Electronic Supplementary Material (ESI) for Chemical Communications.
This journal is © The Royal Society of Chemistry 2021

Supplementary Information for "Ultrafast diffusion-based unmixing of ^{1}H NMR spectra"

Rituraj Mishra, Achille Marchand, Corentin Jacquemmoz and Jean-Nicolas Dumez*

Université de Nantes, CNRS, CEISAM UMR6230, 4400 Nantes, France

Contents

1. Derivation of the frequency sweep for quadratic spacing of the effective gradient area

In spatially encoded diffusion NMR, the data is modelled as:

$$S(z) = S_0 \exp(-D\Delta'(K(z))^2) \quad \text{(S1)}$$

where *D* is the diffusion coefficient, Δ' is the effective diffusion delay and is spatially independent due to the same direction of two frequency swept pulses (Fig. 1a), $K(z)$ is the position-dependent effective gradient area multiplied by the gyromagnetic ratio. Spatial encoding of diffusion is achieved with the combined application, to an initially transverse magnetisation, of a 180° frequency swept pulse and a magnetic-field gradient pulse of duration T_e, followed by a gradient pulse of duration T_e This spatial encoding block is shown in Fig. 1 of the main text.

Consider an ensemble of uncoupled nuclear spins *I* = ½, submitted to such spatial encoding block. An expression for *K(z)* can be derived by assuming that spins flip instantaneously at a time $t_{flip}(z)$ when their resonance frequency matches with the offset of the pulse. Under this assumption, the effective precession time is

$$\tau(z) = 2T_e - 2t_{flip}(z) \quad \text{(S2)}$$

which results in

$$K(z) = \gamma G \tau(z) = \gamma G(2T_e - 2t_{flip}(z)). \quad \text{(S3)}$$

Most spatial encoding experiments rely on a linearly-swept pulse, which results in a linear spacing of $t_{flip}(z)$, and $K(z)$, as a function of z. In order to achieve the quadratic spacing required for DECRA processing, the following relations has to be verified:

$$\frac{\partial(\tau(z))^2}{\partial z} = c \quad \text{(S4)}$$

where c is a constant, which gives

$$(\tau(z))^2 = cz + c_0 \quad \text{(S5)}$$

where c_0 is another constant, and

$$\tau(z) = \sqrt{cz + c_0} \quad \text{(S6)}$$

The values of *c* and c_0 can be found by imposing that $\tau\left(-\frac{L}{2}\right) = 0$,and $\tau\left(\frac{L}{2}\right) = 2T_e$, where *L* is the length of the spatial region swept by the pulse. This gives:

$$\tau(z) = 2T_e\sqrt{\frac{z}{L} + \frac{1}{2}} \quad \text{(S7)}$$

Comparing Eqs S2 and S7, the requirement to achieve the desired quadratic spacing is to have a flip time:

$$t_{flip}(z) = T_e\left(1 - \sqrt{\frac{z}{L} + \frac{1}{2}}\right) \quad \text{(S8)}$$

On the other hand, at the time of the flip, the offset of pulse is equal to that of the spins at position *z*:

$$\nu_{rf}\left(t_{flip}(z)\right) = -\frac{\gamma}{2\pi}Gz. \tag{S9}$$

Replacing z by $-2\pi\nu_{rf}\ (t_{flip})\ /\gamma G$ in Eq. S8 gives

$$t_{flip}(z) = T_e\left(1 - \sqrt{\frac{-2\pi\nu_{rf}\ (t_{flip}(z))}{\gamma \mathrm{GL}} + \frac{1}{2}}\right) \tag{S10}$$

$$t_{flip}(z) = T_e\left(1 - \sqrt{\frac{\nu_{rf}\ (t_{flip}(z))}{\mathrm{BW}} + \frac{1}{2}}\right) \tag{S11}$$

where BW $= -\gamma GL/2\pi$ is the bandwidth of the pulse.

Eq. 11 is valid for $z \in [-\frac{L}{2}\ \frac{L}{2}]$ and $t \in [0\ 2T_e]$. Therefore, quadratic spacing can be achieved with the pulse whose offset is given by

$$\nu_{rf} = \left(\frac{(T_e - t)^2}{T_e^2} - \frac{1}{2}\right)\mathrm{BW} \tag{S12}$$

The amplitude (A) and phase $(\varphi_{rf}\ (t))$ of the pulse is then given by

$$A = \sqrt{\frac{d\nu_{rf}}{dt}} = \sqrt{\frac{2(T_e - t)\mathrm{BW}}{T_e^2}} \tag{S13}$$

$$\phi_{rf}(t) = 2\pi\int \nu_{rf}dt = 2\pi\mathrm{BW}t\left(\frac{t^2}{3T_e^2} - \frac{t}{T} + \frac{1}{2}\right) \tag{S14}$$

In practice the pulse shape is smoothed for the first and last 10% by multiplication by a sine envelope. An example of pulse shape is shown in Fig. S2, for BW = 110 kHz and T_e = 1.5 ms.

The final expression for $K(z)$ is obtained by replacing $t_{flip}(z)$ by

$$t_{flip} = T_e\left(1 - \sqrt{\frac{-\gamma Gz}{2\pi\mathrm{BW}} + \frac{1}{2}}\right) \tag{S15}$$

in Eq. S3. This gives

$$(K(z))^2 = 4\gamma^2G^2\left(\frac{1}{2} - \frac{\gamma Gz}{2\ \pi\ BW}\right). \tag{S16}$$

Note that the effective diffusion delay Δ', which accounts for diffusion during the spatial encoding block, in principle has a different expression for the newly derived pulse. We have, however, for simplicity, kept the expression $\Delta' = \Delta\ - T_e$.

2. Multivariate processing

In SPEN DOSY, the dataset $X(z, \nu)$ represents the signal as a function of position, *z*, and frequency, ν, with the encoding phase at position *z*, given by Eq. S16. X is a $m \times n$ matrix with m spectra each having n points. In multivariate processing methods, the data is decomposed in a set of component spectra and component decays. Ideally,

$$X = CP^T \tag{S17}$$

where C is an $m \times r$ matrix of component decays, P is an $n \times r$ matrix of component spectra, and r is the number of components. The methods aim to find estimates of the component spectra and of the associated diffusion coefficients.

2.1 DECRA processing

The processed SPEN DNMR data was fed to the DECRA processing code adapted from the implementation by Nilsson and co-workers,[1,2] which is based on a script by Antalek.[3,4] A brief description of the method is as follows.

Following the work of Kubista, Scarminio[5,6] and Windig, Antalek,[3,4] the data set X, obtained after sampling the data with constant quadratic spacing in k-dimension, i.e., $K_{n+1}^2 - K_n^2 = constant$, is divided into two reduced data sets, A and B, by deleting the first and the last row respectively from X. The dataset A and B can be mathematically represented as

$$A = CP^T \tag{S18}$$

$$B = C\beta P^T \tag{S19}$$

where C is now an $(m - 1) \times r$ matrix of component decays, P is an $(n \times r)$ matrix of component spectra, A and B are $(m - 1) \times n$ matrices and β (an $r \times r$ diagonal matrix), is a scaling factor between data set A and B. Equation 19 and 20 leads to an eigenvalue equation

$$Z^*\beta = (\overline{U}^T B V S^{-1}) Z^* \tag{S20}$$

where $Z^* = SV^T Z$, $Z = (P\,T)^\dagger$. U, V, and S are left eigenvector, right eigenvector, and diagonal matrix respectively obtained after singular value decomposition of dataset A solution of this equation provides eigenvalue β and eigenvectors Z^* . The pure spectral components and their concentration profiles are obtained by $P^T = (V\,S^{-1}\,Z^*)^\dagger$ and $C = UZ^*$

The eigenvalue matrix β is utilized to obtain the diffusion constant of the pure components. The attenuation of the signals from the two dataset A and B can be written with the help of equation.

$$\frac{S_A}{S_0} = e^{-D\Delta'\left(K_n^2\right)_A} \tag{S21}$$

$$\frac{S_B}{S_0} = e^{-D\Delta'\left(K_n^2\right)_B} \tag{S22}$$

Dividing equation 21 by 22 results in matrix of eigenvalues, i.e., β which provides the diffusion constants of the corresponding components as shown in following equation

$$\beta = e^{D\Delta'\xi K_n^2} \quad \text{(S23)}$$

$$D = \frac{log\beta}{\Delta'\xi K_n^2} \quad \text{(S24)}$$

where D is the matrix of diffusion constants and $\xi K_n^2 = (K_n^2)_B - (K_n^2)_A$ which is a constant as the phase is sampled quadratically.

2.2 Speedy Component Resolution (SCORE) Processing

The processed SPEN DNMR data was fed to a SCORE processing code adapted from the implementation by Nilsson and co-workers.[1,2] A brief description of the method, adapted from Ref.[7], is as follows.

In SCORE processing, estimates for the component spectra P and decays C are obtained iteratively by minimising the error

$$E = CP^T - X \quad \text{(S25)}$$

This involves the following:

- **Initialisation**

The matrix data, $X\ (m \times n)$, is summed over n to have a vector form dataset, $X_{vec}\ (m \times 1)$. This vector dataset is fitted with an equation of the form

$$y = a_1 e^{-a_2\Delta' K(z)^2}, \quad \text{(S26)}$$

where $\Delta' = \Delta - T_e$, $K(z) = -2\gamma G(2T_e - 2t_{flip}(z))$, and a_1 and a_2 are adjustable parameters.

The initial guesses for the diffusion coefficients are then prepared as a linear grid of *r* elements centered on the best-fit parameter for a_2, and used to start the optimization.

- **Outer Optimization**

A non-linear least-square fit is used to find the values of the diffusion coefficients that minimize the sum of the squares of the difference $|X - I|^2$ between the experimental dataset, X, and the model dataset I provided by the inner optimization.

- **Inner optimization**

By using estimated diffusion coefficients, components decays, C^-, are calculated with equation (S26) for all of the components (using $a_1 = 1$). An estimate of the component spectra P^+ is then obtained by solving the system of linear equation

$$X = C^- P^{+T} \quad \text{(S27)}$$

In GNAT this is achieved by using a MATLAB function "\" (mldivide), which returns the least-square solution.

The component decays and spectra are used to obtain a model dataset as

$$I = C^- P^{+^T} \quad \text{(S28)}$$

- **Completion**

The default convergence parameters of the GNAT were used here for the non-linear least-square fit. Once the optimization is complete the diffusion constant after minimization are again obtained by solving the system of linear equations

$$X = CP^T \quad \text{(S29)}$$

Results obtained through SPEN SCORE are shown in Fig. S5 and Fig.S6 for ethanol-butan-2-ol and sucrose-propan-1-ol mixtures respectively.

3 Materials and methods

3.1 Sample preparation

Mixtures of ethanol and butan-2-ol in D_2O (M1), and sucrose and propan-1-ol in D_2O (M2) are used to perform spatially encoded diffusion NMR experiments.

For M1, 27 µL of butanol and 17µL of ethanol are solubilised in 556 µL D_2O to obtain approximate 500 mM concentration of each of the component.

For M2, 103 mg of sucrose is dissolved in a solution of 45 µL propanol and 555 µL D_2O to prepare approximate 500 mM concentration of both of the components in D_2O.

Each mixture was transferred to a standard 5 mm NMR tube for SPEN DNMR experiments. For conventional DNMR experiments, a Shigemi tube restricted to a length of 8 mm was used, to mitigate the effect of gradient non-uniformity.

A 600 µL sample of 90 % H_2O and 10 % D_2O Was prepared by mixing 60 µL D_2O in 540 µL H_2O to record reference spatial profiles.

3.2 NMR spectroscopy

All the experiments were carried out on a Bruker spectrometer operating at a Larmor frequency of 500.13 MHz, equipped with a triple-axis gradient, broadband inverse-detection probe. The temperature was set at a nominal value of 298 K and the diffusion time was 100 ms for all the DOSY experiments, except when specified.

For SPEN experiments, the diffusion encoding parameters were chosen to sweep a 10.2 mm region of the sample with the frequency swept pulses (for both linear and quadratic sweep). The pulses had a bandwidth of 110 kHz, a duration of 1.5 ms, and were applied together with a rectangular gradient, with a short ramp of 25µs placed symmetrically around the gradient, of 0.2538 T/m. The acquisition consisted of a train of bipolar gradient pulses, with an amplitude of ±0.29 T/m and a duration of 332.8 µs for each gradient pulse, resulting in a spectral width of 3 ppm. 256 loops were acquired, resulting in an acquisition time of 170.4 ms. Spatial encoding was performed along the Z axis. Coherence selection gradients were applied along the Y axis, with amplitude 0.041 T/m for a and 0.077 T/m for b, and a duration of 1000 µs. The spoiler gradient f was applied with 0.052 T/m on the X-axis and 0.061 T/m on the Y -axis, and a duration of 1000

μs. The compensation gradient was g_1 = - f. Gradients c, g_2 and g_3 were not used. The WET pulse sequence block was used for solvent suppression.

The reference spatial profiles are recorded with 10% H_2O and 90% D_2O sample. The reference profile $S_{\mathrm{ref},1}$ is an estimate of the probe's sensitivity and is obtained with the pulse sequence shown in Fig. S3a, which consists of an imaging or spectroscopic imaging acquisition after an excitation pulse. The reference profile $S_{\mathrm{ref},2}$corresponds to the effect of the pair of frequency swept pulses on the magnitude of the magnetisation. It is obtained by dividing the outcome of the pulse sequence shown in Fig. S3b by $S_{\mathrm{ref},1}$. While $S_{\mathrm{ref},2}$is close to 1 for linear sweeps for the region of interest, we find that dividing by $S_{\mathrm{ref},1}$ is important in the case of the quadratic sweep. An improved choice of amplitude modulation (Eq. S13) may alleviate this constraint.

Conventional experiments were recorded using a stimulated echo sequence with bipolar gradient pulses (stebpgp1s Bruker sequence with additional lock stabilisation gradients). For the data analysed in Table S1, the diffusion- encoding gradient values consisted of a quadratic ramp ranging from 0.065 T/m to 0.522 T/m. The duration of the gradient pulses was 1000 μs. The 2D spectra were acquired with 16384 points in the direct dimension, a spectral width of 4504.5 Hz, a relaxation delay of 5 s, 16 gradient steps, and 8 scans per increment, resulting in a total experiment duration of 14 min 56s. For the data analysed in Fig. S7 , the diffusion- encoding gradient values consisted of a quadratic ramp ranging from 0.065 T/m to 0.587 T/m with duration of 1300 μs was used, and the diffusion time was 90 ms. The 2D spectra were acquired with 30998 points in the direct dimension, a spectral width of 5000 Hz, a relaxation delay of 3 s, 16 gradient steps, and 8 scans per increment, resulting in a total experiment duration of 13 min 44 s. This experiment was additionally carried out with diffusion-encoding parameters optimised to provide an example of multivariate processing of conventional DOSY data, which has limited performance on this system, likely because of gradient non-uniformity.

3.3 Data Processing

The experimental data was processed with MATLAB R2018b (The Mathworks), using a combination of custom-written code and code from the GNAT.[2] The raw data is pre-processed before univariate or multivariate analysis.

The experimental raw data is baseline corrected (to remove a potential DC offset), rearranged into a 2D matrix, where the first dimension is the k dimension (corresponding to the acquisition gradient area multiplied by the gyromagnetic ratio), and the second dimension is t (the acquisition time). This data, in (k, t), space, is apodized in both dimensions. A Hann window is used in the k dimension, and a sine window is used in the time dimension. Both dimensions are zero-filled with 2048 points.

2D Fourier transform of the data yields spectroscopic imaging data in (z, ω) space. However, several additional steps are required to account for: i/ the effect of chemical shift offsets during acquisition and encoding, ii/ the spatial profile of the probe's sensitivity and of the frequency swept pulse.

To compensate for the effect of chemical-shift offsets during acquisition, the data in (z, ω) space is Fourier transformed along the first dimension, multiplied by $\exp(-ik\Delta\omega/(\gamma G_a))$ (where G_a is the acquisition gradient, and $\Delta\omega = \omega - \omega_R$, with ω_R the carrier frequency), and inverse Fourier transformed along the second dimension. After this step, the nominal value of *z* for a pixel corresponds to its physical position. The data is then divided by the reference profile $S_{\mathrm{ref},1}$, corresponding to the probe's sensitivity.

To compensate for the effect of chemical-shift offsets during encoding, the data in (k, t) is Fourier transformed along the second dimension, multiplied by $\exp(-ik\Delta\omega/(\gamma G_e))$, where G_e is the encoding gradient, and inverse Fourier transformed along the second dimension. After this step, the nominal value of *z* for a pixel is such that the effective gradient area given by Eq. S16 is correct. The data is then divided by the reference profile $S_{\mathrm{ref},2}$, corresponding to the RF selectivity profile.

The resulting, pre-processed (z, ω) is ready for univariate or multivariate processing. For univariate processing, integration regions are defined for each peak from a 1D spectrum obtained as a slice of the data, and the spatial region used for the fit is chosen manually from a 1D spatial profile obtained as a slice of the data. Then, for each peak, the spatial profiles are fitted with Eq. 1 of the main text. The DOSY representation was obtained using Gaussian lineshapes with maxima corresponding to the calculated *D* value and linewidths set by the error of the fit. For multivariate processing, the full spectral points are used for DECRA processing, and the spatial range is selected manually. The result of DECRA processing is found to be sensitive to the choice of spatial range. Failure of DECRA separation can be diagnosed with the presence of negative peaks in the separated components and complex values for the diffusion coefficients.

4 Supplementary figures

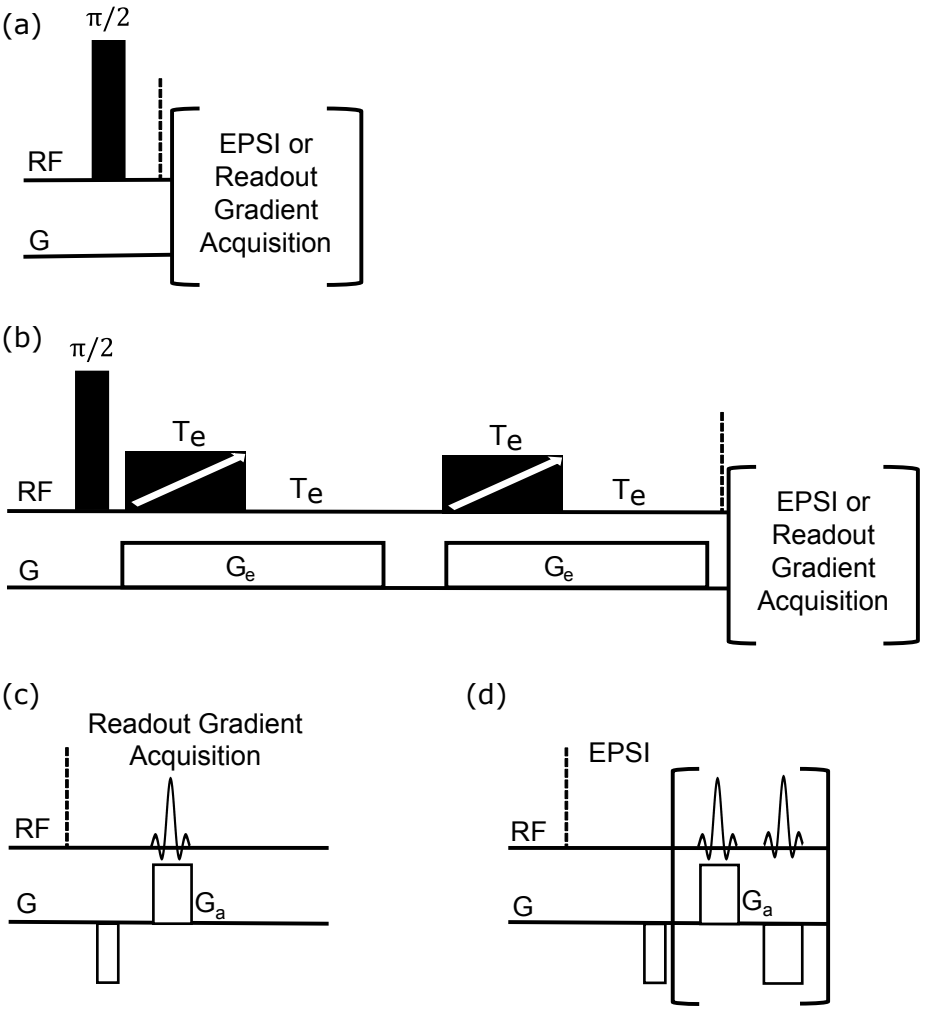

Figure S1: Pulse sequences to obtain reference spatial profiles. Pulse sequence (a) is used to record reference profile S_{ref1}, which is an estimate of the probe's sensitivity. Pulse sequence (b) is used to ($S_{ref2} \times S_{ref1}$), where S_{ref2} is the effect of a pair of frequency-swept pulses on the magnitude of the magnetisation. The acquisition block may consist of (c) a single readout gradient, or (d) echo-planar spectroscopic imaging (EPSI)

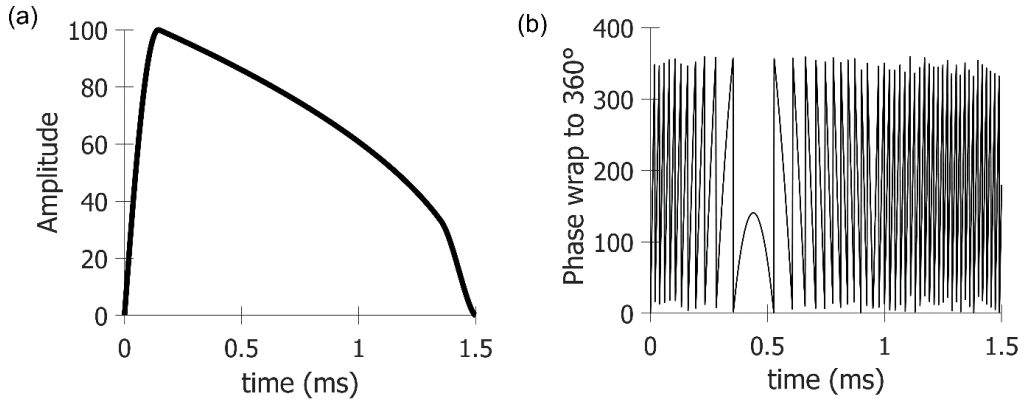

Figure S2 (a) Amplitude and (b) phase wrapped to 360^0 of a quadratic-sweep pulse with a bandwidth of 110 kHz and a length of 1.5 ms.

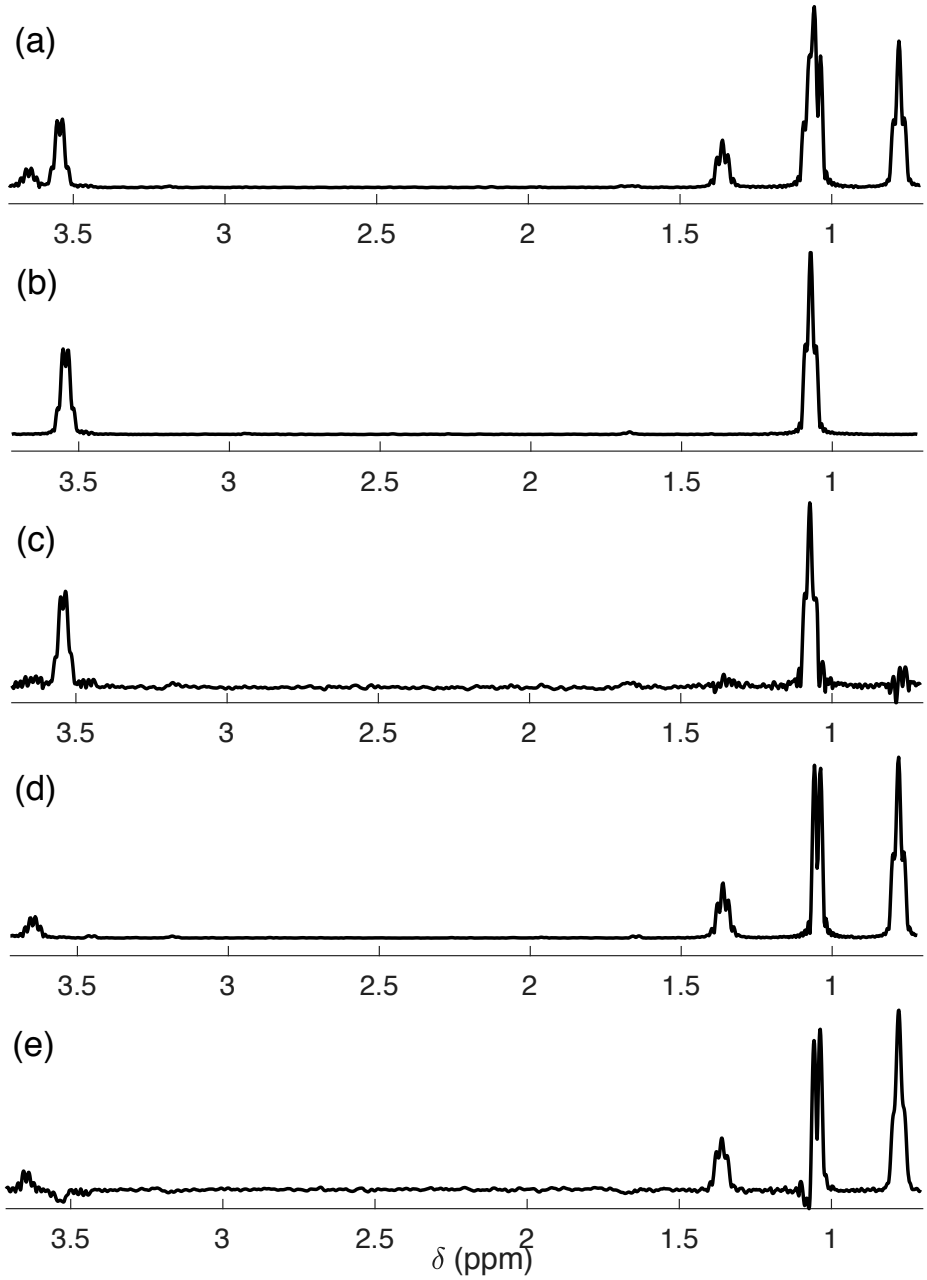

Figure S3 DECRA processing of SPEN DNMR data for a mixture of ethanol and butan-2-ol in D_2O, compared to the spectra of the compounds acquired separately. The residual HDO peak is suppressed by a WET pulse sequence block. (a) 1D spectrum of a mixture of ethanol and butan-2-ol in D_2O. (c, e) Components obtained by DECRA processing of SPEN DNMR data. (b, d) 1D spectra of (b) ethanol and (d) butan-2-ol alone in D_2O. The 1D spectra are obtained as slices of EPSI data.

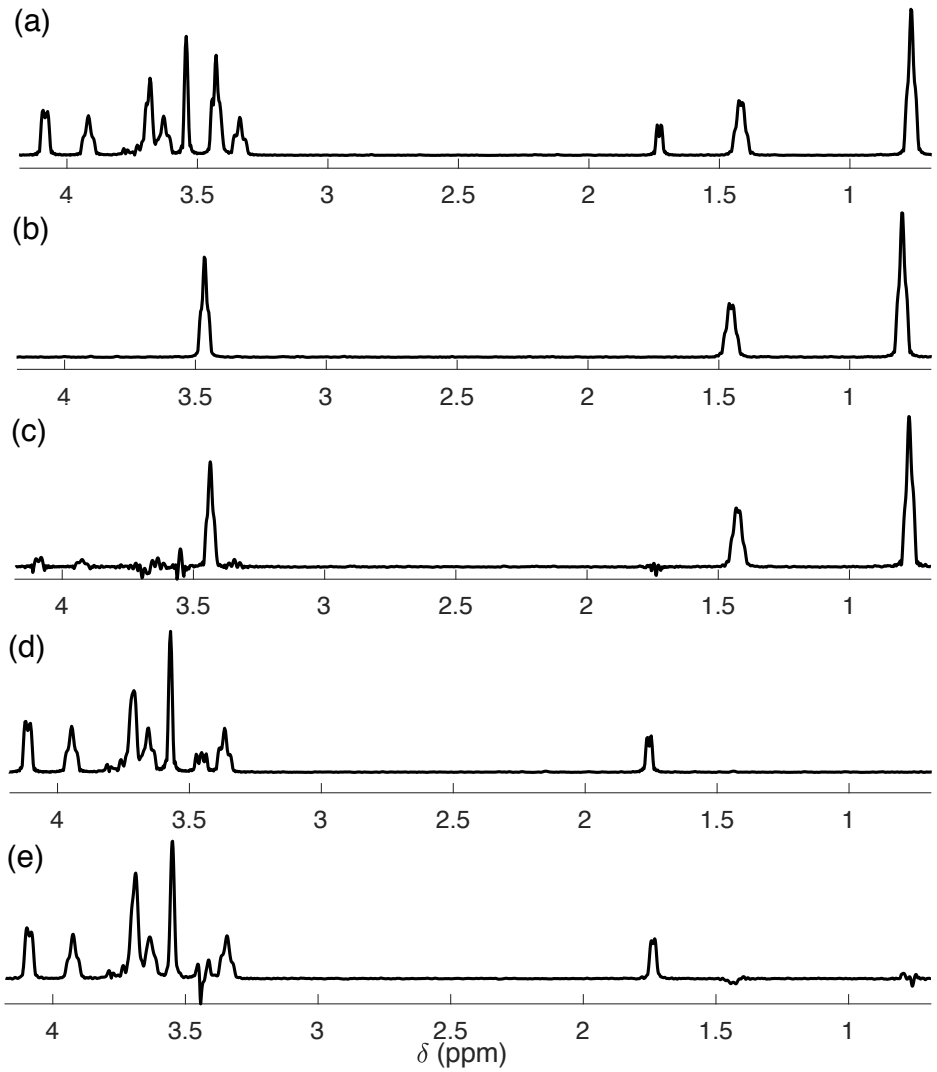

Figure S4 DECRA processing of SPEN DNMR data for a mixture of propan-1-ol and sucrose in D_2O, compared to the spectra of the compounds acquired separately. The residual HDO peak is suppressed by a WET pulse sequence block. (a) 1D spectrum of a mixture of propan-1-ol and sucrose in D_2O. (c, e) Components obtained by DECRA processing of SPEN DNMR data. (b, d) 1D spectra of (b) propan-1-ol and (d) sucrose alone in D_2O. The 1D spectra are obtained as slices of EPSI data.

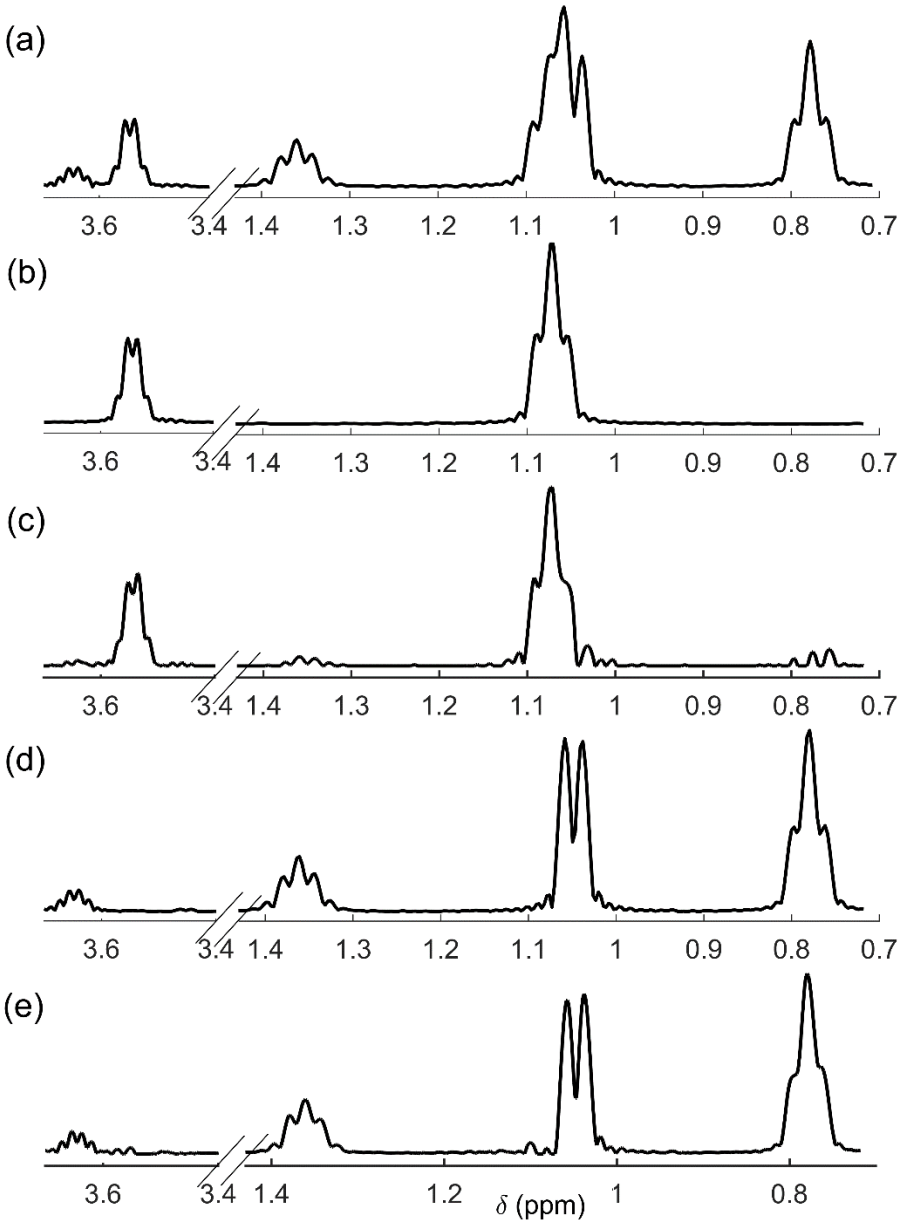

Figure S5 SCORE processing of SPEN DNMR data for a mixture of ethanol and butan-2-ol in D_2O, compared to the spectra of the compounds acquired separately. The residual HDO peak is suppressed by a WET pulse sequence block. (a) 1D spectrum of a mixture of ethanol and butan-2-ol in D_2O. (c, d) Components obtained by SCORE processing of SPEN DNMR data. (b, d) 1D spectra of (b) ethanol and (d) butan-2-ol alone in D_2O. The 1D spectra are obtained as slices of EPSI data.

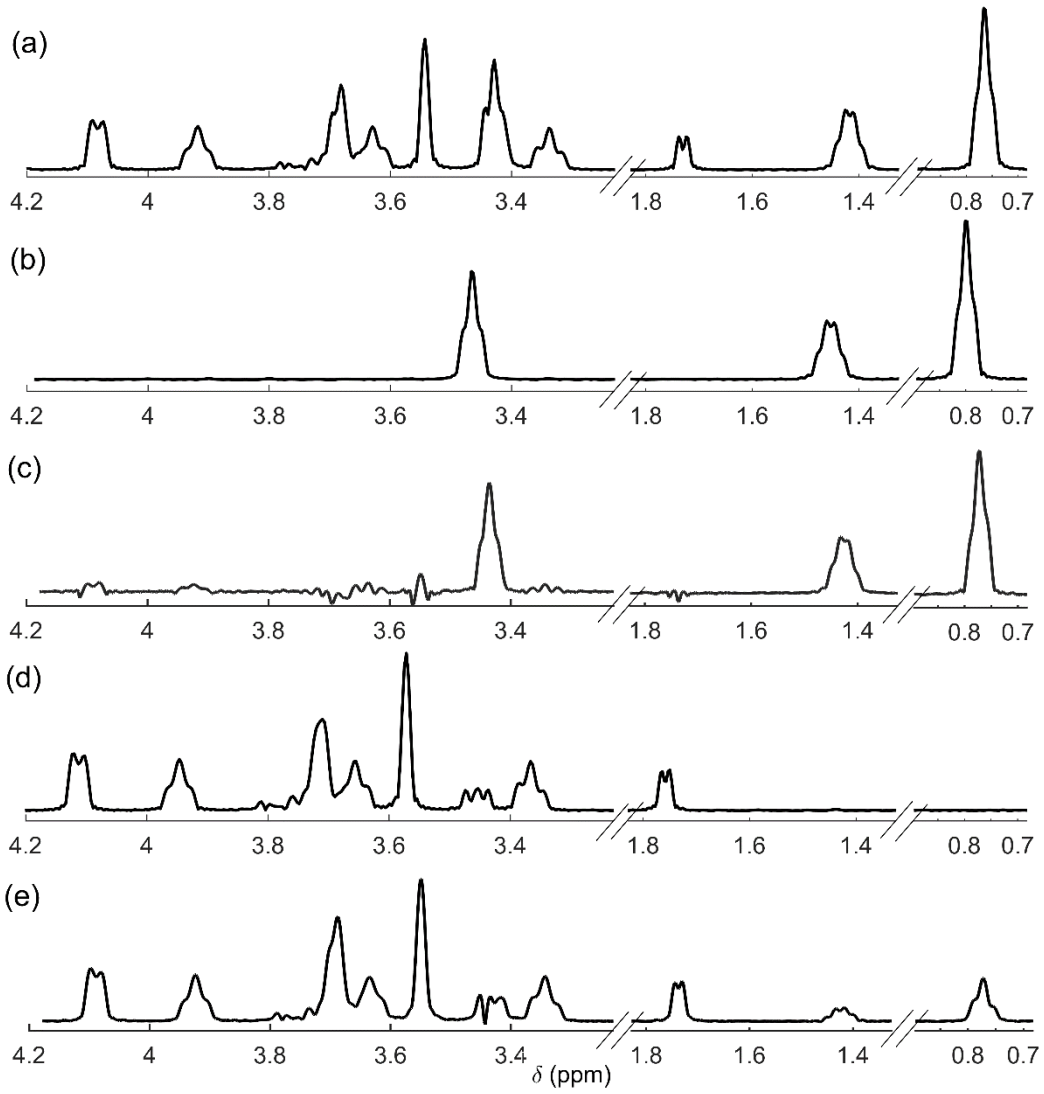

Figure S6 SCORE processing of SPEN DNMR data for a mixture of propan-1-ol and sucrose in D_2O, compared to the spectra of the compounds acquired separately. The residual HDO peak is suppressed by a WET pulse sequence block. (a) 1D spectrum of a mixture of propan-1-ol and sucrose in D_2O. (c, e) Components obtained by SCORE processing of SPEN DNMR data. (b, d) 1D spectra of (b) propan-1-ol and (d) sucrose alone in D_2O. The 1D spectra are obtained as slices of EPSI data.

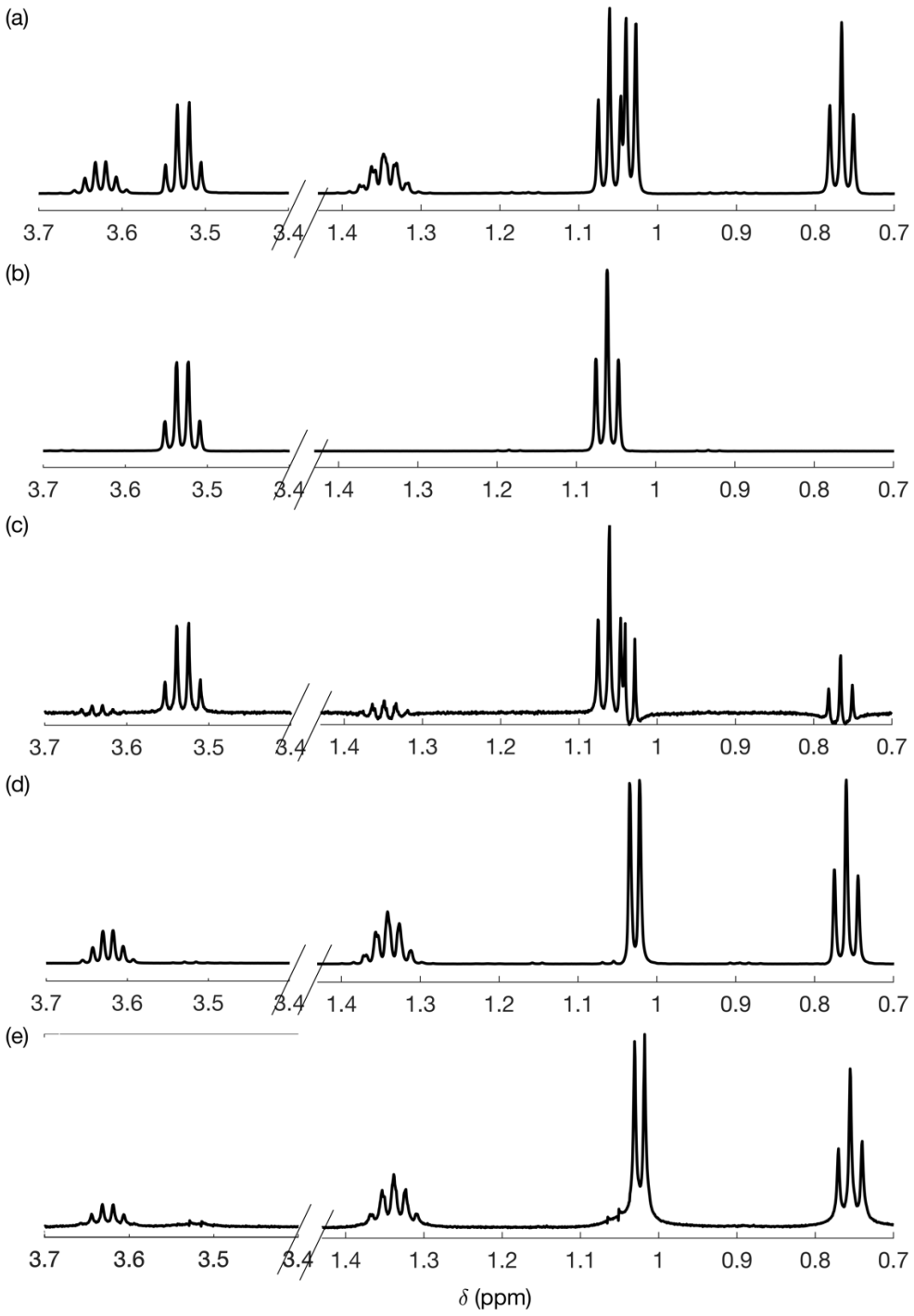

Figure S7 DECRA processing of conventional DOSY data for a mixture of ethanol and butan-2-ol in D_2O, compared to the spectra of the compounds acquired separately. The residual HDO peak is omitted for the better comparison between SPEN DNMR and conventional DOSY. (a) 1D spectrum of a mixture of ethanol and butan-2-ol in D_2O. (c, d) Components obtained by DECRA processing of conventional DOSY data. (b, d) 1D ^{1}H-NMR spectra of (b) ethanol and (d) butan-2-ol alone in D_2O.

5 Supplementary Table

Table S1 Diffusion coefficients for the ethanol butanol mixture, from univariate (single-exponential fit) and multivariate (DECRA) analysis.

Components	δ (ppm)	Diffusion coefficients[a] (x $10^{10} m^2.s^{-1}$)			
		Conventional	Spatially encoded[b]		
			Univariate		DECRA[c]
			Linear spacing	Quadratic spacing	
ethanol	3.47	8.67 ($\pm$0.02)	9.29 ($\pm$0.07)	9.38 ($\pm$0.08)	9.11 ($\pm$0.08)
butan-2-ol	3.60	6.44 ($\pm$0.07)	7.29 ($\pm$0.10)	7.25 ($\pm$0.09)	6.49 ($\pm$0.52)
	1.28	6.23 ($\pm$0.02)	6.90 ($\pm$0.06)	7.02 ($\pm$0.10)	
	0.73	6.22 ($\pm$0.02)	7.06 ($\pm$0.07)	6.91 ($\pm$0.06)	
overlap	0.97	7.01 ($\pm$0.06)	8.26 ($\pm$0.06)	8.00 ($\pm$0.06)	

[a] For univariate processing, the uncertainty is estimated from the non-linear least square fit.

[b] The values obtained with SPEN experiments are about 10% larger than with conventional experiments. This systematic error is due in part to gradient non uniformity.

[c] For DECRA there is one value per component, instead of one value per peak for univariate processing. The uncertainty is estimated from the repetition of three experiments.

6 References

1. Nilsson, M. The DOSY Toolbox: A new tool for processing PFG NMR diffusion data. *J. Magn. Reson.* **200**, 296–302 (2009).

2. Castañar, L., Poggetto, G. D., Colbourne, A. A., Morris, G. A. & Nilsson, M. The GNAT: A new tool for processing NMR data. *Magn. Reson. Chem.* **56**, 546–558 (2018).

3. Antalek, B. Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **14**, 225–258 (2002).

4. Windig, W. & Antalek, B. 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. *Chemom. Intell. Lab. Syst.* **37**, 241–254 (1997).

5. Scarminio, I. & Kubista, M. Analysis of Correlated Spectral Data. *Anal. Chem.* **65**, 409–416 (1993).

6. Kubista, M. A new method for the analysis of correlated data using procrustes rotation which is suitable for spectral analysis. *Chemom. Intell. Lab. Syst.* **7**, 273–279 (1990).

7. Nilsson, M. & Morris, G. A. Speedy component resolution: An improved tool for processing diffusion-ordered spectroscopy data. *Anal. Chem.* **80**, 3777–3782 (2008).