RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

HD-2D: Routine high-dispersion two-dimensional NMR spectra at no extra cost.

Juan A. Aguilar[1*], Raquel Belda[1], Adolfo Botana[2] and Alan M. Kenwright[1]

[1] Durham University, Chemistry Department (Durham, United Kingdom); [2] Agilent Technologies UK, (Oxford, United Kingdom)

Abstract

The synergic use of Pure Shift NMR techniques and Compressive Sensing allows the recording of two-dimensional NMR spectra with much higher dispersion (effective resolution) than can be achieved using conventional techniques. This gain requires no additional hardware and no increase in experiment time. We show that this combination of techniques can be implemented routinely and illustrate the significant advantages that result, enabling analyses that would otherwise have been extremely problematic.

Introduction

Throughout science the concept of "resolution" (the minimum separation necessary between two features before they become distinguishable) is a key parameter in determining the quality of a measurement. The fine details of what constitutes "resolution" vary with the measurement technique under consideration but classic examples demonstrating the importance of resolution include the Hubble space telescope (where optics problems leading to initially disappointing resolution were corrected by astronauts in orbit, leading to higher resolution than had previously been possible on earth and resulting in a number of important discoveries in optical astronomy), and

the development of super-resolution microscopy (permitting resolution beyond the previously assumed wavelength limit, for which the 2014 Nobel Prize in Chemistry was awarded to Eric Betzig, W.E. Moerner and Stefan Hell).

In NMR experiments the effective resolution may be limited by a number of factors, some of which relate to the instrumentation, some to the intrinsic properties of the sample, and some to the experimental parameters. For example, resolution may be limited by the achievable magnetic field homogeneity and stability. In modern instruments these factors are rarely limiting, and linewidths of less than 1 Hz for a singlet are readily achieved in proton spectra and maintained within a variation of less than the linewidth for periods of several hours or even days.

Resolution may also be limited by the intrinsic linewidths of the signals in cases where this is greater than the limit imposed by the magnetic field homogeneity. In this context the effective "linewidth" may be defined as the spectral range (in Hz) occupied by the signal corresponding to a single proton site in the molecule. This is different to the achievable instrumental resolution (width of a single line) since the effective "linewidth" is often largely independent of the magnetic field homogeneity. For example, in the ^{1}H spectra of paramagnetic lanthanide complexes, the observed linewidths are determined by paramagnetic broadening and may be tens or hundreds of Hz. [1] Nevertheless, useful spectra, including 2D experiments, can be obtained from these samples because the relevant spectral window is increased to hundreds of ppm, which gives a workable dispersion. In the more familiar case of ^{1}H spectra of diamagnetic small molecules, the effective "linewidth" for a particular multiplet signal is often determined by the width of the coupling pattern for that signal. While this varies considerably from site to site, an average over a large number of sites would typically be of the order of 10 Hz. Note that this value (in Hz) is independent of the magnetic field strength, while the separation between different signals (in Hz) scales linearly with magnetic field strength. This is one of the reasons why NMR spectroscopists have sought ever higher magnetic field strengths - precisely because higher field strength increases the available dispersion in ^{1}H spectra.

Finally, resolution may be limited by experimental parameters, particularly by what is commonly referred to as the "digital resolution", meaning the difference (in Hz) between two discrete data points in the digitised spectrum. With modern computers and instruments this should not be any sort of limitation in 1-dimensional spectra, but in 2-dimensional experiments it may still be a factor for two reasons. The first is that the digital resolution in the final spectrum is directly related to the number of increments acquired, at least for classical, linearly incremented 2-dimensional spectra, and the duration of the experiment is also directly related to the number of increments so, in classical 2-dimensional experiments, doubling the digital resolution in F_1 doubles the experiment time. The second is that the total digital resolution in both dimensions is related to the total size of the data set, so doubling the digital resolution in both F_1 and F_2 requires an acquired data set four times as big. While this may appear to be a trivial consideration given the availability of cheap computing power, it is still the case that a number of software packages commonly used for NMR data processing struggle with data sets larger than 4k x 4k.

In this article we look at real examples where the resolution available in classical 2-dimensional experiments would be inadequate on any realistic timescale, and show how the synergic use of Pure Shift NMR techniques and Compressive Sensing can increase the effective resolution in two dimensional experiments by more than an order of magnitude without increasing the experiment time required.

The ideas presented here are the realization of suggestions made in a paper we published in 2015, [2] in which the combined use of non-uniform sampling techniques, and of Pure Shift methods were proposed. A few publications have appeared elsewhere that make some use of such a combination, [3,4] but without fully exploiting the synergistic potential of the combined techniques.

Here, we discuss the appropriate use of these techniques to produce spectra with very high resolution reliably and routinely. Due to the high degree of dispersion (effective resolution) present,

we refer to the resulting spectra as “HD-2D”, while we use the terms “Compressed NMR” to describe the synergic combination of Pure Shift and Compressive Sensing techniques.

Results and Discussion

The effective resolution in two-dimensional NMR spectra of small molecules (where T_2 is typically long and therefore not a limiting factor) is determined by a combination of the dispersion in the spectrum and the digital resolution. The dispersion depends on the separation between signals of interest and the “linewidth” of those signals. In the majority of modern two-dimensional experiments the directly detected nucleus is proton. As has already been mentioned, the effective "linewidth" for a particular multiplet signal is often determined by the width of the coupling pattern for that signal. Partly as a result of this and partly as a result of largely historical constraints on data set size, the majority of automated set-up macros for two-dimensional experiments will set the digital resolution in the directly detected dimension to be of the order of 6 or 7 Hz per point. This keeps the size of the FID per transient small and means that effects due to proton-proton J coupling are largely obscured in the resulting spectrum.

For a classical, uniformly sampled, two-dimensional experiment the digital resolution in the indirectly detected dimension is purely a function of the number of increments used and the spectral window covered. For a given spectral window the determining factor is therefore the number of increments, which is directly proportional to the experiment time. Even if there is sufficient experiment time available, acquiring homonuclear two-dimensional proton experiments to high digital resolution in the indirectly detected dimension may still not increase the effective resolution because of the limits on dispersion caused by J-coupling, as discussed above for the directly detected dimension. (See Supplementary Information, Figure S1).

One possibility for overcoming restrictions on effective resolution due to J-couplings are so-called Pure Shift techniques. "Pure Shift" refers to spectra in which multiplets have been reduced to singlets. The Pure Shift description has often been used as a synonym for homonuclear decoupling but, strictly speaking, the production of a true Pure Shift spectrum requires the removal of any multiplicity, homonuclear or heteronuclear. Removing the effects of heteronuclear coupling due to protons is routine, but eliminating the effects of proton-proton homonuclear coupling has proved to be one of the hardest problems in NMR history (or at least one of the longest running). Only recently have practical solutions that consistently deliver good results been reported. [4,5] Among the latest Pure Shift techniques, PSYCHE [6] seems to be particularly well suited for homonuclear experiments as its sensitivity is largely independent of the bandwidth to be decoupled, but techniques based on the elegant Zangger-Sterk method, as well as BIRD-based ones, are also useful, [7–10] the latter particularly for producing real-time Pure Shift HSQC experiments. [11,12] Using Pure Shift techniques, it is possible produce spectra where the effective “linewidth” for a particular signal in both dimensions is a few Hz or less. This holds out the prospect of greatly increased dispersion in the spectrum, but raises significant potential problems in terms of achieving the necessary digital resolution.

A possible solution to the digital resolution problem lies in the use of "Compressive Sensing". Compressive Sensing [13–15] is a mathematical body of work that states that when signals are sparse or compressive, objects such as spectra, [16] images, etc. can be reconstructed from fewer data points than the Shannon-Nyquist theorem [17,18] requires by collecting a set of incoherent measurements and then reconstructing the object using an algorithm that minimises the ℓ_1-norm of the object. [19] In such cases, the number of data points (m) needed to properly reconstruct the object is given by:

$$m \geq \mu . C . S . log(n_i)$$

where μ represents the coherence of the sampling method, C is a (small) constant, S describes an S-sparse signal matrix, and n_i is the size of the object (spectrum) to be reconstructed. S is the number of coefficients necessary to describe the relevant features (signals) in a compressed representation.

If we assume Lorentzian line shapes and infinite (or at least very good) signal to noise ratio, then each signal in a one-dimensional spectrum can be described by four parameters (frequency, intensity, line width, and phase). This would reduce to 3 parameters for a phase-insensitive representation. The total value for *S* in one dimension of a phase-sensitive two-dimensional spectrum is therefore approximately given by the total number of features (peaks) to be reconstructed times four (number of parameters per "peak"). There are two features of this equation that are immediately relevant to the problem at hand. The first is that the number of incoherently sampled points required (m) scales as the log of the size of the object (spectrum) to be reconstructed, so that the method is ideally suited to large objects. For example, reconstructing an 8192 increment two-dimensional spectrum using Compressive Sensing requires the acquisition of only 44 % more increments than the reconstruction of a 512 increment spectrum. (Eq. 2)

$$log(512 * 16)/log(512) = 1.44$$

In contrast, increasing the digital resolution in the indirectly detected dimension by a factor of 16 using classical data acquisition methods would require a 1500% increase in experiment time. The second point of interest is that the value of *S* decreases if there are fewer lines in the spectrum, so Pure Shift spectra can be reconstructed from fewer samples than their fully coupled analogues. This means that the combination of Pure Shift techniques and Compressive Sensing is synergic; Pure Shift techniques allows Compressive Sensing to work more effectively by reducing the number of lines in the spectrum, thereby reducing the value of *S*, while Compressive Sensing opens the way to the levels of digital resolution required to take full advantage of the increase in spectral dispersion offered by Pure Shift techniques. Finally, the increased digital resolution in F1 allows a safer use of covariance, as resolved signals in F1 avoid the production of spurious correlations.

Determining the absolute minimum number of increments required (m) is not straightforward and in reality is probably not appropriate on a sample by sample basis (anymore than determining the optimum repetition rate is appropriate on a sample by sample basis). However, as rough guidance, if the data is sampled in a fully incoherent manner, the value of μ approaches 1, while the value of *C*

is generally considered to be less than 1 (values around 0.3 are typically quoted). So assigning $\mu .C$ a value of 1 is a conservative approach. The minimum value of m is then surprisingly low, depending on the value of S. This assumes adequate signal to noise ratio (S/N) for reconstruction. This refers to the signal to noise ratio for the entire data set, so for weak samples we have the choice of either acquiring close to the minimum number of increments at high S/N by acquiring more transients, or acquiring a greater number of increments at more moderate S/N (assuming that T_2^* effects are not too severe, as is usually the case for small molecules). We have typically chosen to acquire the minimum number of transients consistent with phase-sensitive two dimensional spectra with suppression of axial artefacts (2 scans), but acquire more increments (typically 512) than equation (1) would suggest was the minimum necessary as a way of increasing signal to noise. We have found that these parameters deliver reliably robust results at very high effective resolution for a wide range of samples; we have run more than 100 different samples using these conditions in our NMR service and all can be considered as "real world" samples rather than carefully selected model samples.

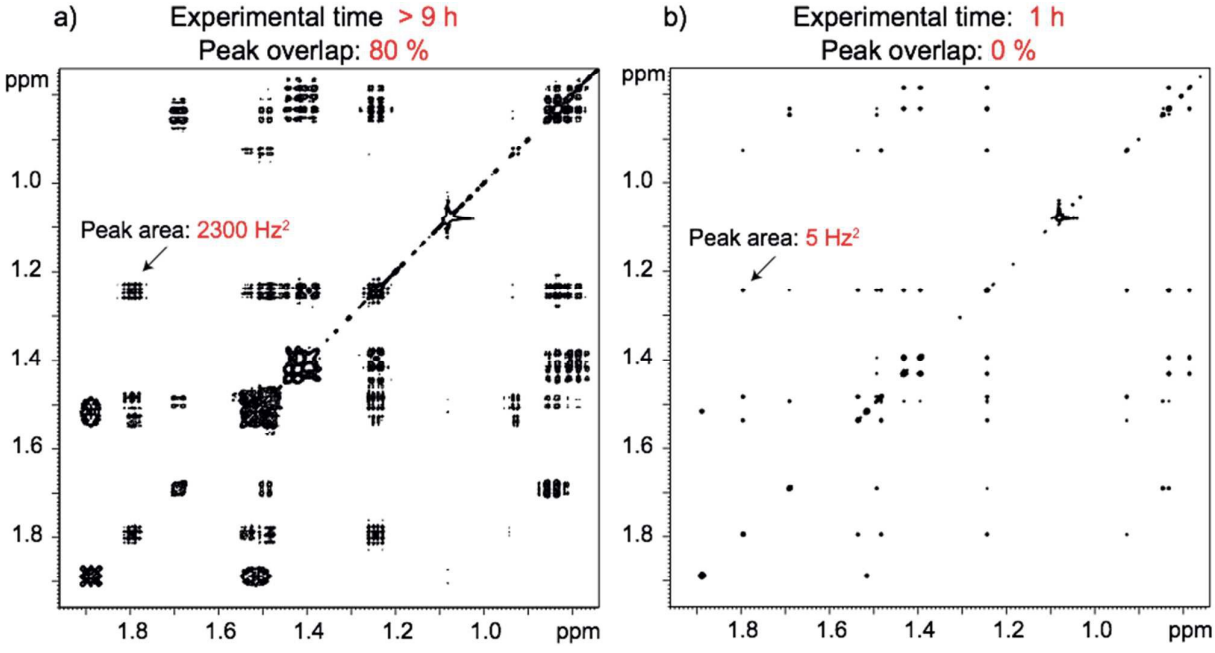


Figure 1. The alkyl regions of spectra of the product mixture produced by an attempted tosylation of testosterone. On the left, a conventional TOCSY was acquired to using a large number of

increments. More than 9 hours of spectrometer time were used to collect it. In spite of this, 80 % of the peaks overlap with at least one other peak. In contrast, on the right, the combination of Pure Shift NMR and Compressive Sensing is able to produce two-dimensional correlation peaks that occupy $1/500^{th}$ of the area in just $1/8^{th}$ of the time, with no peak overlap. Some peaks visible in (b) are not visible in (a), but they are actually present in the latter although they are too small to be seen at the threshold used. Spectrum (b) was acquired using a F1-decoupled PSYCHE-TOCSY; the results were covariance processed.

A good example of the utility of the combined techniques is shown in Figure 1. We were asked to try to analyse the product(s) of the tosylation of testosterone, which was problematic due to extensive peak overlap. The figure shows the stark contrast between a conventionally acquired TOCSY experiment, and the corresponding Pure Shift analogue acquired using Compressive Sensing (HD-2D). The additional information accessible in the HD-2D spectrum is immediately obvious and allowed the ready identification of components in the mixture. In this case, purely for comparison, the conventionally acquired spectrum was acquired to the same digital resolution in both dimensions as the HD-2D spectrum and therefore took considerably longer to run, but it should be noted that even if the conventional TOCSY had been acquired using Compressive Sensing (thereby reducing the total experiment time considerably), the information content would still have been inferior because the area of a typical cross-peak is measured in hundreds (possibly thousands) of Hz^2, while the corresponding cross peak in the HD-2D spectrum has an area of < 5 Hz^2. In the region of the conventional spectrum shown, 80% of the peaks have at least partial overlap. In the HD-2D analogue, there is no peak overlap.

Following the discussion above, the HD-2D spectrum (Figure 1(b)) was recorded using 512 incoherent complex increments at 2 scans per increment, making a total of 2048 individual FID's recorded. The next step is to determine how many increments are needed to achieve a digital resolution of about 1 Hz in the indirect dimension, considering that signals are rarely narrower than this (disregarding multiplicity). In our case around 4096 linear increments (classical acquisition

method) would be necessary to digitize a 6 KHz window to around that level, so the 512 incoherently sampled increments we collected needed to be reconstructed to the equivalent of 4096 linear increments. These conditions were used to produce the covariance [20,21] processed PSYCHE-TOCSY [22] spectrum shown in Figure 1b. Similar conditions and considerations have been used to produce other HD-2D NMR spectra. Incoherent sampling was implemented using a Poisson distribution as a means of avoiding large gaps in the sampling schedule. [23] For the ℓ_1-norm based reconstruction method, we have used *Iteractive Soft Thresolding (IST)* [23–25] but other algorithms, based on different principles, have also been suggested. [26–28] In addition to *IST*, we tried a greedy algorithm (*CLEAN*) as an example of the latter. [29] In all the examples we tried *IST* seems to produce better spectra than *CLEAN*, although *CLEAN* seems to require less computing time.

The utility of the combined techniques is not limited to homonuclear experiments. In heteronuclear correlation experiments such as HSQC, homonuclear decoupling in the proton dimension still eliminates the multiplicity, and therefore increases the effective resolution in the final HD-2D spectrum. At the same time, Compressive Sensing can greatly increase the digital resolution in the indirect (typically carbon) dimension in the same amount of experiment time. An example of heteronuclear HD-2D is shown in Figure 2, which shows spectra obtained from a mixture of isomers of an organometallic complex. In this case, both the conventional and the HD-2D experiment took approximately 1 hour each to run, but the effective resolution is improved by an order of magnitude in each dimension of the HD-2D spectrum. The impact this has can be appreciated by comparing the insets on the two-dimensional spectra, and by comparing the one-dimensional projections.

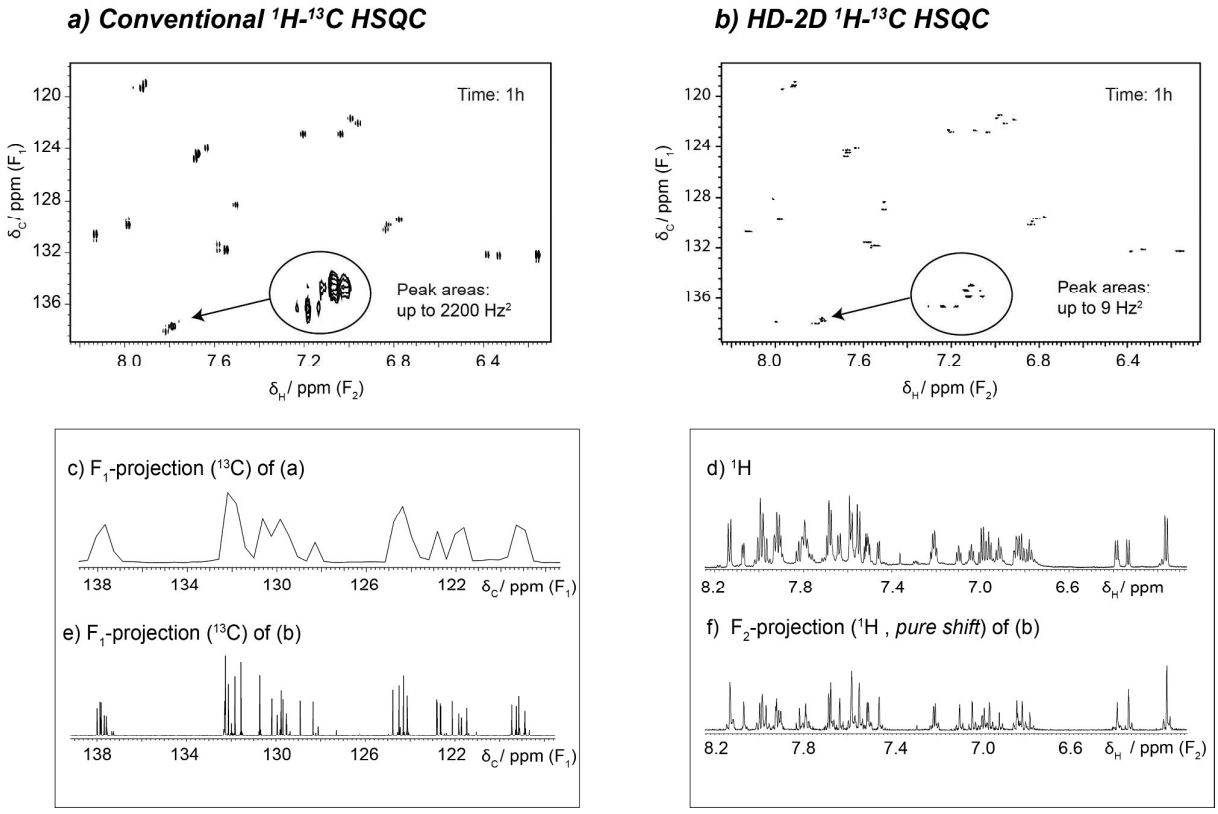


Figure 2. In (a) a 1024 increment ^{1}H-^{13}C HSQC is shown. It takes 1 h to acquire and shows peak areas up to 2200 Hz^2. In (b), where a *Compressed* NMR version has been used, peaks are reduced to a maximum of 9 Hz^2. Here Compressive Sensing has facilitated a high digital resolution in the carbon dimension. Under a classical data acquisition regime the carbon digital resolution in (b) would have taken 17 hours to acquire, whereas here it took just 1 hour. Compare the carbon projections of the conventional HSQC (c) with that of the compressed one (d). In addition, Pure Shift NMR has eliminated the multiplet limitation of the proton dimension increasing the achievable resolving power and the sparseness of the spectrum. Compare the proton spectrum (e) with the proton projection of the compressed HSQC (f). In (f) some signals show sidebands produced because large chunks were used to produce the real-time Pure Shift data. Reducing the size of the chunks by a factor of 3 would effectively eliminate the sidebands.

Experimental

The zTOCSY of Figure 1a was produced using a 600 MHz Varian spectrometer equipped with an Agilent OneNMR Probe able to deliver a maximum pulsed field gradient of 62 G cm^{-1}. Two scans per increment were collected each comprising 8192 complex data points and a spectral width of 6 kHz. 4096 increments were acquired. The repetition time was 1.7 s, of which 0.7 s comprised the acquisition time. Zero-quantum artefacts were attenuated as previously described. [30] An 80 ms DIPSI2 mixing composite pulse was used. [31] 128 dummy scans were used. The total experimental time was 9 h 40 min. The compressed version was produced in 1 h and 7 min, using the same experimental conditions but introducing the following modifications: i) the zTOCSY pulse sequence was replaced by the PSYCHE-zTOCSY sequence. The PSYCHE pulse was created using a 10° WURST180 double sweep of 30 ms. [32] A pulsed field gradient (0.8 G $cm^{-1)}$ was kept on for the whole duration of the adiabatic pulse. ii) The traditional sampling schedule (4096 increments) was replaced with a random one based on a Poisson distribution of 512 increments. After reconstruction, this equates to a 4096 increment spectrum. In the case of Figure 1, covariance was used after reconstructing the spectrum, to produce the doubly decoupled spectrum shown. VNMRJ version 4.2 (Agilent) was used to acquire and reconstruct the spectra. The reconstructed spectra were then covariance processed using TopSpin 2.1 (Bruker). Details of the reconstruction methods and parameters used can be found in the Supplementary information section.

Figure 2 was produced using a 700 MHz Varian spectrometer equipped with an Agilent OneNMR Probe able to deliver a maximum pulsed field gradient of 62 G cm^{-1}. In Figure 2, an HSQC was acquired using two scans per increment, each comprising 2010 complex data points. 128 dummy scans were used. The spectral width of the proton dimension spanned 3.5 kHz. The spectral width of the carbon dimension spanned 36 kHz. The repetition time was 1.3 s, of which 0.3 s comprised the acquisition time. Compressive Sensing (a Poisson distribution of 640 randomly distributed increments) was used to acquire a data set that, after, reconstruction yielded a data set equivalent to 10240 increments. This would have required 17 h if it had been acquired conventionally, but Compressive Sensing made it possible to acquire the data in just 1h. In order to simulate what the

resolution would have been had just one hour of conventional data been acquired, only the first 1024 increments of the reconstructed experiments were processed to generate Figure 2a. Figure 2b was produced using the same conditions, but using all the available increments and replacing the conventional HSQC with a real-time Pure Shift version. [11] The real-time acquisition collected 96 ms long blocks of data, apart from the first and the last blocks acquired which were half that duration. Some sidebands can be seen around isolated signals of the proton dimension due to the fact that the data blocks are long. These sidebands can be reduced by acquiring blocks of data of a third of the duration used here. The raw data can be found in the supporting information section. VNMRJ version 4.2 (Agilent) was used to acquire and reconstruct the spectra. Details of the reconstruction methods and parameters used can be found in the Supplementary information section.

The data sets containing the raw data used in this paper can be found in doi:10.15128/8g84mm259

Conclusions

In two-dimensional spectra of small molecules, Pure Shift techniques and Compressive Sensing are synergic and their combined use can routinely provide two-dimensional spectra with effective resolution that is orders of magnitude higher than spectra from conventionally acquired analogous experiments that take the same amount of spectrometer time. The use of this synergic combination (Compressed NMR) allows us to address problems that would be extremely problematic otherwise.

Acknowledgements

The authors thank Prof. Andrew Beeby and Drs Steven Cobbs, Ross Davidson and Maria Czyzewska for providing the samples used in the Figures, and for permission to use the results. The authors also thank the University of Durham for support.

References

1.S. Aime, M. Botta, and G. Ermondi, *Inorganic Chemistry*, 1992, **31**, 4291–4299.

2.J. A. Aguilar, P. Király, R. W. Adams, M. Bonneau, E. J. Grayson, M. Nilsson, A. M. Kenwright, and G. A. Morris, *RSC Advances*, 2015, **5**, 52902–52906.

3.J. Saurí, T. Parella, and R. T. Williamson, *Magnetic Resonance in Chemistry*, 2015. DOI: 10.1002/mrc.4322

4.L. Castañar, R. Roldán, P. Clapés, A. Virgili, and T. Parella, *Chemistry - A European Journal*, 2015, **21**, 7682–7685.

5.K. Zangger, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2015, **86-87**, 1–20.

6.M. Foroozandeh, R. W. Adams, N. J. Meharry, D. Jeannerat, M. Nilsson, and G. A. Morris, *Angewandte Chemie International Edition*, 2014, **53**, 6990–6992.

7.Lokesh and N. Suryaprakash, *Chemical Communications*, 2014, **50**, 8550.

8.J. Mauhart, S. Glanzer, P. Sakhaii, W. Bermel, and K. Zangger, *Journal of Magnetic Resonance*, 2015, **259**, 207–215.

9.P. Sakhaii, B. Haase, W. Bermel, R. Kerssebaum, G. E. Wagner, and K. Zangger, *Journal of Magnetic Resonance*, 2013, **233**, 92–95.

10.J. A. Aguilar, M. Nilsson, and G. A. Morris, *Angewandte Chemie International Edition*, 2011, **50**, 9716–9717.

11.L. Paudel, R. W. Adams, P. Király, J. A. Aguilar, M. Foroozandeh, M. J. Cliff, M. Nilsson, P. Sándor, J. P. Waltho, and G. A. Morris, *Angewandte Chemie International Edition*, 2013, **52**, 11616–11619.

12.K. J. Donovan and L. Frydman, *Angewandte Chemie International Edition*, 2014, **54**, 594-598.

13.D. L. Donoho, *Information Theory, IEEE Transactions on*, 2006, **52**, 1289–1306.

14.E. J. Candes and M. B. Wakin, *IEEE Signal Processing Magazine*, 2008, **25**, 21–30.

15.E. J. Candès, J. Romberg, and T. Tao, *Information Theory, IEEE Transactions on*, 2006, **52**, 489–509.

16.K. Kazimierczuk and V. Y. Orekhov, *Angewandte Chemie International Edition*, 2011, **50**, 5556–5559.

17.C. E. Shannon, *Proceedings of the IEEE*, 1998, **86**, 447–457.

18.H. Nyquist, *Proceedings of the IEEE*, 2002, **90**, 280–305.

19.E. Candès and J. Romberg, *Inverse problems*, 2007, **23**, 969–985.

20.R. Brüschweiler and F. Zhang, *The Journal of Chemical Physics*, 2004, **120**, 5253.

21.G. A. Morris, J. A. Aguilar, R. Evans, S. Haiber, and M. Nilsson, *Journal of the American Chemical Society*, 2010, **132**, 12770–12772.

22.M. Foroozandeh, R. W. Adams, M. Nilsson, and G. A. Morris, *Journal of the American Chemical Society*, 2014, **136**, 11867–11869.

23.S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, and G. Wagner, *Journal of Biomolecular NMR*, 2012, **52**, 315–327.

24.A. S. Stern, D. L. Donoho, and J. C. Hoch, *Journal of Magnetic Resonance*, 2007, **188**, 295–300.

25.W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *SIAM Journal on Imaging Sciences*, 2008, **1**, 143–168.

26.R. Chartrand, *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, 2009, 262–265.

27.S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, *Journal of Communications and Networks*, 2013, **15**, 443–456.

28.K. Kazimierczuk and V. Y. Orekhov, *Journal of Magnetic Resonance*, 2012, **223**, 1–10.

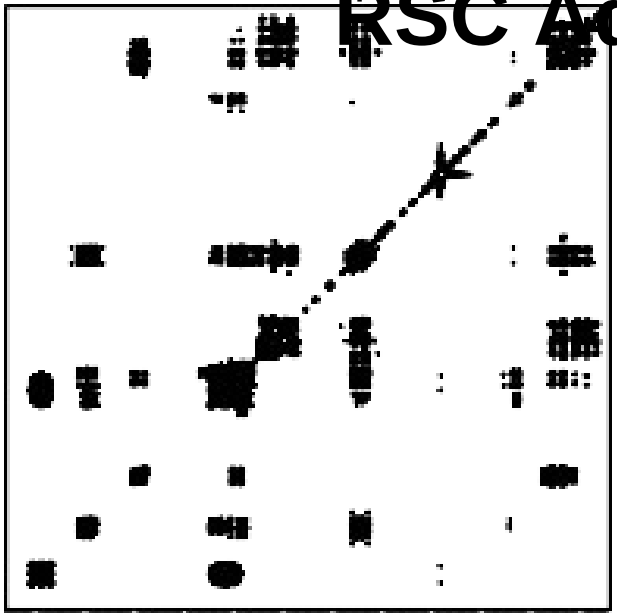
29.E. Kupče and R. Freeman, *Journal of Magnetic Resonance*, 2005, **173**, 317–321.

30.M. Thrippleton and J. Keeler, *Angewandte Chemie International Edition*, 2003, **42**, 3938–3941.

31.A. J. Shaka, C. J. Lee, and A. Pines, *Journal of Magnetic Resonance*, 1988, **77**, 274–293.

32.E. Kupče and R. Freeman, *Journal of Magnetic Resonance*, 1995, **115**, 273–276.

Experimental time > 9 h
Peak overlap: 80 %



Experimental time: 1 h
Peak overlap: 0 %