Diffusion NMR and trilinear analysis in the study of reaction kinetics

Mathias Nilsson,*[a] Maryam Khajeh,[a] Adolfo Botana,[a] Michael A. Bernstein[b] and Gareth A. Morris[a]

Received (in Cambridge, UK) 20th November 2008, Accepted 10th December 2008
First published as an Advance Article on the web 15th January 2009
DOI: 10.1039/b820813a

Measurement of diffusion-weighted NMR spectra as a function of time allows the time-dependence of concentration and the isolated spectrum to be found for each component in a reaction, without prior assumptions about spectra, kinetics or diffusion behaviour, by data decomposition using the PARAFAC algorithm.

Nuclear magnetic resonance (NMR) is frequently employed to study reaction kinetics. NMR can provide detailed structural information about (and often identify) the chemical entities involved in a reaction, and as it is non-invasive and non-destructive, the kinetics of an intact mixture can be studied in real time directly in the NMR tube[1] (alternatives exist for the study of reaction conditions that cannot be duplicated in an NMR probe[2]). Reaction monitoring by quantitative NMR works best when each component in a reaction mixture has at least one well-resolved resonance; the change in peak integral can then be used directly to determine the kinetic behaviour.[3] When no resolved peaks are available, as is quite common, the extraction of kinetic data becomes much more challenging, and it is often impossible to identify individual reaction components, let alone determine their concentrations. We demonstrate here that by adding diffusion information to the NMR experiments, the spectrum, time evolution and diffusion data can be recovered for each component in a reaction mixture. Because the data are trilinear (*i.e.* vary independently in three dimensions, here diffusional attenuation, time evolution and chemical shift) they can be decomposed using a PARAFAC[4] (parallel factor analysis) algorithm, and it is therefore possible to analyse the data without the need for fitting to a predetermined model, and without having to constrain the data to fit either the reaction kinetics or the diffusional attenuation.

The study of reactions is an example of the general case of mixture analysis by NMR. It is well known that it can be frustrating to study intact mixtures by NMR, as it is often difficult to assign resonances unambiguously to given mixture components. It is expensive, tedious and time-consuming to separate components physically (*e.g.* by chromatography) before subjecting them to NMR, and frequently it is the study of the intact mixture itself that is of interest (as for reaction monitoring). Therefore it is desirable to develop NMR methods that can recover the required information from intact mixtures. Some of the most powerful NMR methods currently available, commonly referred to as DOSY (diffusion-ordered spectroscopy) experiments, are based on diffusion;[5–9] these are most effective where each component in a mixture has a unique rate of diffusion. The diffusion of molecules can be measured by recording the signal attenuation in a pulsed field gradient NMR experiment,[10] typically by incrementing the gradient strength in a pulsed field gradient spin or stimulated echo. It was recognised early on that the results of such experiments can be used to distinguish the signals from different molecular species.[11] The decays of individual NMR signals are typically fitted to a model function, and the fitted diffusion coefficient is then used to correlate the signals of individual molecular species. In high resolution DOSY,[5,6] this is done by fitting each peak individually (implicitly assuming that there is no spectral overlap), while in multivariate methods the whole bandwidth is fitted simultaneously.[7–9] The model function used is typically some form of the Stejskal–Tanner equation,[10] which describes the effect of pulsed field gradient on signal amplitude; for best results, the equation can be extended to include the effects of imperfect field gradient uniformity.[6,12]

DOSY and kinetic studies by NMR have much in common: both rely on fitting variations in signal amplitude to suitable model functions, and in both cases it is far easier to analyse experimental data when the NMR signals of individual species are well resolved. DOSY data and timecourse spectra are bilinear: signal intensity I is measured as a function of two variables, frequency and gradient amplitude, and frequency and time, respectively. In a bilinear dataset, the theoretical intensity I_i for a given signal i is the product of the signal variation as a function of two different variables, $I_i(p,q) = P_i(p)Q_i(q)$. Thus in DOSY, if the spectrum of component i is $S_i(f)$ and its signals attenuate as a function of gradient g according to $A_i(g)$, then $I_i(f,g)$ is the product of $S_i(f)$ and $A_i(g)$. The experimental dataset is a tensor of rank 2, and may be represented as a sum over N components i of outer products of two vectors $\boldsymbol{S}_i$ and $\boldsymbol{A}_i$, plus some residual $\boldsymbol{E}$:

$$\boldsymbol{I} = \sum_{i=1}^{N} \boldsymbol{S}_i \boldsymbol{A}_i + \boldsymbol{E} \qquad (1)$$

In analysing bilinear data with spectral overlap it is common to use multivariate methods to help resolve the component spectra (and diffusion/kinetics).[7–9,13–16] Unfortunately such analyses suffer from *rotational ambiguity*: any linear combination of the true functions P_i, or the true functions Q_i, gives an equally good fit to the experimental data. For bilinear analysis it is therefore necessary to apply constraints,[17] for example non-negativity and/or known/hypothesised kinetic models, to

[a] *School of Chemistry, University of Manchester, Oxford Road, Manchester, UK M13 9PL*
[b] *AstraZeneca R&D Charnwood, Bakewell Rd, Loughborough, Leics, UK LE11 5RH. E-mail: mathias.nilsson@manchester.ac.uk*

allow the true solutions to be selected out from the infinite range of linear combinations. This problem can be avoided, and a model-free fit obtained by PARAFAC decomposition, if trilinear data $I_i(p,q,r) = P_i(p)Q_i(q)R_i(r)$, in which I varies linearly with P, Q and R, can be measured. Adding a diffusion dimension to a bilinear dataset can create a trilinear structure.[18] Recording NMR spectra as a function both of time and of gradient amplitude, *i.e.* measuring a timecourse of DOSY spectra, gives just such a dataset. No prior knowledge of the component spectra, diffusion behaviour or kinetics is needed; the only requirement is that the spectrum $S_i(f)$, diffusional attenuation $A_i(g)$ as a function of gradient g, and concentration profile $C_i(t)$ of each species be independent of each other, so that the signal intensity $I_i(f,g,t) = S_i(f)A_i(g)C_i(t)$. The experimental dataset is now a rank 3 tensor:

$$\boldsymbol{I} = \sum_{i=1}^{N} \boldsymbol{S}_i \boldsymbol{A}_i \boldsymbol{C}_i + \boldsymbol{E} \qquad (2)$$

To demonstrate the value of using diffusion encoding in the NMR study of a reacting mixture we have chosen the well-known acid hydrolysis of maltose to glucose.[19] An aqueous solution of maltose (5.5% w/w) in 33% (w/w) sulfuric acid was prepared, with 0.15% (w/w) pivalic acid as a reference compound. Hydrolysis was carried out at 50 °C in a thick-walled NMR tube (to prevent convection; i.d. 2.2 mm) in a 400 MHz Varian Inova instrument, using a 5 mm diameter indirect detection probe equipped with a *z*-gradient coil allowing gradient pulses up to 30 G cm^{-1}. 98 DOSY experiments were carried out over the course of the reaction, in a total of 41 h 49 min. Each DOSY experiment used the Oneshot sequence[20] with 32 transients at each of 6 gradient levels, spaced equally in gradient squared, ranging from 3.0 to 27.3 G cm^{-1}. The data were then Fourier transformed, phase corrected, baseline corrected, reference deconvoluted[17] using the pivalic acid signal, and the solvent (HOD) peak was removed by digital filtering, all using the manufacturer's VnmrJ software, before export to MATLAB. PARAFAC analysis was performed with the MATLAB N-Way toolbox.[21,22] Fig. 1 shows a subset of the experimental spectra as a function of time and gradient level. Small variations in receiver sensitivity during the experiment were corrected for by normalising the integral of the spectrum for each gradient level using the average area of the pivalic acid reference peak.

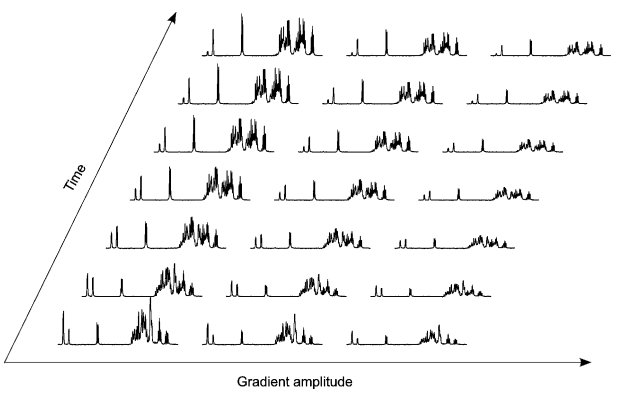

Fig. 1 A subset of the raw experimental data. For the time evolution every 16th spectrum is shown, and for the decay with gradient amplitude (caused by diffusion) the first three gradient levels are shown.

PARAFAC fitting was carried out for the spectral region 3.1–5.5 ppm with only one assumption, that there were two components. The resultant fit accounted for 99.8% of the variance in the data, yielding statistical components $S_i(f)$, $A_i(g)$ and $C_i(t)$ representing the spectrum, signal decay as a function of gradient strength, and time evolution for the reactant and product, respectively.

One great advantage of PARAFAC is that, if the assumption of trilinearity holds, the fitted components obtained should have physical relevance, *i.e.* should in this case be the true spectrum, diffusional attenuation and concentration timecourse. Where prior information exists, therefore, it is possible to assess directly the quality of the PARAFAC decomposition, for example by comparing the spectra of reaction components obtained by PARAFAC with the spectra of the pure materials. As can be seen in Fig. 2, in this case the fitted spectra are virtually identical to the spectra of pure maltose and pure glucose, confirming that trilinear decomposition into two components was successful.

Because the relative scaling of the three multiplicands in the trilinear model is arbitrary, to obtain true relative concentrations it is necessary to ensure that the other two modes in the model, $S_i(f)$ and $A_i(g)$, are normalised. Where, as here, the structures of the reaction components are known this is straightforward: each PARAFAC spectral mode $S_i(f)$ is normalised to have an integral proportional to the number of protons involved, and each diffusion mode $A_i(g)$ is normalised so that it extrapolates to unity at zero gradient g. Multiplying the remaining raw modes $C_i(t)$ by the normalisation factors by which the $S_i(f)$ and $A_i(g)$ were divided then gives $C_i(t)$ modes, which are directly proportional to concentration.

The net result for the experimental data of Fig. 1 is shown in Fig. 3B, and as expected gives an excellent fit to first order kinetics. The PARAFAC results can, since well-resolved anomeric signals are available, be compared with concentration profiles obtained by direct integration of the respective

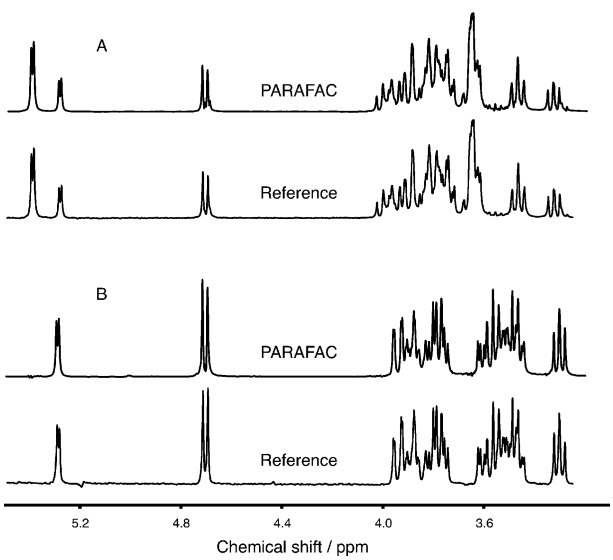

Fig. 2 Reference spectra of pure materials and spectra obtained from the data of Fig. 1 by PARAFAC for maltose (A), and glucose (B).

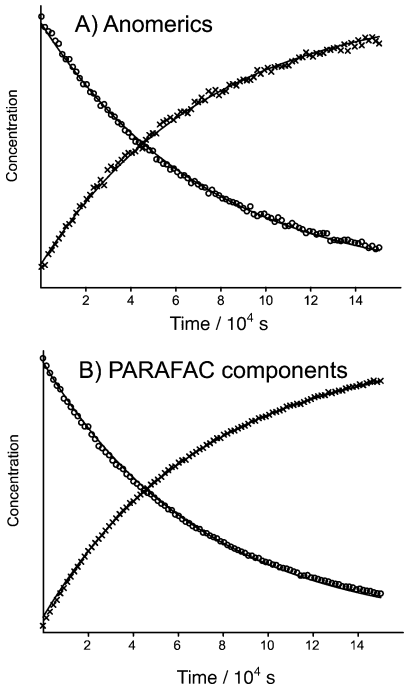

Fig. 3 Experimental data and non-linear least squares fits to first order kinetics for the acid hydrolysis of maltose to glucose. (A) relative concentrations of maltose (integral of the inner anomeric signal at 5.4 ppm, decaying curve) and glucose (sum of the integrals of the terminal alpha and beta peaks between 4.6 and 5.3 ppm minus the integral of that at 5.4 ppm, rising curve). Estimated rate constant $k = 1.36 \pm 0.02 \times 10^{-5}$ s^{-1}. (B) Normalised PARAFAC[4] components $C_i(t)$ for maltose (decaying) and glucose (rising); $k = 1.40 \pm 0.01 \times 10^{-5}$ s^{-1}.

anomeric signals (Fig. 3A); while there is excellent agreement, the scatter for the PARAFAC result is, as expected, significantly smaller. Although for this example the full range of chemical shifts, including the well-resolved anomeric signals, was used, essentially identical results were obtained when only the highly overlapped region between 3.1 and 4.1 ppm was fitted, confirming that both fully-resolved spectra and kinetic information can be recovered even when the experimental data contain no resolved peaks. The diffusion modes $A(g)$ show the expected near-Gaussian form, with diffusion coefficients for maltose and glucose of 5.4 ± 0.2 and $6.5 \pm 0.2 \times 10^{-10}$ m^2 s^{-1}, respectively. The PARAFAC decomposition is remarkably robust; similar results can be obtained using only two of the six gradient increments measured, and/or with many fewer time points, and/or with much poorer signal-to-noise ratio.

From this model study it is clear that by adding diffusion information to an experimental timecourse study and using multi-way methods to decompose the results, it is possible not only to obtain good kinetic data irrespective of whether any resolved signals are available, but also to recover the NMR spectra of individual reaction components. In principle, one could obtain by this method the NMR spectra of intermediates that are difficult or impossible to isolate. The fundamental requirement is that each reaction component show a different diffusion coefficient and a different timecourse; even where this is not fully met, a hybrid analysis in which PARAFAC is constrained using prior knowledge (*e.g.* spectral non-negativity) can still succeed. It should, however, be stressed that for a trilinear PARAFAC decomposition, no assumptions are needed about the form of the spectra, the diffusional attenuation, or the kinetics. In principle, the method described should be applicable to mM concentrations, and offer a limiting time resolution of a few tens of seconds.

Support from the Engineering and Physical Sciences Research Council (grant ref. EP/D05592X, EP/E057888/1 and EP/E05899X) is gratefully acknowledged. MK thanks the EPSRC and AstraZeneca for an Industrial CASE studentship, and AB thanks the EPSRC for a project studentship.

Notes and references

1 D. Kaufman, C. Sterner, B. Masek, R. Svenningsen and G. Samuelson, *J. Chem. Educ.*, 1982, **59**, 885–886.
2 M. A. Bernstein, M. Stefinovic and C. J. Sleigh, *Magn. Reson. Chem.*, 2007, **45**, 564–571.
3 J. Chrastil, *Comput. Chem.*, 1988, **12**, 289–292.
4 R. Harshman, *UCLA Work. Pap. Phonet.*, 1970, **16**, 1–84.
5 C. S. Johnson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1999, **34**, 203–256.
6 G. A. Morris, in *Encyclopedia of Nuclear Magnetic Resonance*, ed. D. M. Grant and R. K. Harris, John Wiley & Sons Ltd, Chichester, 2002, vol. 9, Advances in NMR, pp. 35–44.
7 M. Nilsson and G. A. Morris, *Anal. Chem.*, 2008, **80**, 3777–3782.
8 L. C. Van Gorkom and T. M. Hancewicz, *J. Magn. Reson.*, 1998, **130**, 125–130.
9 W. Windig and B. Antalek, *Chemom. Intell. Lab. Syst.*, 1997, **37**, 241–254.
10 E. O. Stejskal and J. E. Tanner, *J. Chem. Phys.*, 1965, **42**, 288–292.
11 P. Stilbs, *Anal. Chem.*, 1981, **53**, 2135–2137.
12 P. Damberg, J. Jarvet and A. Gräslund, *J. Magn. Reson.*, 2001, **148**, 343–348.
13 E. Bezemer and S. Rutan, *Anal. Chim. Acta*, 2002, **459**, 277–289.
14 A. de Juan and R. Tauler, *Anal. Chim. Acta*, 2003, **500**, 195–210.
15 R. Binstead, L. Stultz and T. Meyer, *Inorg. Chem.*, 1995, **34**, 546–551.
16 R. A. Binstead, C. W. Chronister, J. F. Ni, C. M. Hartshorn and T. J. Meyer, *J. Am. Chem. Soc.*, 2000, **122**, 8464–8473.
17 J. Jaumot, V. Marchan, R. Gargallo, A. Grandas and R. Tauler, *Anal. Chem.*, 2004, **76**, 7094–7101.
18 H. T. Pedersen, M. Dyrby, S. B. Engelsen and R. Bro, in *Annual Reports On NMR Spectroscopy*, ed. G. A. Webb, Academic Press, London, 2006, vol. 59, pp. 207–233.
19 J. N. BeMiller and R. K. Mann, *Carbohydr. Res.*, 1966, **2**, 70–79.
20 M. D. Pelta, G. A. Morris, M. J. Stchedroff and S. J. Hammond, *Magn. Reson. Chem.*, 2002, **40**, S147–S152.
21 C. A. Andersson and R. Bro, *Chemom. Intell. Lab. Syst.*, 2000, **52**, 1–4.
22 R. Bro, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 149–171.

