Anal. Chem. **2008,** *80,* 3777–3782

Speedy Component Resolution: An Improved Tool for Processing Diffusion-Ordered Spectroscopy Data

Mathias Nilsson* and Gareth A. Morris

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

Diffusion-based NMR techniques (e.g., diffusion-ordered spectroscopy, DOSY), which can be used to distinguish between the signals of different components of a mixture, are steadily gaining in popularity. When processing data from a DOSY experiment it is often desirable to reconstruct the spectra of individual components; here, multivariate methods that take advantage of the covariance between the resonances of a given component can often be advantageous. This paper presents a minor variation on the established CORE method, speedy component resolution (SCORE), that gives a major improvement in performance. In common with CORE it can use any experimental sampling scheme and is adaptable to different experimental decay shapes, but unlike CORE it is very fast and relatively insensitive to starting guesses. The method is demonstrated on a mixture of quinine, geraniol, and camphene in deuteriated methanol, where all four component spectra can be extracted in less than 15 s.

Diffusion-based NMR experiments can be very useful in mixture analysis, as under favorable conditions both the spectra of the individual components and their respective diffusion coefficients and concentrations can be recovered.[1–3] Such pulsed field gradient NMR experiments are often referred to generically as diffusion-ordered spectroscopy (DOSY) experiments, after the most common processing scheme. In a DOSY experiment, a series of spectra is measured with increasing gradient strength and hence increasing diffusional attenuation, and the signal decay is fitted to some form of the Stejskal−Tanner equation[4]

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

where S is the spin or stimulated echo signal amplitude, S_0 is the amplitude in the absence of diffusion, γ is the magnetogyric ratio, g is the gradient amplitude, and Δ' is the diffusion time corrected for the effects of the finite gradient pulse width δ. In DOSY, each individual peak is fitted to eq 1 and a two-dimensional plot is calculated with the NMR spectrum as one dimension and the diffusion coefficient as the second, where the peak shape in the diffusion dimension is a Gaussian centered on the fitted diffusion coefficient with a peak width determined by the statistics of the fit. In the high-resolution approximation,[5] each peak is assumed to belong to a single species, and its decay is fitted to a single exponential; when this assumption holds, the resolution in the diffusion dimension can be better than 0.5%. However, spectral overlap is the norm in all but the simplest mixtures, and makes interpretation of a DOSY plot much more difficult: where resonances overlap, the resulting apparent diffusion coefficient is a compromise value between those of the compounds involved. Although an expert user can often recover a lot of useful information from such a DOSY plot, complementary processing approaches can be very helpful.

Alternatives to high-resolution DOSY range from fitting signal decays to a mixture of exponentials to fitting to a continuous distribution. Practical implementations include biexponential fitting,[6] SPLMOD,[7] CONTIN,[8] and MaxEnt.[9] Unfortunately resolving superimposed exponentials in experimental data is far from straightforward,[10] and all of these methods lead to much greater uncertainty in the exponential parameters than does monoexponential fitting. A further complication in resolving superimposed signal decays is that hardware limitations mean that experimental data never exactly follow the theoretical pure exponential dependence on the square of gradient amplitude g seen in eq 1. The commonest source of such systematic errors is spatial nonuniformity of the pulsed field gradient.[11] Nonuniform gradient (NUG) effects cause a monodisperse sample to give a nonexponential signal decay that is indistinguishable from the behavior of a polydisperse sample in a perfect gradient.[6] Thus, given a sufficient signal-to-noise ratio, a multiexponential fit will always fit experimental data better than a monoexponential fit, even for a single component. In a DOSY spectrum constructed using multiexponential fitting, this results in false peaks in the diffusion dimen-

* To whom correspondence should be addressed. E-mail: mathias.nilsson@manchester.ac.uk.

(1) Johnson, C. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203–256.
(2) Morris, G. A. In *Encyclopedia of Nuclear Magnetic Resonance, Advances in NMR*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons, Ltd.: Chichester, U.K., 2002; Vol. 9, pp 35−44.
(3) Stilbs, P. *Anal. Chem.* **1981**, *53*, 2135–2137.
(4) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42*, 288–292.
(5) Barjat, H.; Morris, G. A.; Smart, S.; Swanson, A. G.; Williams, S. C. R. *J. Magn. Reson., Ser. B* **1995**, *108*, 170–172.
(6) Nilsson, M.; Connell, M. A.; Davis, A. L.; Morris, G. A. *Anal. Chem.* **2006**, *78*, 3040–3045.
(7) 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..
(8) Provencher, S. W. *Comput. Phys. Commun.* **1982**, *27*, 229–242.
(9) Delsuc, M. A.; Malliavin, T. E. *Anal. Chem.* **1998**, *70*, 2146–2148.
(10) Istratov, A. A.; Vyvenko, O. F. *Rev. Sci. Instrum.* **1999**, *70*, 1233–1257.
(11) Hurd, R. E.; Deese, A.; Johnson, M. O.; Sukumar, S.; van Zijl, P. C. M. *J. Magn. Reson., Ser. A* **1996**, *119*, 285–288.

10.1021/ac7025833 CCC: $40.75 © 2008 American Chemical Society
Published on Web 04/12/2008

sion.[6] Fortunately, it is possible to determine a much better approximation to the experimental decay shape by mapping the spatial dependence of field gradient strength and signal amplitude for a given probe and pulse sequence, using experiments on a sample of known diffusion coefficient. Integration of the calculated decay as a function of position over the sample active volume yields the model signal decay; this can then be parametrized, for example, as a sum of two error functions,[12] or as has been used previously by the authors,[2,6,13–15] as an exponential of a power series:

$$S = S_0 \exp\left(-\sum_{n=1}^{N} c_n \sigma^n\right) \tag{2}$$

where

$$\sigma = D\gamma^2\delta^2 g^2 \Delta' \tag{3}$$

In many processing schemes, this new equation can be used as the target function, instead of the original Stejskal−Tanner equation (eq 1) and can alleviate problems with false positives in bi- and multiexponential fitting.[6]

All of the above methods are univariate: each signal peak is treated individually. However, in the absence of magnetization exchange all resonances from a certain compound attenuate in exactly the same way, whether exponentially or not, as a function of gradient strength. With the use of multivariate processing, this covariance can be exploited to improve the separation of the signals of different components, treating the contribution of a given mixture component to the experimental data as a complete spectrum rather than as a set of uncorrelated spectral points. The covariance advantage makes it simpler to separate the contributions of different components to an overlapped signal decay, by using information from other, nonoverlapped (or just differently overlapped) signals.

The basis of most multivariate methods is to attempt to represent a DOSY data set $X(f,g)$ recording signal as a function of frequency f and pulsed field gradient amplitude g using an ideal model function

$$I(f,g) = \sum_{i=1}^{M} S_0^i(f)\, C^i(g) \tag{4}$$

constructed from a set of M spectra $S_0^i(f)$ and decay functions $C^i(g)$ that are varied systematically to minimize the variance between $X(f,g)$ and $I(f,g)$. In matrix form this corresponds, as illustrated in Figure 1, to decomposing the data set $\mathbf{X}$ into a set of component spectra $\mathbf{S}$ and the corresponding decays $\mathbf{C}$

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \tag{5}$$

where $\mathbf{E}$ is the experimental error (noise) and $^{\mathrm{T}}$ denotes the transpose. Matrices $\mathbf{X}$ and $\mathbf{E}$ have dimensions ng × nf, where ng is the number of different gradient amplitudes used and nf is the number of points in the NMR spectrum; $\mathbf{C}$ has dimensions ng × nc, where nc is the number of components into which the data are to be decomposed; and $\mathbf{S}$ has dimensions nc × nf. Because of the very large number (≥nc (nf + 1)) of independent parameters required in such a fitting, a brute force least-squares fit is rarely practical and indirect methods are required. The most commonly used multivariate methods for DOSY data processing are DECRA (direct exponential curve resolution algorithm[16,17]), CORE (component resolved NMR[18,19]), and MCR (multivariate curve resolution[20]). In these methods, the results are normally displayed in the form of fitted component spectra, sometimes along with the fitted decay and often with the diffusion coefficients and (relative) concentrations.

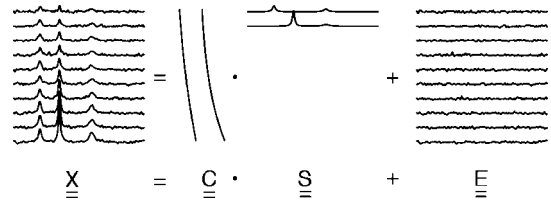

Figure 1. Schematic illustration of the decomposition of an experimental data set **X** into component spectra **S** and decays **C**.

Table 1. Fitting Times for Multivariate Analysis of DOSY Data Using Different Algorithms

	algorithm		
decay function	CORE[18,19]	DECRA	SCORE
pure exponential eq 1	72 min	<1 s	13 s
NUG corrected eq 2	not available[a]	not available[b]	16 s

[a] Algorithm could readily be adapted for NUG correction. [b] Partial NUG correction is possible.[15]

All three methods require a certain amount of user input; in particular, in each case the user has to specify the number of different components with which the algorithm is try to fit the experimental data. DECRA decomposition, in which the original data set is split up in two (theoretically) proportional sets, is based on the generalized rank annihilation method. This allows for an analytical solution, which makes it extremely fast (see Table 1) compared with the other methods. The processing is relatively straightforward and can be easily carried out using MATLAB scripts published by the originators;[16,17] it is also available in the processing software from the major spectrometer companies. The only user input needed is the number of components to be fitted. However, there are two major restrictions, as a consequence of the need to divide the original data set into two proportional halves. First, the experimental data must show pure exponential signal decay for each component; and second, the data must be sampled in equal steps of gradient squared. Unfortunately, real experiments do not show perfect exponential behavior. In practice,

(12) Damberg, P.; Jarvet, J.; Gräslund, A. *J. Magn. Reson.* **2001**, *148*, 343–348.
(13) Nilsson, M.; Morris, G. A. *Magn. Reson. Chem.* **2006**, *44*, 655–660.
(14) Nilsson, M.; Morris, G. A. *Chem. Commun.* **2007**, *933*, 935.
(15) Nilsson, M.; Morris, G. A. *Magn. Reson. Chem.* **2007**, *45*, 656–660.
(16) Antalek, B. *Concepts Magn. Reson.* **2002**, *14*, 225–258.
(17) Windig, W.; Antalek, B. *Chemom. Intell. Lab. Syst.* **1997**, *37*, 241–254.
(18) Stilbs, P.; Paulsen, K. *Rev. Sci. Instrum.* **1996**, *67*, 4380–4386.
(19) Stilbs, P.; Paulsen, K.; Griffiths, P. C. *J. Phys. Chem.* **1996**, *100*, 8180–8189.
(20) Van Gorkom, L. C. M.; Hancewicz, T. M. *J. Magn. Reson.* **1998**, *130*, 125–130.

DECRA decomposition is therefore always somewhat flawed, the most obvious manifestation being crosstalk between the fitted spectra. The principal source of nonexponential signal decay is, as noted above, nonuniformity of the pulsed field gradients (often loosely referred to as gradient nonlinearity). In order to minimize the effects of imperfect gradients, data can be recorded only from the middle of the active region[16] of the sample, but unfortunately such slice selection[21] comes at a severe cost in signal-to-noise ratio. Where the range of diffusion coefficients of interest is relatively narrow, a more effective strategy is to tailor the choice of gradient strengths used to give equal or nearly equal steps in signal attenuation.[15]

In CORE, just as in DECRA, assumptions are made about the forms of the decay functions $C^i(g)$, but in the case of CORE the latter are not necessarily restricted to exponentials[13,18,19] but may be other functions such as the NUG compensated function[13] of eq 2, or functions suitable for modeling signal attenuation in polydisperse samples.[22] The algorithm uses an outer optimization loop for the decay parameters governing the functions $C^i(g)$ and an inner loop for the spectral amplitudes $S^i(f)$. Calculation proceeds from an initial guess as to the parameters (e.g., diffusion coefficients D^i) governing the functions $C^i(g)$, which define a set of M decay curves. The signal decay as a function of g at each frequency point in succession of the experimental data set $X(f,g)$ is then fitted using the set of decay curves to yield an approximate set of M spectra $S^i(f)$, and the sum of the squares of the differences between the calculated data spectral decay $I(f,g)$ and the experimental data $X(f,g)$ is calculated. The inner and outer optimization loops are then alternated, varying the decay parameters until no further reduction in the sum of squares is obtainable. In matrix terms this corresponds to finding the matrices **C** and **S** in eq 5, respectively the decay functions and the spectra for the M different components, such that the sum of the squares of **E** (the square of the Euclidean norm $||\mathbf{X} - \mathbf{CS}^T||$) is a minimum. A notably clear analysis of the problem of fitting experimental NMR diffusion data, and of the case for algorithms of the CORE class, is given in ref 18.

The original CORE[18,19] algorithm was written in Fortran, uses the STEPIT algorithm[23] in both optimization loops, and can be downloaded from the originator's homepage.[24] The great strengths of CORE are that it uses prior knowledge (or more commonly prior hypothesis) parsimoniously but effectively and can be adapted to any predetermined decay function. For the analysis of diffusion NMR data this means that compensation for nonuniform pulsed field gradients and/or decay functions appropriate to polydisperse samples can easily be incorporated in the kernel of the algorithm. In addition, physically sensible constraints, such as nonnegative spectra, can readily be accommodated by the algorithm. The principal drawback of the current implementation of CORE, and one possible reason for its unjustified comparative neglect, is that fitting times for realistic data set sizes can be prohibitive (see Table 1). Some improvement can be made by excluding from analysis spectral regions that mainly contain noise, but even on a modern PC processing times can be many hours.

In contrast to CORE and DECRA, MCR is a so-called soft modeling method in which no assumption is made about the form of the signal decay. A general MCR package for MATLAB is available for download,[25,26] as well as one more specifically tailored to DOSY.[27,28] Many different algorithms under the umbrella name of MCR are in current use;[29] MCR-ALS (alternating least squares) is the method currently most used in DOSY.[20,27,30–33] This algorithm takes a set of either spectra ($\mathbf{S}^-$) or decays ($\mathbf{C}^-$) as initial guesses and estimates a new $\mathbf{C}^+$ or $\mathbf{S}^+$ in an ALS loop, iterated until convergence and employing constraints (e.g., nonnegativity) as appropriate, such that

$$\mathbf{C}^+ = \mathbf{XS}^-(\mathbf{S}^{-T}\mathbf{S}^-)^{-1} \tag{6}$$

or

$$\mathbf{S}^+ = \mathbf{X}^T\mathbf{C}^-(\mathbf{C}^{-T}\mathbf{C}^-)^{-1} \tag{7}$$

As no assumption is made about the decay shapes **C**, MCR should in principle be able to accommodate the effects of NUG automatically. However, in practice it is frustratingly hard to resolve **S** and **C** such that they correspond to the real component spectra and decays. This difficulty arises because in the bilinear model of eq 4 any linear combination of **C** and **S** can give an equally good fit to **X**, a problem often referred to as rotational ambiguity, and therefore the fitted components rarely represent the true spectra. The application of (physically sensible) constraints in the ALS loop is an attempt to break this ambiguity. Arguably the best results obtained with MCR have been in combination with hard modeling, for example, where the elements of **C** are limited to an exponential decay in the ALS step[31] (an approach that could easily be extended to use a NUG decay shape such as eq 2). MCR is often reasonably quick and is potentially very powerful for DOSY data decomposition, but the user is often left struggling to find the close starting guesses that are vital for a good result. A variety of methods is available for finding such starting values, but most are difficult to use for the novice; examples used for DOSY data include PCA-Varimax,[20] IPCA,[33] and OPA.[31]

In this investigation we note that the inner optimization loop of CORE is a linear, not a nonlinear, problem. Changing CORE to use a linear algorithm for the inner optimization results in a speed improvement of several orders of magnitude, improves the accuracy of the final decomposition, and reduces the sensitivity to starting guesses. In view of the substantial reduction in processing time, we refer to this modification of the CORE approach as Speedy COmponent REsolution, or SCORE. Writing

(21) Tillett, M. L.; Lian, L. Y.; Norwood, T. J. *J. Magn. Reson.* **1998**, *133*, 379–384.
(22) Ostlund, J. A.; Nyden, M.; Stilbs, P. *Energy Fuels* **2004**, *18*, 531–538.
(23) Chandler, J. P. *Quantum Chemistry Program Exchange* **1976**, *11*, 307.
(24) Stilbs, P. http://gamma.physchem.kth.se/~peter/.
(25) Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 101–110.
(26) http://www.ub.es/gesq/mcr/mcr.htm.
(27) Huo, R.; Geurts, C.; Brands, J.; Wehrens, R.; Buydens, L. M. C. *Magn. Reson. Chem.* **2006**, *44*, 110–117.
(28) http://www.cac.science.ru.nl/research/software/mulvado/lic.html.
(29) de Juan, A.; Tauler, R. *Crit. Rev. Anal. Chem.* **2006**, *36*, 163–176.
(30) Huo, R.; van de Molengraaf, R. A.; Pikkemaat, J. A.; Wehrens, R.; Buydens, L. M. C. *J. Magn. Reson.* **2005**, *172*, 346–358.
(31) Huo, R.; Wehrens, R.; Buydens, L. M. C. *J. Magn. Reson.* **2004**, *169*, 257–269.
(32) Huo, R.; Wehrens, R.; Buydens, L. M. C. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 9–19.
(33) Huo, R.; Wehrens, R.; van Duynhoven, J.; Buydens, L. M. C. *Anal. Chim. Acta* **2003**, *490*, 231–251.

the inner loop optimization step in matrix form, for an iteration step with a given decay matrix $\mathbf{C}^-$, an improved estimate of the spectrum matrix $\mathbf{S}^+$ is given by

$$\mathbf{S}^+ = \mathbf{X}/\mathbf{C}^- \quad (8)$$

Because the matrix division represents an overdetermined problem, $\mathbf{S}^+$ is identical to the result of a least-squares fitting of $\mathbf{C}^-$ to $\mathbf{X}$ (e.g., by singular value decomposition). The parallel between eqs 7 and 8 brings out the relationship between CORE (or SCORE) and MCR; CORE may be regarded as a hard modeling version of MCR in which starting guesses are determined by our prior knowledge of the data. Normally, no further constraints are needed, but it may sometimes be useful to use nonnegativity, at the expense of a considerable reduction in speed. In the original CORE implementation, initial guesses may be specified for the decay parameters; here, the diffusion coefficients yielded by DECRA were used as starting points for both CORE and SCORE.

EXPERIMENTAL SECTION

Experimental measurements were carried out nonspinning on a 400 MHz Varian Inova spectrometer, using a 5 mm diameter indirect detection probe equipped with a *z*-gradient coil allowing gradient pulses of up to 30 G cm^{-1}. A sample containing camphene (20 mM), geraniol (23 mM), and quinine (19 mM), dissolved in methanol-d_4 with TMS as a chemical shift reference, was used in all experiments. No sample temperature control was used; experiments were carried out at the probe quiescent temperature in a room air conditioned at about 20 °C. Nominal gradient strengths were calibrated as recommended by the manufacturer.

For the acquisition of DOSY data, the Oneshot[34] pulse sequence was used with a diffusion delay Δ of 0.2 s and a net diffusion-encoding pulse width δ of 2 ms (i.e., each bipolar pulse pair consisted of two 1 ms gradient pulses) with 30 gradient strengths ranging from 3.0 to 27.3 G cm^{-1} chosen to give equal steps in nominal gradient squared. A spectral width of 6000 Hz was used, and 16 384 complex data points were acquired for each gradient strength. ^{1}H reference spectra were acquired using 16 transients at a spectral width of 6000 Hz. The DOSY data were corrected for instrumental and experimental imperfections by reference deconvolution[35,36] with the standard instrument software (Vnmr6.1C) using the TSP peak as reference and a target line shape of a pure Lorentzian of 3 Hz width at half-height.

With the exception of the original CORE processing,[18,19] which used the CORE 3.7 software downloaded from http://gamma.physchem.kth.se/~peter/, all other processing (including Fourier transformation, phasing, apodization, baseline correction, DOSY, DECRA, and SCORE) was carried out in MATLAB, using code written in-house and available from the authors on request. All spectra were baseline-corrected before further analysis. The MATLAB code for DECRA was based on the script published by Antalek,[16] as previously described.[15] The SCORE algorithm was implemented in MATLAB using a simplex algorithm for the outer loop and the direct matrix division of eq 8 for the inner loop. The power series coefficients for correction of the effects of nonuniform field gradients using eq 2 were obtained as described previously.[6,13] All calculations were performed on a standard PC with a single-core 3 GHz CPU and 3 Gb of RAM, using MATLAB 7.4 (2007a) and the latest version (3.7) of CORE. Multivariate analyses assumed four components ($M = 4$); the low concentration of residual water in the sample ensures that the HDO/CD_3OH and CD_2HOD resonances show similar diffusion coefficients and may be treated as a single component.

The authors are in the process of assembling all of the above (and more) processing tools for DOSY data into a user-friendly graphical user interface, currently undergoing alpha testing, to be made freely available in due course.

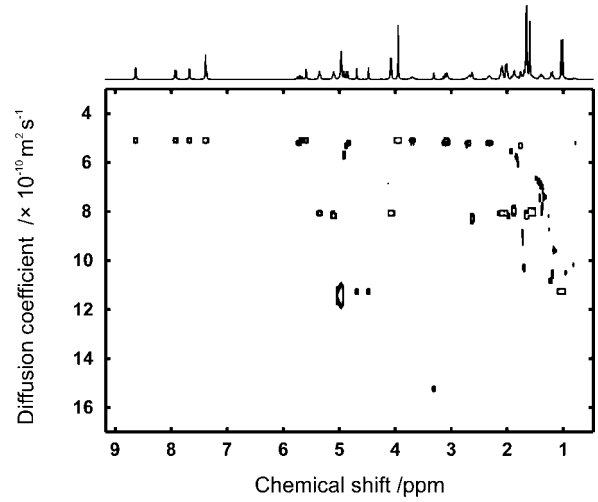

Figure 2. ^{1}H DOSY spectrum of a solution of geraniol, camphene and quinine in methanol-d_4 with TMS as a chemical shift reference.

RESULTS AND DISCUSSION

The complications associated with spectral overlap are evident in the standard DOSY spectrum in Figure 2, where the signals around 1−2 ppm shows compromise values for the diffusion coefficients, as well as at 4.8 ppm where overlap with the HOD signal causes slowly diffusing species to appear at higher apparent diffusion coefficients. The advantages of using multivariate methods for these data are clear from Figure 3, where the component spectra have been resolved with DECRA (d), SCORE (e), and CORE (a); the first two methods extract components that are in very good agreement with the reference spectra (right), but the components extracted by CORE are significantly less good approximations to the true component spectra. The SCORE components (10 overlaid fits from different starting points, vide infra) reproduce the true spectra almost perfectly, while the components resolved by DECRA show a significant amount of cross-talk, most clearly seen at 7−9 ppm in the "geraniol" component. This cross-talk arises because the nonuniformity of the pulsed field gradients makes the diffusional decay of the signals deviate from exponentiality. Compensating for this is straightforward in SCORE and CORE[13]

(34) Pelta, M. D.; Morris, G. A.; Stchedroff, M. J.; Hammond, S. J. *Magn. Reson. Chem.* **2002**, *40*, S147−S152.
(35) Morris, G. A. In *Encyclopedia of Nuclear Magnetic Resonance, Advances in NMR*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons, Ltd.: Chichester, U.K., 2002; Vol. 9, pp 125−131.
(36) Morris, G. A.; Barjat, H.; Horne, T. J. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 197–257.

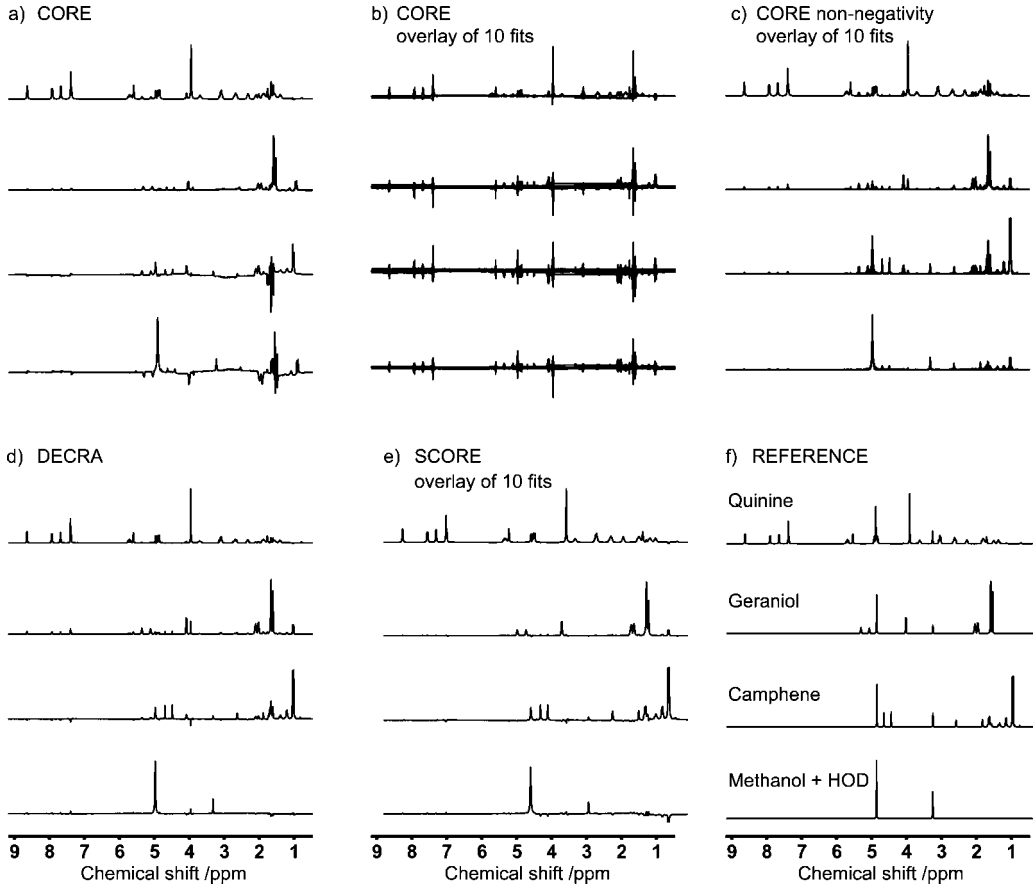

Figure 3. Component spectra extracted from the ^{1}H DOSY data of Figure 2 using DECRA, CORE, and SCORE, together with reference spectra from individual solutions of geraniol, camphene, and quinine in methanol-d_4. In b, c, and e, overlays of the component spectra obtained using 10 different random sets of starting values are shown.

(and potentially in MCR), but only limited compensation is possible in DECRA.[15] The standard CORE implementation used here does not correct for nonuniform gradients, so the components extracted using it would be expected to show similar cross-talk to that seen in the components derived using DECRA, but this effect is masked by the poorer quality of the fitted components. The set of four components extracted with CORE that is shown in Figure 3a was obtained using the diffusion coefficients estimated by DECRA as starting points and is representative of the best results obtained with this data set using CORE; as noted below, starting with less good guesses for the component diffusion coefficients typically gives poorer results. Table 2 summarizes the diffusion coefficients estimated by the different algorithms.

Table 2. Diffusion Coefficients Obtained Using the Different Algorithms[a]

	algorithm			
component	CORE	SCORE	DECRA	HRDOSY
quinine	4.77	5.06	4.16	5.12 ± 0.0022
geraniol	8.02	8.19	7.86	8.10 ± 0.0036
camphene	11.0	11.8	12.4	11.2 ± 0.0067
methanol/HOD	16.5	15.9	17	15.3 ± 0.011

[a] The values for CORE and DECRA were obtained without, and those for SCORE with, compensation for nonuniform gradients. The data from HRDOSY used NUG correction and were obtained from one well-resolved signal per component: quinine (8.63 ppm), geraniol (4.08 ppm), camphene (1.02 ppm), and methanol/HOD (3.30 ppm). The estimated errors quoted are twice the standard errors obtained from the relevant fitting algorithm, but do not take into account the blanket uncertainty, estimated at ±2%, deriving from temperature variation and calibration errors.

For all of the multivariate methods used here it is important for a successful fit to have very good quality data (i.e., minimal systematic errors and a high signal-to-noise ratio). To assess the effect of noise on the spectral decomposition, SCORE fitting of the experimental data was repeated with different amounts of synthetic pseudorandom noise added. The original data were corrected for systematic variation in signal phase, frequency, and line shape using reference deconvolution[35,36] and had a signal-to-noise ratio for the highest peak of 130000:1. For signal-to-noise ratios above 10000:1, all the spectral features of the four components extracted were effectively independent of signal-to-noise ratio; below this figure the results degraded progressively until at 1000:1 the component spectra were too poor to be of use, containing less than half of the true signals. In this investigation the relative amplitudes of the spectra are broadly similar, but it worthwhile to note that the difficulty of unraveling mixed data sets such as these increases with the dynamic range spanned by the component amplitudes.

To compare processing speed, the same data were fitted with CORE, SCORE, and DECRA (Table 1). The total number of spectral points was 9514, and all 30 gradient levels were used. DECRA and SCORE converged to component spectra close to the reference spectra, but CORE 3.7 gave component spectra which showed significant cross-talk and artifacts. The latter

problems reflect the difficulty of performing a multivariate analysis for four components on DOSY data spanning only a factor of 3 in diffusion coefficient (5.1, 8.2, 11.8, and 15.9 × 10^{-10} m^2 s^{-1}, respectively, for quinine, geraniol, camphene, and methanol). For the data used, SCORE is more than 300 times faster than CORE, making component resolution practical for complete spectra.

To investigate the relative sensitivity to starting values of the CORE and SCORE algorithms, a set of 10 different starting points was used, each consisting of four diffusion coefficients randomly distributed between 1 and 25 × 10^{-10} m^2 s^{-1}; CORE was run both with and without nonnegativity constraints. The results can be seen in Figure 3, where the 10 resultant sets of component spectra are overlaid for CORE (b), nonnegative CORE (c), and SCORE (e). The fit-to-fit variation with SCORE is essentially negligible, but both types of CORE fitting show severe sensitivity to the initial diffusion coefficients; as expected, imposing the constraint of nonnegative signals reduces the sensitivity to starting guesses, but the signal amplitudes remain highly variable. A quantitative comparison was performed by calculating the root-mean-square deviation r from the average fit, relative to the root-mean-square average signal, in each case

$$r = \sqrt{\frac{\sum_{k=1}^{m}\sum_{j=1}^{p}\sum_{i=1}^{n}(X_i - \overline{X})^2}{m\sum_{j=1}^{p}\sum_{i=1}^{n}\overline{X}^2}} \qquad (9)$$

for m fits, p components, and n data points. The r values for the SCORE, CORE, and nonnegativity constrained CORE were 1.9 × 10^{-7}, 2.8, and 0.19, respectively, i.e., for this data set the results of the SCORE algorithm are more than 6 orders of magnitude less sensitive to starting guesses than those of the two variants of CORE.

CONCLUSIONS

Multivariate methods can be very useful for extracting component spectra from DOSY data. SCORE is a minor variation on the established CORE method that gives a major improvement in performance, speeding it up by almost 3 orders of magnitude and reducing its sensitivity to starting guesses while retaining its many advantages. Parallelizing the inner optimization loop would further improve speed, as noted in reference 18.

ACKNOWLEDGMENT

Support from the Engineering and Physical Sciences Research Council (Grant references EP/E057888/1 and EP/E058899X/1) is gratefully acknowledged.

Received for review December 20, 2007. Accepted March 17, 2008.

AC7025833