetero-spectral correlation curve crossed zero at a given point, the output spectrum was given the value of the corresponding location in the synchronous hetero-spectral correlation curve. Otherwise, the output spectrum was given a value of zero at that point.

Due to experimental uncertainty, the diffusion coefficients determined by this approach for different frequencies corresponding to the same molecule have some scatter. Additional processing was therefore required to properly classify signals with respect to their molecule of origin and to objectively assign a diffusion coefficient to each molecule. The next step in the procedure used to generate Figure 4 was to calculate the projection of the spectrum along the diffusion axis. This projection was then smoothed by convolution with a Gaussian function having a width corresponding to the width of the clusters of diffusion coefficients. The position of the signal at each chemical shift in the preliminary spectrum was then adjusted to the location of the closest maximum in the smoothed projection. Finally, the signals were given a Gaussian width in the diffusion dimension corresponding to the width of the smoothing function used originally for the projection.

Figure 4 shows that signals from HOD, methanol, ethanol, sucrose, and SDS can all be distinguished on the basis of their diffusion coefficients. Spurious signals are observed near 3.9 ppm where sucrose and SDS overlap. The apparent diffusion coefficients for these spurious signals are intermediate between the two contributing materials. Low intensity spurious signals also appear near the upfield ethanol triplet signal. These arise from experimental uncertainty beyond the width of the Gaussian convolution function used to classify the diffusion coefficients in the preliminary output spectrum. By inspection of the asynchronous generalized correlation spectrum of this data set, one can determine which signals in the DOSY spectrum are spurious.

It is worthwhile to consider how the values of diffusion coefficients determined by this procedure compare to those determined by least-squares curve fitting. Signals at chemical shifts corresponding to water, methanol, the upfield tripliet of ethanol, the anomeric hydrogen of sucrose, and the upfield triplet of SDS were selected for this analysis. Diffusion coefficients were determined from traditional least-squares curve fitting to a Gaussian function, from positions of zero crossings in the asynchronous hetero-spectral correlation plot, and from the consensus positions determined in the construction of Figure 4. Table 1 lists the diffusion coefficients compared to water for all of these compounds and methods. This shows that when the form of the response curves is known, generalized correlation analysis gives comparable results to least-squares curve fitting and may be useful for quantitative diffusion studies in addition to qualitative separation of signals from different mixture components.

**Differential Asynchronous Decay Spectra.** It is common to compare two NMR spectra or spectral sets to see if one differs from the other in some regard. For example, in ligand−receptor binding studies conducted with NMR, a frequent strategy is to identify differences in diffusion coefficients or relaxation rates between two samples.[19] The control sample is usually a mix of potential ligands, while the experimental sample contains a macromolecular receptor in addition to the ligands. Because of the reduction in diffusion coefficients or change in relaxation rates of ligands which reversibly bind to large receptors, this approach can be used to identify and quantify binding of the highest affinity ligands in the mix. There are many additional examples of situations in which changes in diffusion coefficients, relaxation rates, or other parameters apply to some mixture of components and not to others following some perturbation. It is advantageous to devise a model-independent data reduction strategy that identifies those spectral components whose behavior with respect to an external parameter $p$ changes from one data set to the next. Generalized correlation analysis provides a means for devising such a strategy.

The protocol is best illustrated with a specific example. Consider a pair of nearly identical mixtures. The first (experimental) sample contains an additive that influences the diffusion coefficients of some components, while the second (control) sample does not contain the additive. The samples have otherwise identical compositions. The goal of the analysis is to identify those signals influenced by the presence of the additive. Identification of the affected signals requires acquisition of a series of spectra for both samples from which diffusion coefficients can be determined. One approach to identifying the components that interact with additive is to directly determine the diffusion coefficients of every signal in both samples, and compare the results peak-by-peak. An alternate strategy involves calculation of the generalized correlation matrix in which response curves $y(\omega_m,p)$ and $y(\omega_n,p)$ originate in the experimental and control spectra, respectively. Such spectra will be called *differential* decay spectra.

The desired information is contained in the diagonal elements of the differential asynchronous decay spectrum. These elements correlate slices having the same chemical shift in the two data sets. The intensity of the asynchronous spectrum will be zero if the two slices have identical response curves. The intensity will be nonzero if the two slices have different response curves. Therefore, the diagonal of the differential asynchronous decay spectrum will appear as a simple one-dimensional NMR spectrum with peaks at those positions for which a difference in the response curves exists between the experimental and control samples. This analysis is model independent. It can be applied to diffusion, kinetics, relaxation, electrokinetics, etc. Quantitative determination of decay constants is of course possible by introducing a model data set with an appropriate kernel.

Figure 5, top, shows the diagonal of a differential asynchronous decay spectrum calculated from two samples, one identical with that used to generate Figure 1, the other differing only in that the mixture additionally contained 10 mM of the nonionic surfactant pentaoxyethylene glycol monooctyl ether (C8E5). This is compared to the normal NMR spectrum of the mixture

(19) Lin, M. F.; Shapiro, M. J.; Wareing, J. R. *J. Am. Chem. Soc.* **1997**, *119*, 5249.

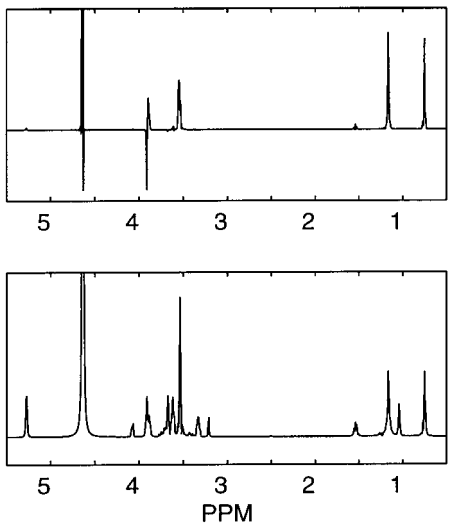


***Figure 5.*** Differential asynchronous decay spectrum from addition of C8E5 (top). The spectrum is compared to the spectrum of the mixture (bottom).

in the lower panel of Figure 5. Introduction of additional surfactant into the mixture is expected to decrease the diffusion coefficient of the surfactant because a larger fraction of the surfactant will reside in the slowly diffusing micellar form. As expected, the peaks in the differential asynchronous decay spectrum occur at frequencies corresponding to signals from the surfactants. The peak near 3.5 ppm arises from overlap of signals from the nonionic surfactant and sucrose. Because of this overlap, the response curves in the experimental and control spectra are different, leading to the appearance of a signal in the differential asynchronous decay spectrum at this position. There is also a feature near 3.9 ppm that corresponds to signals from both SDS and sucrose. The positive aspect of this signal arises from an increase in the diffusion coefficient of SDS. We speculate that the negative aspect arises from a slight change in the chemical shift and/or the line width of this SDS signal that changes the response curve at the chemical shift of sucrose, causing its apparent decay rate to increase. Such slight changes in chemical shifts make a significant contribution to differential asynchronous decay spectra. The time domain data used to generate Figure 5 were truncated to 2048 points before Fourier transformation to eliminate some of these effects. The signal near 4.6 ppm corresponding to water also appears to have contributions from changes in the diffusion coefficient of water and possibly from slight chemical shift or line width changes. Signals from ethanol, methanol, and most of the sucrose signals are absent from the differential asynchronous decay spectrum.

## Discussion

Generalized correlation analysis provides a new approach to analyzing and comparing NMR data sets. The technique extends the applicability of two-dimensional NMR by allowing model-independent analysis of nonperiodic signals as well as model-dependent analysis of such signals. Establishing correlations between signals does not necessarily require a coherence transfer step. The result of such an analysis is a complex two-dimensional spectrum whose real and imaginary parts provide a measure of how similar and how different the signals are, respectively.

The simplest approach calculates the spectrum from response curves originating within a single data set. This was illustrated with a set of diffusion data for a multicomponent mixture, giving an asynchronous spectrum having cross-peaks only among signals with different diffusion coefficients. This particular experiment has a great deal of practical value because, like DOSY, it allows a spectrum of a complex mixture to be resolved into subspectra of individual components based on their differing diffusion coefficients. Partial overlap of signals is apparent from the presence of peak patterns close to the diagonal of the asynchronous spectrum. The spectral appearance is similar to established forms of two-dimensional NMR spectroscopy and should be familiar to practicing spectroscopists. The rules for interpreting the data are quite simple and easily amenable to computer automation. The strategy does not require a high degree of spatial gradient constancy, and it is therefore recommended that this approach be used with many existing gradient probes. We propose the name GECO-DOSY (GEneralized COrrelation Decay Ordered SpectroscopY) for spectra generated with this approach.

Comparison of an experimental data set to a model was also illustrated. In this case, the comparison allows determination of quantitative diffusion coefficients, and provides an alternate approach to the generation of DOSY spectra. Although it is model-dependent, the model may be developed based on interpolation from an *experimental* curve whose *p* axis is stretched or compressed to generate the reference response curve set. Thus, this procedure allows, for example, quantitative interpretation of data acquired with NMR probes having sub-optimal gradient design. Furthermore, when model response curves describe only approximately the response function of the real data, there will still be a model parameter giving an effective match as indicated by a zero asynchronous intensity. The location of this effective match remains useful for qualitative separation of differing components.

Generalized correlation analysis underlies an experimental strategy for identifying mixture components whose behaviors differ with respect to an external perturbation. This was illustrated by changing a surfactant concentration that caused changes in the diffusion coefficient of a subset of solution components. It was straightforward to identify the components whose behavior changed by examining the diagonal trace of the differential asynchronous decay spectrum. This strategy is potentially useful as a method for identifying changes in complex response curves, as might be generated in metabolic and toxicological experiments monitored by NMR. We propose the acronym DAD (Differential Asynchronous Decay) for spectra generated following this procedure.

Finally, it is of value to mention some practical aspects of the use of generalized correlation analysis of NMR data based on our experience so far. The methods appear to place the same demands on the quality of raw spectral data as other processing techniques. Spectrum-to-spectrum phase constancy and flat baselines are no less critical than with other methods. When signals are sufficiently strong and sharp, use of a half sine wave apodization function along with absolute value presentation completely eliminates baseline and phase issues. Applications of hetero-spectral correlation analysis, such as generation of differential asynchronous decay spectra, depend strongly on sample-to-sample reproducibility of chemical shifts if the desired

information is to be obtained from a single one-dimensional slice through the diagonal. Though we have not yet carried out a detailed analysis of the signal-to-noise characteristics of generalized correlation spectra, there does not appear to be an obvious difference compared to other approaches to two-dimensional spectroscopy. In the unlikely event that two signals have truly identical chemical shifts and line widths but different response curves, two-dimensional generalized correlation analysis does not have the ability to separate these signals. If signals from the same molecules are present at other chemical shifts, the overlapping signals will appear different from both of these. If the signals have even slightly different shifts or widths, the presence of near-diagonal peak patterns in generalized correlation spectra provides a powerful diagnostic.

JA011819V