JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Local Covariance Order Diffusion-Ordered Spectroscopy: A Powerful Tool for Mixture Analysis

Adam A. Colbourne, Gareth A. Morris, and Mathias Nilsson*

School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, U.K.

Supporting Information

ABSTRACT: Diffusion-ordered spectroscopy (DOSY) is an important tool in NMR mixture analysis that has found use in most areas of chemistry, including organic synthesis, drug discovery, and supramolecular chemistry. Typically the aim is to disentangle the overlaid, and often overlapped, NMR spectra of individual mixture components and/or to obtain size and interaction information from their respective diffusion coefficients. The most common processing method, high-resolution DOSY, breaks down where component spectra overlap; here multivariate methods can be very effective, but only for small numbers (2−5) of components. In this study, we present a hybrid method, local covariance order DOSY (LOCODOSY), that breaks a spectral data set into suitable windows and analyzes each individually before combining the results. This approach uses a multivariate algorithm (e.g., SCORE or DECRA) to resolve only a small number of components in any given window. Because a small spectral region should contain signals from only a few components, even when the spectrum as a whole contains many more, the total number of resolvable chemical components rises dramatically. It is demonstrated here that complete resolution of component spectra can be achieved for mixtures that are much more complex than could previously be analyzed with DOSY. Thus, LOCODOSY is a powerful, flexible tool for processing NMR diffusion data of complex mixtures.

Diffusion-ordered spectroscopy (DOSY)[1−3] is a method of separating the individual NMR spectra of molecules in a mixture according to their diffusion behavior, which in turn typically depends on size (hydrodynamic radius) and intermolecular interactions.[4−6] Although DOSY is strictly only a data-processing method, the name has been widely adopted to include the pulsed field gradient (PFG) NMR experiments that can be used to produce data for DOSY processing. To measure the diffusion coefficients of the components in a mixture, a series of spectra with increasing gradient strength is recorded. The increasing gradient strength causes progressive attenuation of each signal; the extent of this attenuation depends upon the rate at which the component responsible for the signal diffuses.

A signal with an ideal (purely exponential) decay would be described by the Stejskal−Tanner equation,[7]

$$S = S_0 e^{-D\gamma^2\delta^2 g^2 \Delta'} \quad (1)$$

where S is the signal amplitude, S_0 is the signal amplitude had there been no diffusion, D is the diffusion coefficient, δ is the duration of the gradient pulse, γ is the gyromagnetic ratio, g is the strength of the gradient, and Δ' is the diffusion time corrected for the effects of finite gradient pulse width. Inevitably, the experimental signal deviates to some extent from a pure exponential decay; as noted below, the effect of variation in the magnitude of g across the sample is a major source of such deviation for many probes.[8]

High-resolution DOSY (HR-DOSY)[9] assumes that each signal in the NMR spectrum contains only one component and fits the corresponding decay to some form of eq 1. Provided that there is no spectral overlap, impressive resolution in the diffusion dimension can be achieved using this method; peaks from a single chemical component appear in a typical DOSY spectrum along a horizontal line at the diffusion coefficient for the molecule. However, in all but the simplest of mixtures, some signals overlap. Two or more overlapping signals have superimposed decays that, when fitted to a single exponential, typically (although not always[10]) give a value of D intermediate between those for the species involved. In heavily overlapped spectra, this can completely prevent interpretation. An example of severe overlap can be seen in Figure 1a between 1 and 2 ppm, where peaks from three chemical components overlap to varying degrees, making this region particularly difficult to interpret. All processing of DOSY data is sensitive to any systematic error present; in this investigation, the corrections for the effects of nonuniform PFGs were made where necessary.[11] Reference deconvolution[12] was used to correct for a variety of other sources of error stemming from hardware shortcomings (e.g., phase and frequency inconsistencies).

A variety of approaches can be employed to deal with the consequences of spectral overlap. Extending the 2D DOSY experiment to 3D DOSY (e.g., using HMQC[13,14] or COSY[15,16]) can significantly reduce the overlap. Unfortunately, this requires much longer data acquisition times and more complicated processing and can still fail to resolve complex mixtures. Peak overlap can also be efficiently reduced by collapsing the multiplet structure, as in pure shift NMR;[17] this can be a useful addition to DOSY when sensitivity allows. For cases where overlap is unavoidable, various processing techniques have been developed to enable resolution of the diffusion dimension; these include fitting signal decays to a sum of exponentials[18,19] or even to continuous distributions.[20,21] Such methods are very demanding with respect to the quality of experimental data; these demands rise steeply with the number of components. Thus, in Figure 1b, a DOSY plot has been constructed

Received: January 18, 2011
Published: April 29, 2011

dx.doi.org/10.1021/ja2004895 | J. Am. Chem. Soc. 2011, 133, 7640–7643

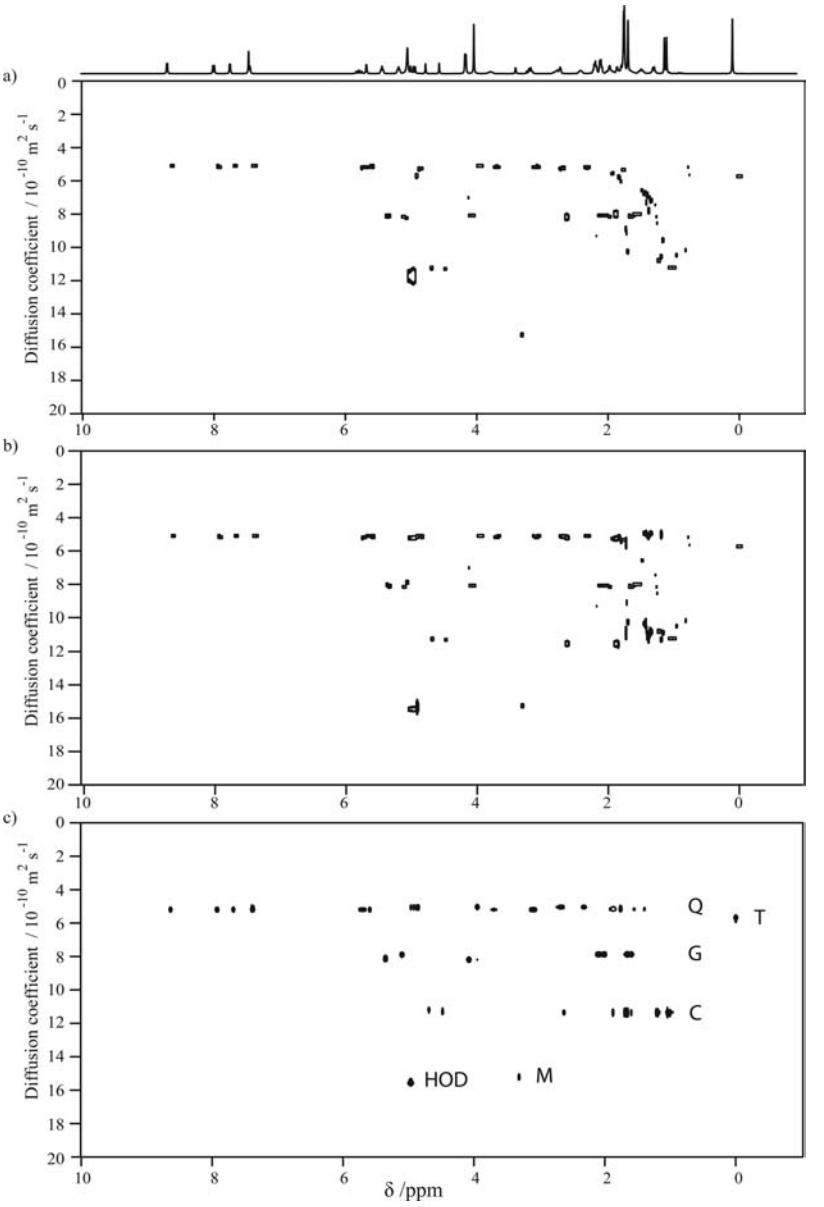

Figure 1. 2D DOSY plots produced by different processing methods, applied to data acquired on a Varian INOVA 400 MHz spectrometer from a mixture of quinine (Q), geraniol (G), and camphene (C) in methanol-d_4 (M) with TSP (T) as a reference. All processing was performed with compensation for the effects of nonuniform field gradients.[8] (a) HR-DOSY[9] fitting; (b) biexponential fitting;[18] (c) LOCODOSY processing using the SCORE[25] algorithm. The considerable overlap at 0−2 ppm impedes the interpretation of the HR-DOSY spectrum (a). In the biexponential fit (b), the overlap is partially resolved, and in the LOCODOSY spectrum (c), all of the signals appear at the correct positions in the diffusion dimension.

using biexponential fitting. It is clearer here than in Figure 1a that there are three components present in the region $(4-12) \times 10^{-10}$ m^2 s^{-1}, but there is still significant ambiguity and the spectrum remains difficult to interpret.

The methods introduced so far are univariate, that is, they analyze one spectral peak or one frequency at a time. Another approach within which several applications have been developed is multivariate processing. Here, rather than just one peak, information from the whole data set is used at the same time to identify similarly attenuating signals across the spectrum and group them as one chemical component (i.e., a single source of variance). The result is to decompose an experimental data set into a set of 1D spectra, each of which (ideally) represents one individual chemical component of the mixture sample.[22−25] Such a multivariate decomposition can be described in matrix form as

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \qquad (2)$$

where the experimental data set ($\mathbf{X}$) is described as a Kronecker product of the decays ($\mathbf{C}$) and the spectra ($\mathbf{S}$), $\mathbf{E}$ is the matrix of residuals, and $^{\mathrm{T}}$ denotes the transpose.

The advantage of this approach is that decay information from peaks in overlap-free regions of the spectrum is used to facilitate the interpretation of the more complex parts, aiding in the resolution of overlapped signals. However, such methods are limited in the number of components they can resolve (3−4 appears to be a practical limit for mixtures of small molecules) and therefore fail for more complex mixtures.

dx.doi.org/10.1021/ja2004895 |*J. Am. Chem. Soc.* 2011, 133, 7640–7643

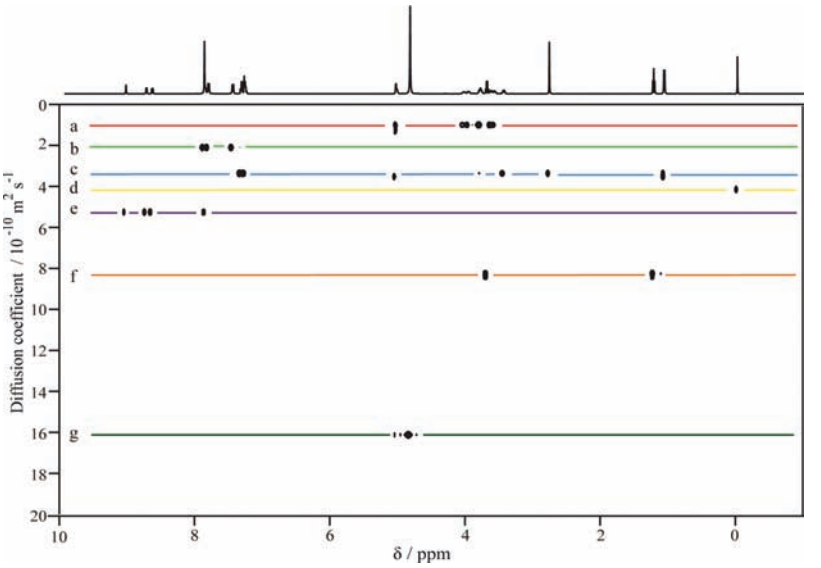

Figure 2. 2D results produced by LOCODOSY processing (using DECRA[22] for the multivariate decomposition) for data acquired on a Bruker Avance II+ 500 MHz spectrometer. The sample contained seven components, all of which were successfully resolved: (a) dextran; (b) tartrazine; (c) ephedrine; (d) TSP; (e) nicotinic acid; (f) ethanol; (g) HOD. The use of either SCORE[25] or DECRA on the entire spectral width failed to resolve the individual component spectra.

Here we present a new hybrid method, local covariance order DOSY (LOCODOSY), that exploits the fact that while a mixture spectrum may contain a large number of components, the number of signals that overlap in any given region of the spectrum is often very small. The principle is simple: take small, separate windows of the spectrum and apply multivariate processing to each individually, with the aim of reducing the number of components required per analysis while retaining the multivariate advantage. The results from processing these separate windows with a multivariate method such as speedy component resolution (SCORE)[25] or the direct exponential curve resolution algorithm (DECRA)[22] may then be combined into a standard DOSY 2D contour plot. The method has been implemented in MATLAB version 2009a and built into the DOSY Toolbox,[26] which can be found on the Web[27] or obtained directly from the corresponding author. Also implemented in the Toolbox alongside LOCODOSY is a set of basic clustering algorithms for automating the aggregation of the spectral fragments resolved by processing into complete component spectra. This produces a 1D spectrum for each identified component that can be compared with the 2D plot. Manual selection and viewing of individual components, as well as control over some key processing parameters, have also been included to maximize flexibility.

The idea of local multivariate fitting of small spectral regions has been described previously,[28] and illustrated with synthetic data, with the aim of improving the statistics of DOSY fitting for nonoverlapped signals. Here, in contrast, we demonstrate the power of local multivariate analysis for the decomposition of overlapping spectra. It is shown that in appropriate cases (no more than three overlapping signals, modest dynamic range), this approach can almost double the number of chemical components cleanly resolvable in real experimental data, as illustrated by Figure 1c and Figure 2.

LOCODOSY requires a spectrum to be divided into suitable windows, each containing the signals of only a few components, before the independent multivariate decompositions are performed. In most cases, this segmentation can be automated (as was done for Figures 1c and 2), although manual division is sometimes advantageous (see Figure 3). To segment the spectrum automatically, a threshold for signal intensity is set by the user (just as in HR-DOSY), and contiguous areas of the spectrum within which the peak intensity is greater than the threshold are identified as windows to be processed.

A number of data points equivalent to ~25 Hz is added to either side of each window to ensure that the bases of all peaks are included. This crude but effective algorithm performed well with the data sets in Figures 1 and 2, but there is ample scope for more complex approaches where appropriate. SCORE and DECRA both require the number of components (i.e., the chemical rank) for each window to be specified. Ideally, determination of the chemical rank should be automatic, as each window in the segmented spectrum could potentially have a different unknown number of components. Many sophisticated methods for the determination of chemical rank are available,[29] but the need for such a method was bypassed here using simple singular value decomposition (SVD). Because the effect of overestimating the number of components in SVD is to cause multivariate fitting to fail in a predictable manner, it is simple to estimate the chemical rank by decreasing the number of components until a satisfactory result is obtained. The symptoms used here to identify an overestimated chemical rank were (i) a fitted D value outside the expected range [$(0-25) \times 10^{-10}$ m^2 s^{-1} here] and/or (ii) a component contributing less than a defined proportion (in principle determined by the experimental noise level) of the described variance (5% here). The multivariate methods used independently and within LOCODOSY also fail when two components have D values that are too similar to be resolved (typically, a difference of ~30% in D is required for resolution); their spectra become combined and show an intermediate value of D, as occurs when peaks overlap in HR-DOSY.

So far we have considered only samples in which the relative concentrations of components are similar. This is where the advantages of LOCODOSY are greatest. Where the data have a high dynamic range, multivariate processing can struggle, as

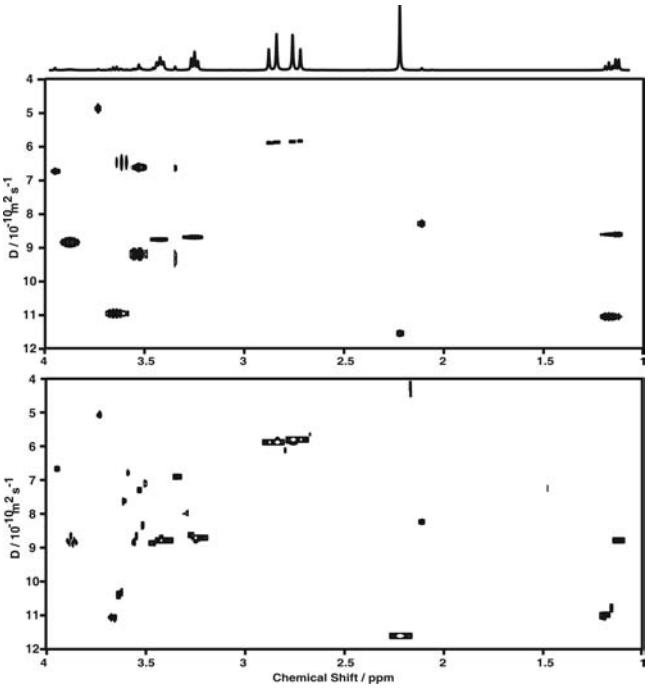

Figure 3. (a) LOCODOSY results for a sample containing 10% D_2O and 90% Red Bull Sugarfree drink, with acetone as a reference. (b) HR-DOSY results for the same sample. Data were acquired on a Varian INOVA 400 MHz spectrometer using presaturation of the water signal.

smaller, less statistically significant signals become increasingly difficult to separate from larger ones. One example of such a mixture is the energy drink Red Bull Sugarfree, in which the detectable signals span 3 orders of magnitude, from the very strong citrate and taurine signals down to the very weak signals of the B vitamins.[30] Here the unsegmented SCORE processing failed, showing very severe cross-talk between resultant spectra, whereas HR-DOSY performs well until peaks begin to overlap (Figure 3b). The use of LOCODOSY, here with manual segmentation to separate regions of low and high intensity (Figure 3a), allowed considerable clarification of the DOSY spectrum and the assignment of several major components (see the Supporting Information).

LOCODOSY is a new and powerful method for processing DOSY data that with automatic segmentation and determination of chemical rank is simple and effective to apply. It allows the resolution of a considerably larger number of mixture components than was previously possible, as well as cleaner separation of overlapped spectra, although the known limits on resolution of component spectra by multivariate methods apply in each window. The data-processing time required depends on the number of windows used and hence the number of decompositions performed, but the use of the relatively fast SCORE and DECRA multivariate fitting algorithms ensures that processing times are modest (typically 2 min).

■ ASSOCIATED CONTENT

Supporting Information. SCORE processing results for the sample in Figure 1 (Figure S1) and assignments for the Red Bull Sugarfree DOSY spectra (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
mathias.nilsson@manchester.ac.uk

■ ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (Grants EP/E05899X/1 and EP/H024336/1). A.A.C. thanks the EPSRC for a DTA studentship.

■ REFERENCES

(1) Johnson, C. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203.
(2) Morris, G. A. In *Encylopedia of Magnetic Resonance*; Harris, R. K., Wasylishen, R. E., Eds.; Wiley: Chichester, U.K., 2009; DOI: 10.1002/9780470034590.emrstm0119.pub2.
(3) Stilbs, P. *Anal. Chem.* **1981**, *53*, 2135.
(4) Evans, R.; Haiber, S.; Nilsson, M.; Morris, G. A. *Anal. Chem.* **2009**, *81*, 4548.
(5) Morris, K. F.; Stilbs, P.; Johnson, C. S. *Anal. Chem.* **1994**, *66*, 211.
(6) Zielinski, M. E.; Morris, K. F. *Magn. Reson. Chem.* **2009**, *47*, 53.
(7) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42*, 288.
(8) Hurd, R. E.; Deese, A.; Johnson, M. O.; Sukumar, S; van Zijil, P. C. M. *J. Magn. Reson., Ser. A* **1996**, *119*, 285.
(9) Barjat, H.; Morris, G. A.; Smart, S.; Swanson, A. G.; Williams, S. C. R. *J. Magn. Reson., Ser. B* **1995**, *108*, 170.
(10) Botana, A.; Aguilar, J. A.; Nilsson, M.; Morris, G. A. *J. Magn. Reson.* **2011**, *208*, 270.
(11) Connell, M. A.; Bowyer, J. P.; Bone, P. A.; Davis, A. L.; Swanson, A. G.; Nilsson, M.; Morris, G. A. *J. Magn. Reson.* **2009**, *198*, 121.
(12) Morris, G. A. In *Encylopedia of Magnetic Resonance*; Harris, R. K., Wasylishen, R. E., Eds.; Wiley: Chichester, U.K., 2009; DOI: 10.1002/9780470034590.emrstm0449.
(13) Barjat, H.; Morris, G. A.; Swanson, A. G. *J. Magn. Reson.* **1998**, *131*, 131.
(14) Wu, D. H.; Chen, A. D.; Johnson, C. S. *J. Magn. Reson., Ser. A* **1996**, *123*, 215.
(15) Nilsson, M.; Gil, A. M.; Delgadillo, I.; Morris, G. A. *Chem. Commun.* **2005**, 1737.
(16) Wu, D.; Chen, A.; Johnson, C. S., Jr. *J. Magn. Reson., Ser. A* **1996**, *121*, 88.
(17) Nilsson, M.; Morris, G. A. *Chem. Commun.* **2007**, 933.
(18) Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A. *Anal. Chem.* **2006**, *78*, 3040.
(19) 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; pp 304−319.
(20) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 229.
(21) Delsuc, M. A.; Malliavin, T. E. *Anal. Chem.* **1998**, *70*, 2146.
(22) Antalek, B. *Concepts Magn. Reson.* **2002**, *14*, 225.
(23) Stilbs, P.; Paulsen, K. *Rev. Sci. Instrum.* **1996**, *67*, 4380.
(24) Van Gorkom, L. C. M.; Hancewicz, T. M. *J. Magn. Reson.* **1998**, *130*, 125.
(25) Nilsson, M.; Morris, G. A. *Anal. Chem.* **2008**, *80*, 3777.
(26) Nilsson, M. *J. Magn. Reson.* **2009**, *200*, 296.
(27) http://personalpages.manchester.ac.uk/staff/mathias.nilsson/software.htm (accessed Jan 18, 2011).
(28) Stilbs, P. *J. Magn. Reson.* **2010**, *207*, 332.
(29) Wasim, M.; Brereton, R. G. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 133.
(30) Simpson, A. J.; Shirzadi, A.; Burrow, T. E.; Dicks, A. P. *J. Chem. Educ.* **2009**, *86*, 360.