Journal of Magnetic Resonance 207 (2010) 332–336

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

RECORD processing – A robust pathway to component-resolved HR-PGSE NMR diffusometry

Peter Stilbs *

Physical Chemistry and Industrial NMR Centre, Royal Institute of Technology (KTH), Teknikringen 36, SE-100 44 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 13 August 2010
Revised 24 September 2010
Available online 7 October 2010

Keywords:
Self-diffusion
PGSE
CORE analysis
DOSY
Diffusometry

ABSTRACT

It is demonstrated that very robust spectral component separation can be achieved through global least-squares CORE data analysis of automatically or manually selected spectral regions in complex NMR spectra in a high-resolution situation. This procedure (acronym RECORD) only takes a few seconds and quite significantly improves the effective signal/noise of the experiment as compared to individual frequency channel fitting, like in the generic HR-DOSY approach or when using basic peak height or integral fitting. Results from RECORD processing can be further used as starting value estimates for subsequent CORE analysis of spectral data with higher degree of spectral overlap.

1. Introduction

Spectral component separation in NMR, as based on PGSE data [1] has in later decades become quite popular in wide application fields. The typical purpose is either or both of (a) to maximize the quality and precision of component self-diffusion coefficients, to be further used for physico-chemical considerations of various kinds – or (b) to achieve analytical-type detection or separation of the component spectra from an unknown mixture in solution.

A number of varying strategies in this area have later been suggested [2–16]. The central problem is a commonly recurring one in chemistry and physics – to separate multi-exponential decay curves. This task is well-known to be notoriously difficult even for just doubly exponential experimental data at already marginal levels of experimental statistical noise (see e.g. [17]).

For PGSE NMR experiments on multi-component systems the situation is normally brighter, owing to the initial spectral separation through standard Fourier transformation of the raw data. For fully separated peaks, the individual signal decay amplitudes should become mono-exponential at constant rf pulse spacing, ideally obeying the simplified Stejskal–Tanner Equation [18] (1) or variants thereof. On high field spectrometers signal overlap could be marginal or even absent in practice, and data evaluation may deceptively appear easy and straightforward.

$$A = A_0 \cdot \exp[-(D\gamma^2 g^2 \delta^2 (\Delta - \delta/3))] \quad (1)$$

For improved analysis in an overlapping situation and better signal/noise, the important point is to realize that ideally the spectral bandshapes separated by initial Fourier transformation remain constant in the 2nd time/gradient parameter domain of the experiment. Only their amplitudes should vary, and should ideally obey the S–T relation (1) or some appropriate variant thereof. This is the prior knowledge-type information which constitutes the constraining basis for the so-called CORE global least-squares strategy for separating PGSE component data from the 2nd time/gradient parameter domain into the diffusion dimension [6,7].

For intrinsic numerical and experimental non-ideality reasons listed below, brute-force CORE processing of multi-component PGSE data to achieve objectives (a) and (b) above are likely to fail to some extent for three or more components, however. A natural countermeasure would then be to manually fit parts of the spectral information separately to single or double exponential functions, for further full spectral reassembly at a later stage.

An existing generic approach named "High-Resolution DOSY" [4,19] is based on a strategy of related kind, i.e. typically a 8–32 k data set is analyzed frequency for single- or double component PGSE spectral contributions. Normally, such "H-R DOSY" is applied without any stabilizing interrelation between nearby spectral regions, like in the CORE approach. In its original implementation [4] peak maximum fitting of (filtered) spectral data is followed by DOSY spectrum construction based on peak segmentation. Also applied were reference deconvolution in the frequency dimension, and corrections for non-ideal Stejskal–Tanner type dependence (Eq. (1)) on experimentally applied PGSE pulse parameters. Some other variants are found in the literature or in software implementations by the leading NMR instrument vendors.

* Fax: +46 87908207.
E-mail address: peter@physchem.kth.se

1090-7807/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2010.09.019

Merging CORE-type processing with the HR-DOSY-type spectral analysis through "sequential region evaluation" introduced here (acronym named RECORD-REgion COmponent Resolved Diffusion) is straightforward in a high-resolution situation. One simply considers each spectral peak (typically 10–100 frequency channels wide, depending on the number of data points and spectral range) as a common data entity. Region selection is made through some kind of more or less sophisticated algorithm.

It turns out that a fitting procedure of this kind does becomes very robust and reliable in high-resolution PGSE applications. Like in standard CORE processing one also effectively increases the Signal/Noise of the experiment by a great deal, through full use of interconnected spectral information. Of course, for less-resolved spectral situations, the RECORD approach is less useful or inapplicable, and one has to revert to the original type of CORE processing. A combination of both, i.e. a final CORE one based on initial results from a RECORD pre-processing seems the most effective approach in such a situation.

One should bear in mind that many additional factors may prevent the accurate experimental evaluation or complete spectral disentanglement in all PGSE applications. Reasons for systematic experimental deviation from the S–T relation include non-constant magnetic field gradients or rf fields across the sample, radiation damping effects, magnetic field gradient or rf pulse imperfections, convective flow, spectral drift, poor signal/noise, baseline distortions and eddy current effects.

2. Experimental-simulations

Comments on stability of various fitting procedures [14,16], including the original CORE strategy [6,7] have been made in the literature in recent years. However, various implementations of the CORE into computer code could behave differently with regard to stability and speed, for a number of reasons. The SCORE pathway modification [16] of the CORE fitting procedure has numerical difference in detail, although the general fitting strategy remains the

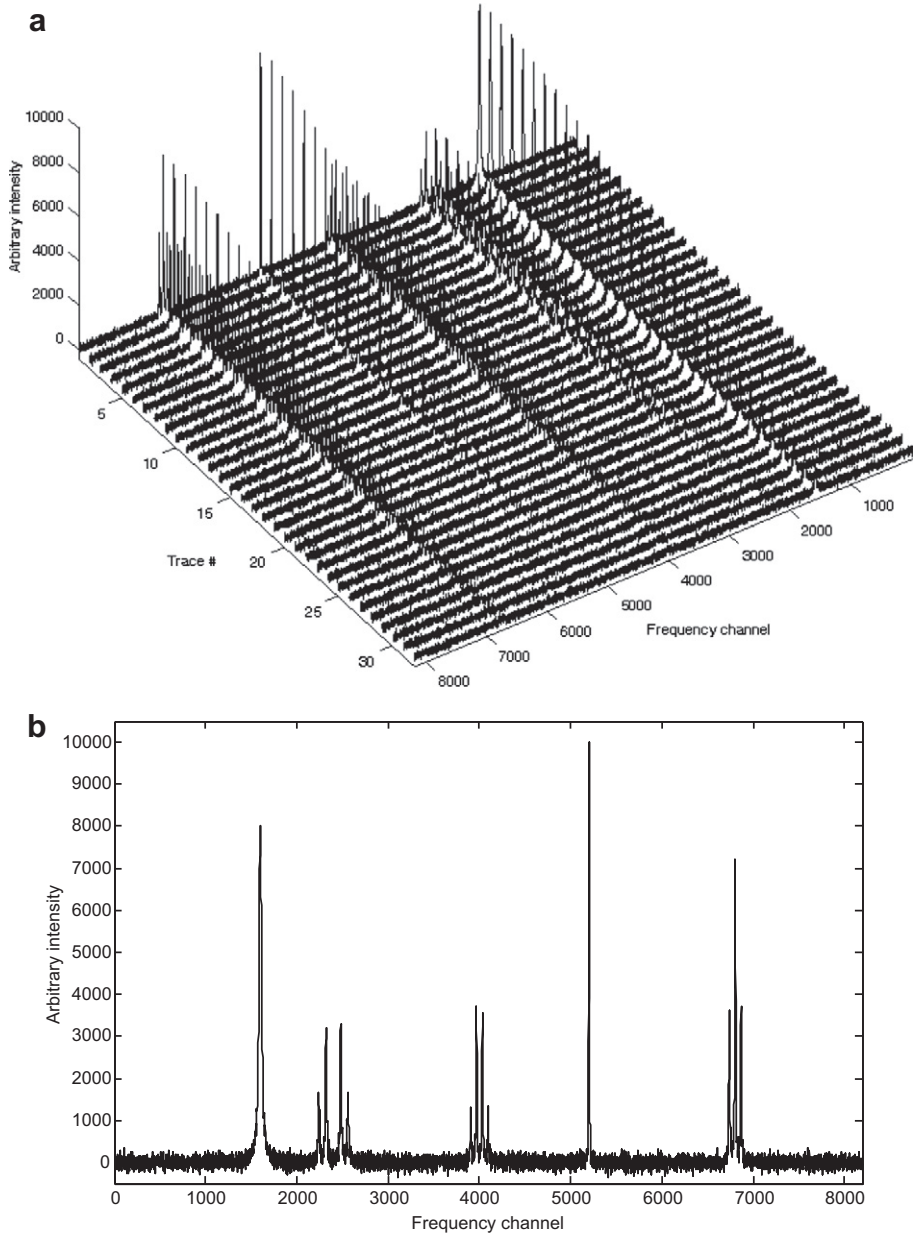

Fig. 1. Overviews of the synthetic data set subjected to analysis in the present paper. The 1D spectrum is the first data trace in the 2D data set.

same in principle. Also, at least two independent Matlab®-based CORE implementations exist [20,21], that do use different minimization algorithms than the original STEPIT one. The actual number of significant digits used in the various computer codes likely matters as well. The original CORE version and its later versions is coded in Double Precision (Real * 8) FORTRAN, which typically uses 15 significant digits in its internal arithmetic.

These issues necessitated some additional comparative data evaluations, with outcomes below that actually contrast earlier conclusions and statements [14,16] about CORE-type processing, as described below.

To allow unbiased and spectrometer-independent numerical examination of these matters the "spectral data" in the present paper are entirely synthetic and in addition only contain random white noise.

The current version of CORE (CORE4) is switchable between the original algorithm and the direct matrix division SCORE-type modification [16] of the inner loop in the CORE. High-quality mathematical library routines are used for the SCORE matrix operations. CORE4 includes the RECORD variant introduced in the present paper, as well as additional options.

CORE4 and supplementary routines are available as a download in PC versions from the Author's web site. The simulations in the present paper were made on a single-core 2.66 GHz DELL Dimension 8250 with 2 GB memory, using the Windows XP Service pack 3 operating system. Note that CORE4 is a standalone console-type program and that it does not have a Windows-type graphical user interface (GUI). A separate Matlab® installation is essential for spectral pre-processing and graphical results display.

3. Spectral data set

Through a separate FORTRAN program, a series of 32 8 k proton PGSE data was simulated, using the Stejskal–Tanner Equation (1). The following parameters were used: magnetic field gradient pulse separation (Δ) 60 ms, magnetic field gradient pulse duration (δ) 2 ms and magnetic field gradient strength (g) 0.0–3.41 T m^{-1}, in 32 equidistant steps. These gradient settings were applied to a composite Lorentzian bandshape spectrum of a singlet (s1), an AB-type spectrum (AB), an "ethyl group" (ethyl), and again a second singlet (s2). The corresponding singlet/multiplet centre frequencies in "channel units"/self-diffusion coefficients (in units of 10^{-11} m^2 s^{-1}) were set to be 1600/1.0 (s1), 2400/2.0 (AB), 4000/1.5 and 6800/1.5 (ethyl) and 5200/2.5(s2), respectively.

The bandwidths of the peaks/multiplets were made to vary by a factor of 5, corresponding to "T_2:s, in inverse frequency channel units", of 0.01(s1), 0.02(AB), 0.03(ethyl) and 0.05(s2), and relative amplitudes correspondingly adjusted to create a reasonable spectral intensity balance.

Finally, Gaussian noise was added to provide the resulting data set illustrated in Fig. 1. Frequency channels where the data intensity was less than 8% of that of the biggest peak in the spectrum were subsequently ignored in the all of the data analyses described below, leaving 475 frequency channels out of the original 8192 for actual analysis. Depending on signal amplitude and bandwidth, the individual peaks effectively were then defined through 10–100 frequency channels.

As seen, the signal/noise for the weakest signals in the data set at zero echo attenuation is only of the order of three, but a reliable

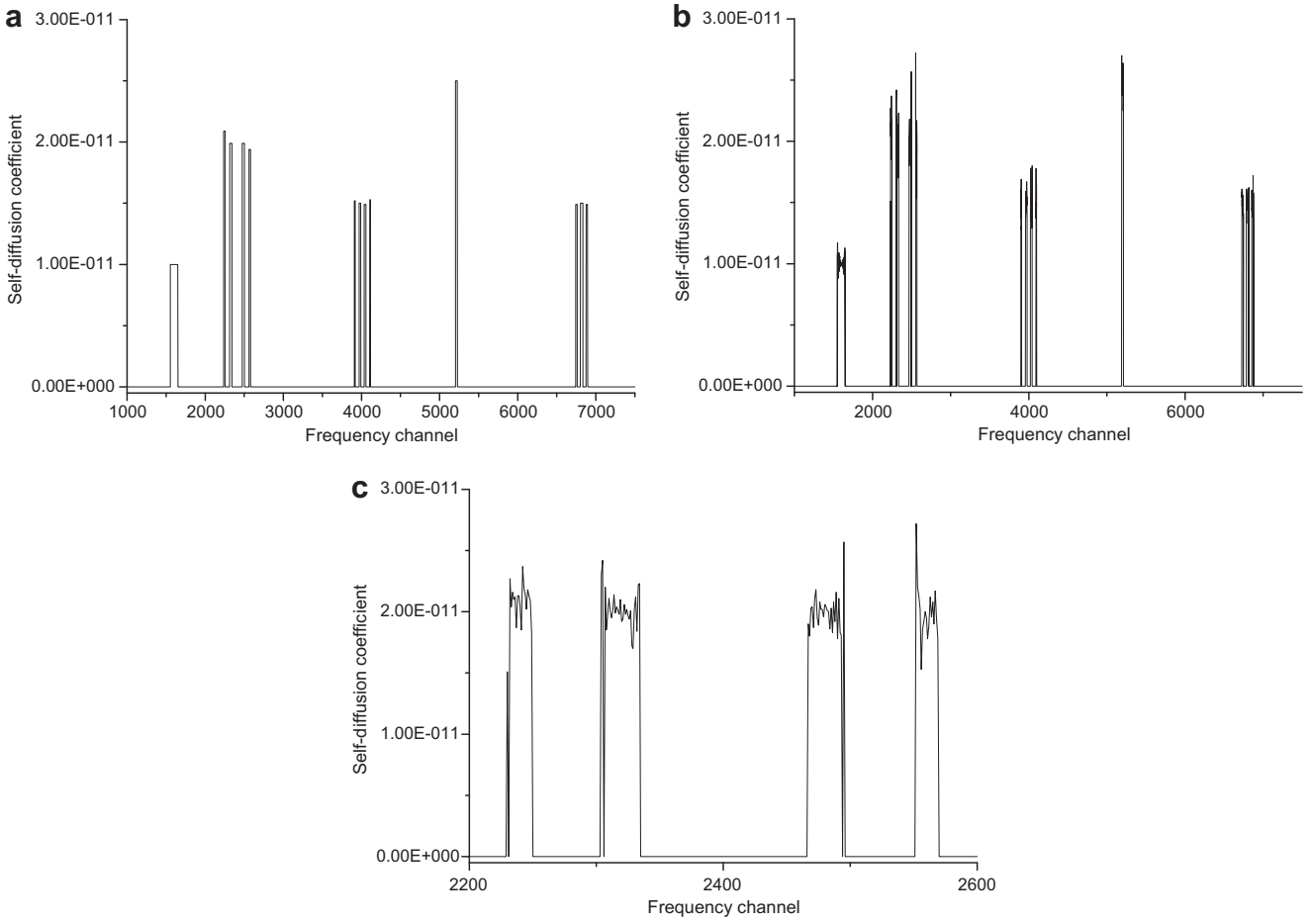

Fig. 2. 1D representations of results from (a) RECORD processing (b) Individual frequency channel "HR-DOSY" processing (c) detail of (b).

analysis can still be made through the CORE and RECORD approaches, as described in the following.

4. RECORD processing and analysis

Fig. 2a illustrates the outcome of RECORD processing of the data set illustrated in Fig. 1 (with spectral low-intensity areas ignored, as described in the previous subsection). Iteration starting values were arbitrarily set way off (1.0×10^{-9} m^2 s^{-1}) and reset to that value for each frequency channel. Total simulation and disk write time was like 5s.

Clearly, four rather distinct self-diffusion coefficients are discernible, with intensity weighted average values corresponding to 1.00(s1), 2.00(AB), 1.50 (ethyl) and 2.50(s2), using three digit numeric result precision. Despite the considerable spectral noise, the RECORD analysis results are seen to be virtually identical to the actual underlying data.

For comparison, a similar "HR-DOSY" analysis is shown in Fig. 2b, with expanded detail in the "AB-region" in Fig. 2c. Note the considerable fitting noise (individually 10–20% of actual data) throughout. It is naturally less in spectral areas with high initial intensity and higher in the spectral wings. In essence, the varying results in the centre of the respective regions approach what would result from routine "peak height fitting" using standard spectrometer, or off-line external software.

DOSY-type displays of the RECORD processing and "HR-DOSY" results above are seen in Fig. 3a and b. The differing relative stability of the two approaches becomes even more obvious in such a graphical form.

At this point some comments on the concept of "DOSY" are justified, also considering the deep and widespread confusion regarding this matter in current literature. DOSY is not a "method". It is rather an accepted acronym for combined multi-frequency data displays of results from PGSE or variants thereof, and is a statistical construct. DOSY displays do not add anything, apart from a visual results overview – that may at times be useful. Original DOSY implementations, where interpolation between inverse Laplace transformations for component separation for individual spectral frequencies is made, even degrade the process of optimal data evaluation severely.

The link between the data in Fig. 2a and b and the 2D-visualization in Fig. 3a and b is nothing else than a simple numerical transformation, combined with a user-added arbitrary or estimated error-limit mimicking bandshape (normally Gaussian) in the diffusion dimension. Apart from this, there is a 1:1 correspondence between the data in Fig. 2a and b and those in 3a and b, respectively. Such data transformation requires less than 10 lines of computer code (see e.g. the Manual supplied with the CORE4 distribution).

5. CORE processing and analysis

The same data were then subjected to a full CORE analysis for four components, either by the original double least-squares procedure and through the SCORE modified one.

The same data selection as in the RECORD and "HR-DOSY" processing was applied, i.e. 475 out of the 8192 frequencies of the data set in Fig. 1 were actually considered. Two comparative runs were made – one with self-diffusion coefficient starting estimates that were approximately in the right region (i.e. 5.0, 2.0, 6.0 and 1.5×10^{-11} m^2 s^{-1}), and one with starting values from the RECORD analysis above (i.e. matching the correct values already). The results are summarized in Table 1.

It may appear surprising at first sight that the fits in the Table are nowhere near the quality of the RECORD or HR-DOSY runs above. There is a considerable difference numerically between one-exponential and freely interrelated multi-exponential optimizations, however.

The approximate 80% statistical confidence interval estimates of the component D-values were also calculated, using existing Monte-Carlo routines in CORE. As expected, they are much larger than those from RECORD processing. Typically, the results were of the order of 5% of the corresponding component's D-value.

More importantly, there was also much statistical covariance in between extracted components in a full 4-exponential CORE analysis. That is particularly evident in DOSY-type displays (not shown here) corresponding e.g. to the data analysis in row one of Table 1. Although there were dominant contributions at the true combinations of chemical shift/frequency channel and D-value, crosstalk between various component contributions was like 10–30%. That makes signal component attribution attempts via DOSY-type results displays virtually meaningless. In the RECORD-type analysis, or in a manual isolated region CORE analysis using single exponentials, the situation is obviously very different indeed, as seen in Fig. 3a.

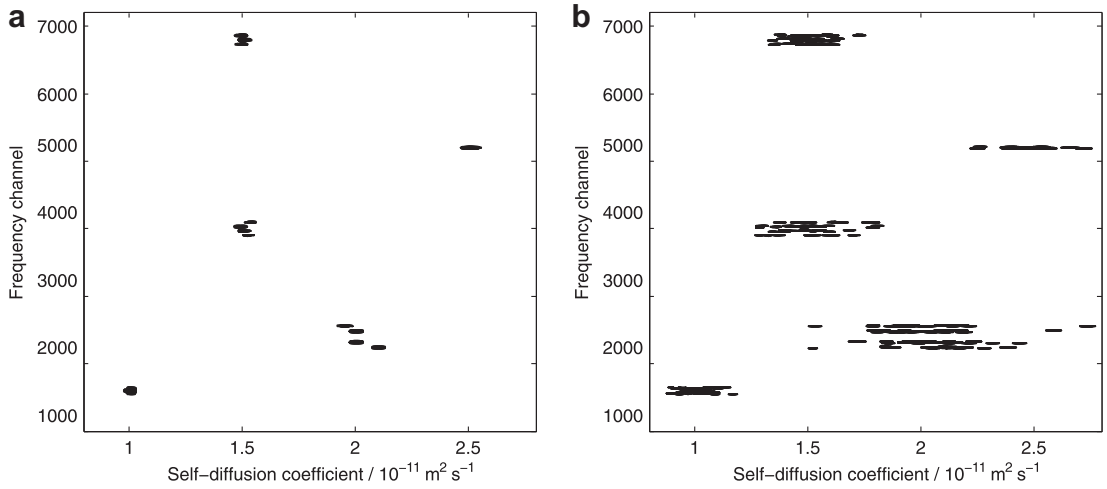

Fig. 3. 2D representations ("DOSY-type PGSE results displays", albeit with interchanged x and y axes, as compared to widespread DOSY conventions) of the data from (a) RECORD processing (b) Individual frequency channel "HR-DOSY" processing. In these maps, an artificial Gaussian bandshape component of 1% width at half height has been introduced in the diffusion dimension. That is more than the actual statistical error distribution of the RECORD results (a) and for clarity considerably less than the wide results spread in (b).

Table 1
List of CORE and SCORE 4-parameter fitting results (D:s are given in 10^{-11} m^2 s^{-1}).

Type	D1	D2	D3	D4	Time/s
CORE[a]	0.95 (1.00)	1.46 (1.50)	1.96 (2.00)	2.71 (2.50)	380
CORE[b]	0.95	1.52	[c]	2.50	500
SCORE[a]	1.04	[c]	[c]	2.56	10
SCORE[b]	1.04	[c]	[c]	2.56	12

[a] Using RECORD-based starting values (almost exact; given in parenthesis in row 1).
[b] Using arbitrary approximate estimates as starting values (see text).
[c] Diverged to high or low *D*-value, with very small fitting amplitude.

As also summarized in Table 1, the SCORE-based algorithm is considerably faster throughout than the original CORE implementation, for which multi-component analyses steeply become quite time-consuming, SCORE-type computing time seems almost independent of the number of components, but at one or even two components they are similar from any practical point of view.

Finally, note that CORE processing of the present data set involves a quite demanding direct four-component analysis of rather noisy data. The underlying fitting complexity here is actually on the borderline of what unguided and unsegmented CORE/SCORE-type data extraction can handle. Contradictory to earlier findings and statements [14,16], the original CORE procedure happens to perform more reliably here than its SCORE-variant, however.

From an algorithm point of view on should note that there is no actual difference between the original CORE one and the SCORE-type modification. As already mentioned in the Introduction, the original CORE analysis uses a numerical brute-force approach in the inner loop least-squares bandshape optimization at a given set of *D*-values from the outer loop least-squares *D*-value optimization. The SCORE-variant replaces the inner loop spectral amplitude least-squares iteration cycle with an equivalent matrix division solution. Although computer-intensive as well, the latter approach is faster, and dramatically so for many spectral components [20]. And for the particularly demanding data set here the differing outcomes of the two types minimizations are likely just coincidal.

6. Conclusions

It is demonstrated that very robust and reliable spectral component separation can be achieved in seconds of computing time through the RECORD procedure in a high-resolution situation, even at low signal/noise.

The RECORD processing approach is evidently not directly applicable to spectral data sets with a higher degree of spectral overlap, but RECORD results and region identifications do serve as rather optimal starting value estimates for some subsequent full CORE analysis. With a more "intelligent" region identification algorithm than the simple-minded one in the present implementation, the direct applicability of the RECORD processing approach is considerably extended above the processing example in the present paper. Versions 4c and later of CORE (September 2010) do have provisions for external interfacing with GUI-interactive Matlab® procedures for this purpose, and for indicating spectral regions that are less well described by mono-modal diffusion (as in the RECORD-type analysis) than others.

Acknowledgments

This work has been supported by the Swedish Research Council (VR) and the Knut and Alice Wallenberg foundation.

References

[1] P. Stilbs, Molecular self-diffusion coefficients in Fourier transform nuclear magnetic resonance spectrometric analysis of complex mixtures, Anal. Chem. 53 (1981) 2135–2137.
[2] K.F. Morris, C.S. Johnson Jr., Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy, J. Am. Chem. Soc. 114 (1992) 3139–3141.
[3] D. Schulze, P. Stilbs, Analysis of multicomponent FT-PGSE experiments by multivariate statistical methods applied to the complete bandshapes, J. Magn. Reson. 105 (1993) 54–58.
[4] H. Barjat, G.A. Morris, S. Smart, A.G. Swanson, S.C.R. Williams, High-resolution diffusion-ordered 2D spectroscopy (HR-DOSY) – a new tool for the analysis of complex mixtures, J. Magn. Reson. Ser. B 108 (1995) 170–172.
[5] B. Antalek, W. Windig, General rank annihilation method applied to a single multicomponent pulsed gradient spin echo NMR data set, J. Am. Chem. Soc. 118 (1996) 10331–10332.
[6] P. Stilbs, K. Paulsen, P.C. Griffiths, Global least-squares analysis of large correlated spectral data sets: application to component-resolved FT-PGSE NMR spectroscopy, J. Phys. Chem. 100 (1996) 8180–8189.
[7] P. Stilbs, K. Paulsen, Global least-squares analysis of large, correlated spectral data sets and application to chemical kinetics and time-resolved fluorescence, Rev. Sci. Instrum. 67 (1996) 4380–4386.
[8] D. Wu, A. Chen, C.S. Johnson Jr., Three-dimensional diffusion-ordered NMR spectroscopy The homonuclear COSY-DOSY experiment, J. Magn. Reson. Ser. A 121 (1996) 88–91.
[9] W. Windig, B. Antalek, Direct exponential curve resolution algorithm (DECRA): a novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles, Chemometr. Intell. Lab. Syst. 37 (1997) 241–254.
[10] C.S. Johnson, Diffusion-ordered nuclear magnetic resonance spectroscopy: principles and applications, Prog. Nucl. Magn. Reson. Spectrosc. 34 (1999) 203–256.
[11] N.M. Loening, J. Keeler, G.A. Morris, One-dimensional DOSY, J. Magn. Reson. 153 (2001) 103–112.
[12] G.A. Morris, Diffusion-ordered spectroscopy (DOSY), in: D.M. Grant, R.K. Harris (Eds.), John Wiley & Sons Ltd., Chichester, UK, 2002. pp. 35–44..
[13] M. Nilsson, A.M. Gil, I. Delagadillo, G.A. Morris, Improving pulse sequences for 3D DOSY: COSY-IDOSY, Chem. Commun. (2005) 1737–1739.
[14] M. Nilsson, G.A. Morris, Correction of systematic errors in CORE processing of DOSY data, Magn. Reson. Chem. 44 (2006) 655–660.
[15] M. Nilsson, G.A. Morris, Pure shift proton DOSY: Diffusion-ordered 1H spectra without multiplet structure, Chem. Commun. (2007) 933–935.
[16] M. Nilsson, G.A. Morris, Speedy component resolution: An improved tool for processing diffusion-ordered spectroscopy data, Anal. Chem. 80 (2008) 3777–3782.
[17] A.A. Istratov, O.F. Vyvenko, Exponential analysis in physical phenomena, Rev. Sci. Instrum. 70 (1999) 1233–1257.
[18] E.O. Stejskal, J.E. Tanner, Spin diffusion measurements. spin echoes in the presence of a time-dependent field gradient, J. Chem. Phys. 42 (1965) 288–292.
[19] M.D. Pelta, H. Barjat, G.A. Morris, A.L. Davis, S.J. Hammond, Pulse sequences for high-resolution diffusion-ordered spectroscopy (HR-DOSY), Magn. Reson. Chem. 36 (1998) 706–714.
[20] N. Hedin, J.B.S. Ng, P. Stilbs, Spectral deconvolution of NMR cross polarization data sets, Nucl. Magn. Reson. 35 (2009) 208–213.
[21] M. Nilsson, The DOSY Toolbox: A new tool for processing PFG NMR diffusion data, J. Magn. Reson. 200 (2009) 296–302.