NMR Spectroscopy

DOI: 10.1002/anie.201100440

Fast Multidimensional NMR Spectroscopy Using Compressed Sensing**

*Daniel J. Holland, Mark J. Bostock, Lynn F. Gladden, and Daniel Nietlispach**

Multidimensional NMR spectroscopic techniques form part of the standard tool kit for the structure determination of macromolecules[1–3] and, increasingly, solid-state materials, such as catalysts.[4,5] Advances in NMR spectroscopy have extended its applicability to large biomolecules,[6] including protein complexes[7,8] and membrane proteins.[1,9,10] NMR spectroscopy is also a key tool in structural-genomics projects.[11]

Such NMR spectroscopic studies rely on 3D or even higher-dimensional techniques. Limiting for all these studies are the large demands on time for the realization of sufficient experimental sensitivity and resolution. For an n-dimensional experiment processed using discrete Fourier transform (DFT), data points on an $(n-1)$-dimensional grid must be sampled, which leads to prohibitively long experiment times. Consequently, many systems remain beyond the reach of NMR spectroscopy, including, for example, the important family of mammalian G-protein-coupled receptors (GPCRs). There is a pressing need to develop new techniques to enable more efficient data acquisition and processing.

Herein we describe the use of compressed sensing (CS)[12,13] as an alternative reconstruction technique for multidimensional NMR spectroscopy. In an initial step, representative of a situation in which the signal-to-noise (S/N) ratio is high, we quantify the accuracy of the technique when applied to multidimensional NMR spectroscopy with 2D [^{1}H,^{15}N]-HSQC data recorded for the model protein ubiquitin. Subsequently, we demonstrate the performance of the technique on nonuniformly sampled (NUS) 3D HNCA and HN(CO)CA data sets of pSRII, a large transmembrane protein, for which high resolution is required but, owing to slow molecular tumbling, the S/N ratio is low. However, the generality of the CS method means it could also be applied to other undersampled multidimensional experiments used, for example, in solid-state NMR spectroscopy,[2,14] fast metabolomics studies,[15] and chemical-shift imaging.[16]

For decaying data, sparse or nonuniform sampling ensures that the majority of the experiment time is used to sample the signal at low evolution times at which the S/N ratio is highest, to maximize sensitivity, while increasingly sparse sampling at long evolution times, at which the S/N ratio is low, ensures that high resolution is obtained.[17,18] A number of methods have been proposed for the reconstruction of spectra from NUS data.[14,17–21] Perhaps the most established is maximum entropy (ME) reconstruction.[17,18] Other methods of speeding up multidimensional NMR spectroscopy are available.[22–26] However, many of these advanced methods assume a high S/N ratio or require specialist equipment, which limits their applicability.[27] ME has recently been used to reconstruct multidimensional spectra of large proteins, such as the T-TE fragment of EntF[28] and pSRII.[10] We show that CS reconstruction compares favorably with ME.

CS[12,13] is rapidly gaining acceptance in imaging applications[16,29–31] and has also been proposed as a tool for the reconstruction of undersampled, multidimensional NMR spectra;[32] however, to our knowledge, the results presented herein are the first demonstration of the application of CS to the reconstruction of real 3D NMR spectra. The application of CS requires 1) sparse representation of the desired signal in a particular basis and 2) incoherent sampling with respect to that basis. Importantly, an initial estimation of the signal location is not required. Many NMR spectra are ideally suited to CS reconstruction, as they consist of relatively few isolated peaks in the Fourier basis, and therefore are inherently sparse. Furthermore, measurements are performed in the time domain, which has been shown to be optimally incoherent with respect to the Fourier basis.[12]

A large reduction in experiment time is possible through the use of CS. By the application of an exponentially weighted sampling scheme, CS reconstruction of a 2D [^{1}H,^{15}N]-HSQC spectrum of ubiquitin was performed with only 30% of the total 128 complex data pairs recorded in the ^{15}N dimension for DFT processing. This approach reduced the experiment time from 165 min to 50 min (number of scans, ns = 32; see Methods in the Supporting Information). The spectra in Figure 1 resulting from reconstruction by the two methods are almost identical, with all peaks clearly resolved (for the full amide region, see Figure S1 in the Supporting Information). Close-up views of several regions of the spectra (see Figure S1) show that all contour levels for the CS reconstruction lie almost exactly on top of the contour levels for the fully sampled spectrum. Thus, the peak positions as well as the shape and intensity of the peaks are accurately recovered by the use of CS.

To quantify the accuracy of the reconstruction, we calculated the root-mean-squared (RMS) error for the peak positions to be 0.02 ppm and that for the intensities of these peaks to be 3% (see Figure S1). Furthermore, to demonstrate

[*] M. J. Bostock, Dr. D. Nietlispach
Department of Biochemistry, University of Cambridge
80 Tennis Court Road, Cambridge, CB2 1GA (UK)
E-mail: dn206@cam.ac.uk
Dr. D. J. Holland, Prof. Dr. L. F. Gladden
Department of Chemical Engineering and Biotechnology
University of Cambridge, Cambridge, CB2 3RA (UK)

[**] This research was supported by a BBSRC PhD studentship (M.J.B.) and Microsoft External Research (D.J.H.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201100440.

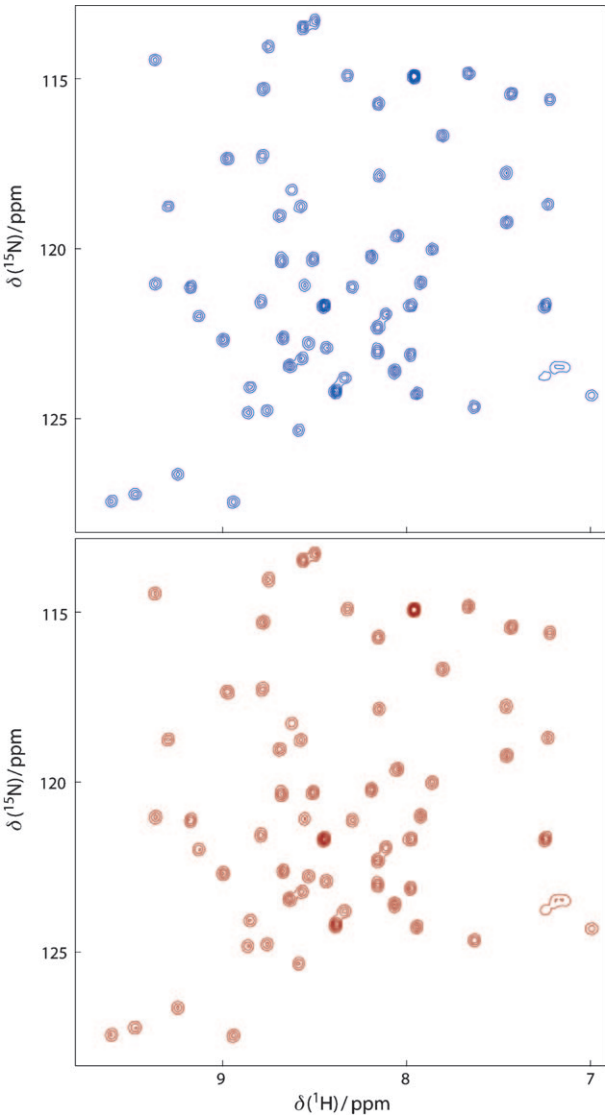

Figure 1. Comparison of an undersampled CS reconstruction (blue) of a [^{1}H,^{15}N]-HSQC spectrum of ubiquitin with the corresponding fully sampled DFT spectrum (red). The CS reconstruction was performed with only 30% of the 128 complex data pairs required to obtain the fully sampled spectrum (see Figure S1 in the Supporting Information for the entire amide region).

the robustness to noise, we added random Gaussian noise to the time-domain data so that the S/N ratio of the fully sampled spectrum was decreased by an order of magnitude (from 150 to 13), as calculated on the basis of the mean peak signal intensity and the standard deviation of the noise. Despite the lower S/N ratio, equivalent to that for a putative experiment time of approximately 30 s, a high-quality undersampled CS reconstruction was obtained (see Figure S2). The RMS error for the peak position of this noisier spectrum increased to only 0.05 ppm. All backbone NH resonances were detected, and only six weak peaks of Arg side-chain $N^{\varepsilon}H^{\varepsilon}$ multiplets with intensities less than three times the noise standard deviation could not be recovered. These initial results demonstrate that CS processing combined with undersampling can result in substantial time saving, and that CS is robust in situations with less favorable S/N ratios (see Figure S3).

Substantially greater sampling reductions are possible with higher-dimensional data sets through undersampling in multiple indirect dimensions. This method was initially tested on the 2D data shown in Figure 1 by artificially removing data points in the direct acquisition dimension and enabled a spectrum of the same quality as that in Figure 1 to be reconstructed from only 10% of the full data matrix (see Figure S4 and Methods in the Supporting Information). Although this approach has no immediate experimental benefit, it indicates the degree of undersampling that should be possible in two indirect dimensions. Consequently, CS was applied to the more practical case of NUS 3D HNCA and HN(CO)CA data sets recorded as [^{1}H,^{15}N]-TROSY implementations with u-[^{2}H,^{15}N,^{13}C]-labeled pSRII, a 241-residue, seven-transmembrane α-helical protein from the archaebacterium *Natronomonas pharaonis* whose structure was previously determined by NMR spectroscopy.[10,33] With a molecular weight of about 70 kDa for the detergent-solubilized protein complex the situation is representative of proteins for which NUS can be highly beneficial both to improve the resolution in a given experiment time and, in the presence of decaying evolution periods, to increase sensitivity. NUS spectra were recorded for 22.8% of the full data matrix, which corresponds to 350 ^{13}C,^{15}N complex pairs. A 2D NUS scheme suitable for a nondecaying constant time period (^{15}N) and a decaying evolution period (^{13}C) was used.[17,18] For the HNCA and HN(CO)CA experiments, the total recording times were 88 (ns = 184) and 132 h (ns = 272), respectively. Time constraints meant that the recording of a full matrix of 48×32 points for the ^{13}C and ^{15}N dimensions was impractical and therefore prevented a comparison with DFT processing. The performance of ME for the processing of NUS 3D data has previously been successfully validated against the use of DFT processing.[17] Therefore, in the absence of a DFT spectrum, the data processed by ME were used as a standard against which to compare the CS method.

Selected strip plots of the reconstructed NUS 3D HNCA and HN(CO)CA spectra shown in Figure 2 illustrate the data quality attainable by the use of CS. High-quality CS data sets were recovered that compare favorably with the corresponding spectra processed by ME (Figure 2). The CS and ME reconstructions of both the HNCA and the HN(CO)CA data are similar for high-intensity peaks (i peaks in HNCA and $i-1$ peaks in HN(CO)CA). No systematic or substantial shift deviations were noticed between the two processing methods, and the spectra derived by both methods were free of systematic signal artifacts. Thus, a successful sequential reassignment of the protein backbone was possible from the CS spectra. Sequential connections between i and $i-1$ residues, representative of the backbone-assignment procedure, are indicated by horizontal arrows.

A quantitative evaluation of nonlinear reconstruction methods is difficult because S/N ratios based on spectral noise have only limited meaning. However, to obtain a more thorough comparison of the CS and ME processing methods, we adjusted the spectra to a minimum contour level at which weak peaks could still be unambiguously discriminated from incoherent artifacts. This threshold, T, provided a qualitative lower limit at which peaks were still readily recognized during

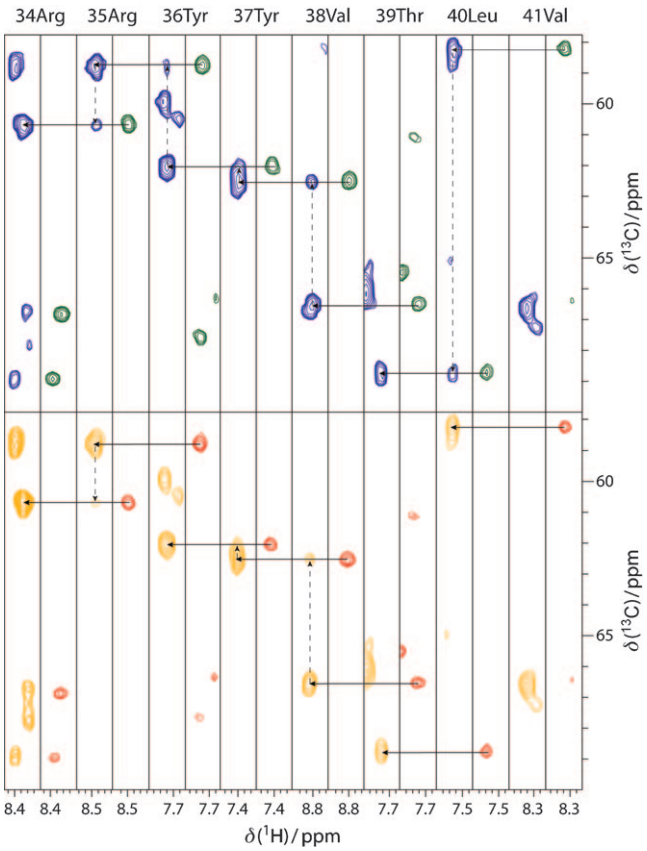

Figure 2. Strip plots of the CS (top) and ME (bottom) reconstructions of the 3D HNCA (CS blue, ME orange) and HN(CO)CA (CS green, ME red) spectra of the protein pSRII. The horizontal arrows indicate sequential assignments of eight residues along the protein backbone. Vertical arrows indicate assignments that could have been derived from the HNCA spectrum directly. Both CS and ME reconstructions were performed with 22.8% of a fully sampled data matrix of 1536 complex pairs, that is, with 350 $^{13}C,^{15}N$ complex pairs.

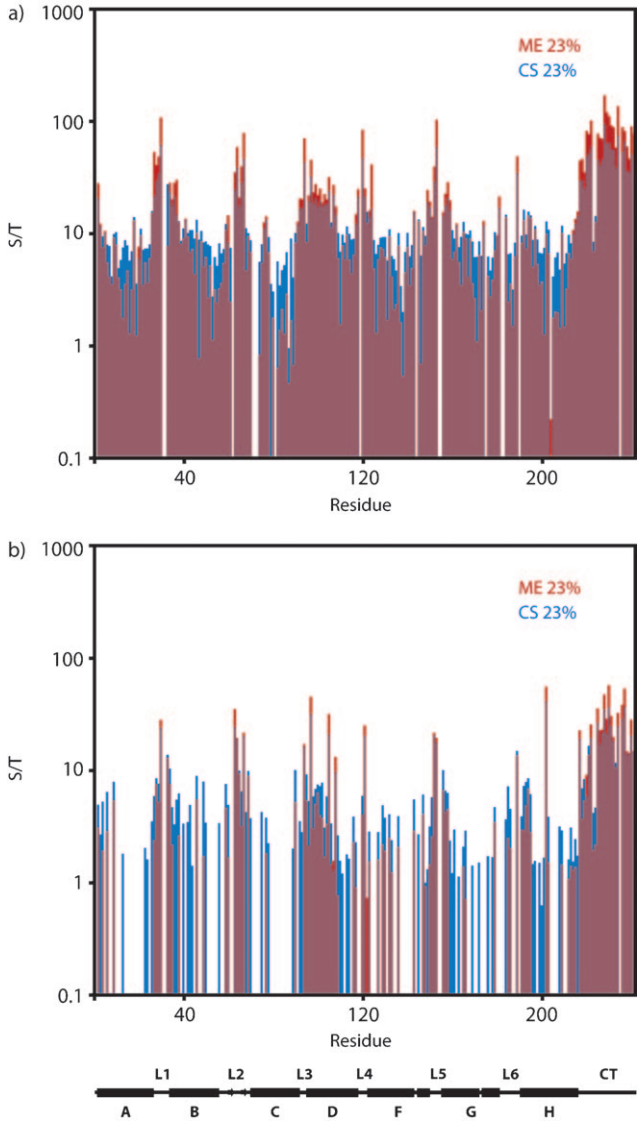

Figure 3. Comparison of peak detectability, expressed as S/T (see text), in a 3D HNCA spectrum processed by CS (blue) and ME (red) for a) the intraresidue *i* correlations and b) the sequential interresidue $i-1$ connectivities of pSRII. Regions in which the CS and ME graphs are overlaid appear in purple. ME and CS reconstructions were performed with 22.8% of the full data matrix of 1536 complex points, that is, with 350 $^{13}C,^{15}N$ complex pairs. In most instances, CS recovered the peaks with higher detectability than ME, as is particularly noticeable in regions in which the peak detectability was low. The blue bars indicate regions in which the peak detectability of CS was greater than that of ME. The pSRII secondary structure is shown below the graphs with α helices represented as bars and β-sheet regions as arrows.

the backbone-assignment procedure. For each reconstruction, the detectability of a peak was then expressed as the ratio of the signal intensity to the threshold, S/T. S/T values for the *i* (Figure 3a) and $i-1$ peaks (Figure 3b) were obtained from an HNCA data set reconstructed by using both ME and CS. In agreement with previous observations, ME boosted peaks that were inherently strong (Figure 3a), for example, those in flexible loops connecting the helices and in the unstructured C-terminal tail. In comparison, the use of CS led to a higher peak detectability (S/T) in regions which are generally less intense, for example, in the well-structured regions of the protein backbone; in these regions, ME processing tends to produce peaks with smaller intensities. For the HN(CO)CA spectrum, the differences are smaller; the measured S/T values remain slightly higher for CS, although the data sets are comparable (see Figure S5). Noticeably, for the weakest group of backbone correlations, the $i-1$ peaks in the HNCA spectrum, CS reconstruction performed markedly better: many peaks were detected that were either weaker or absent in the spectrum reconstructed by ME (Figure 3b). A total of 104 $i-1$ peaks were assigned in the HNCA spectrum reconstructed by ME, in comparison with 145 in the HNCA spectrum reconstructed by CS. Vertical arrows in Figure 2 highlight sequential assignments which could have been conducted by using the HNCA $i-1$ peaks instead of the HN(CO)CA resonances. Although the $i-1$ peaks in an HNCA spectrum are rarely used for assignment purposes, the increased recovery of these generally weaker peaks by CS indicates that data can be processed over a larger dynamic range by CS than by ME.

CS reconstructions of the HNCA and HN(CO)CA spectra were also performed with only 16.1% (247 $^{13}C,^{15}N$

complex pairs) of the full data sets (see Figures S6 and S7) and resulted in recording times of 62 h for the HNCA experiment and 93 h for the HN(CO)CA experiment. Thus, further significant reductions in experiment time of 26 and 39 h, respectively, were possible with respect to the 22.8 % data sets (see Methods in the Supporting Information). In both cases, CS continued to give good results that were comparable in quality with the spectra reconstructed by ME with 22.8 % of the data. Indeed, in the 16.1 % sampled HNCA spectrum processed by CS, 103 $i-1$ peaks were identified: a result comparable to the 104 $i-1$ peaks obtained by the ME reconstruction of 22.8 % of the HNCA data (see Figure S7b).

On the basis of these preliminary results, we propose CS as a powerful method for the reconstruction of undersampled data from multidimensional NMR experiments in a variety of biological and nonbiological contexts. The robustness of CS with respect to noise will enable reduced experiment times for systems with poor S/N ratios (see Figures S3 and S8). Furthermore, the observation that CS can detect the weaker $i-1$ peaks in an HNCA spectrum suggests that this method may be effective for data sets in which the large dynamic range of the signals causes difficulties for other methods, such as ME. We are actively pursuing this area of research with more challenging experiments, such as NOESY spectra with diagonal and cross-peaks of very different intensities. Alternatively, if the dynamic range is less critical, CS can be used to further undersample the data. In the case of 3D measurements of pSRII, good-quality CS reconstructions enabling backbone assignment could be obtained from only 16 % of the data, with a further reduction in experiment time of 30 % relative to the ME reconstruction of spectra of comparable quality. Such additional time savings could be particularly important for inherently unstable samples, for example, GPCRs. Furthermore, the generality of CS should enable the reconstruction technique to be combined with other advanced NMR data-sampling procedures, such as projection–reconstruction[26] or SOFAST[34] NMR acquisition schemes, to further reduce the data-recording time. Consequently, we expect that CS will find many applications in the field of NMR spectroscopy.

Experimental Section

NMR spectroscopic experiments were performed at 25 °C on a Bruker DRX500 spectrometer (ubiquitin) and at 50 °C on a Bruker DRX600 spectrometer (pSRII). TROSY 3D backbone experiments (HNCA, HN(CO)CA) were recorded by using nonuniform sampling schemes generated with the COAST (Configuring Acquisition Sampling Tables) program.[17] Approximate sample conditions: 0.5 mM pSRII, pH 5.9, 60 mM c7-DHPC detergent. Full experimental details are given in the Supporting Information.

Received: January 18, 2011
Revised: April 26, 2011
Published online: June 6, 2011

Keywords: compressed sensing · multidimensional NMR spectroscopy · nonuniform sampling · proteins · structure elucidation

[1] H. J. Kim, S. C. Howell, W. D. Van Horn, Y. H. Jeon, C. R. Sanders, *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 335 – 360.
[2] D. L. Laws, H. L. Bitter, A. Jerschow, *Angew. Chem.* **2002**, *114*, 3224 – 3259; *Angew. Chem. Int. Ed.* **2002**, *41*, 3096 – 3129.
[3] P. L. Rinaldi, *Analyst* **2004**, *127*, 689 – 699.
[4] N. Hedin, R. Graf, S. C. Christiansen, C. Gervais, R. C. Hayward, J. Eckert, B. F. Chmelka, *J. Am. Chem. Soc.* **2004**, *126*, 9425 – 9432.
[5] J. W. Wiench, C. E. Bronnimann, V. S.-Y Lin, M. Pruski, *J. Am. Chem. Soc.* **2007**, *129*, 12076 – 12077.
[6] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366 – 12371.
[7] J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *Nature* **2002**, *418*, 207 – 211.
[8] R. Sprangers, A. Velyvis, L. E. Kay, *Nat. Methods* **2007**, *4*, 697 – 703.
[9] S. Hiller, R. G. Garces, T. J. Malia, V. Y. Orekhov, M. Colombini, G. Wagner, *Science* **2008**, *321*, 1206 – 1210.
[10] A. Gautier, H. R. Mott, M. J. Bostock, J. P. Kirkpatrick, D. Nietlispach, *Nat. Struct. Mol. Biol.* **2010**, *17*, 768 – 774.
[11] T. Szyperski, D. C. Yeh, D. K. Sukumaran, H. N. B. Moseley, G. T. Montelione, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8009 – 8014.
[12] D. L. Donoho, *IEEE Trans. Inf. Theory* **2006**, *52*, 1289 – 1306.
[13] E. J. Candes, J. Romberg, T. Tao, *IEEE Trans. Inf. Theory* **2006**, *52*, 489 – 509.
[14] Y. Matsuki, M. T. Eddy, R. G. Griffin, J. Herzfeld, *Angew. Chem.* **2010**, *122*, 9401 – 9404; *Angew. Chem. Int. Ed.* **2010**, *49*, 9215 – 9218.
[15] S. G. Hyberts, G. J. Heffron, N. G. Tarragona, K. Solanky, K. A. Edmonds, H. Luithardt, J. Fejzo, M. Chorev, H. Aktas, K. Colson, K. H. Falchuk, J. A. Halperin, G. Wagner, *J. Am. Chem. Soc.* **2007**, *129*, 5108 – 5116.
[16] S. Hu, M. Lustig, A. P. Chen, J. Crane, A. Kerr, D. A. C. Kelley, R. Hurd, J. Kurhanewicz, S. J. Nelson, J. M. Pauly, D. B. Vigneron, *J. Magn. Reson.* **2008**, *192*, 258 – 264.
[17] D. Rovnyak, D. P. Frueh, M. Sastry, Z.-Y. J. Sun, A. S. Stern, J. C. Hoch, G. Wagner, *J. Magn. Reson.* **2004**, *170*, 15 – 21.
[18] J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling, S. J. P. Worrall, *J. Magn. Reson.* **1987**, *73*, 69 – 77.
[19] K. Kazimierczuk, W. Koźmiński, I. Zhukov, *J. Magn. Reson.* **2006**, *179*, 323 – 328.
[20] D. Marion, *J. Biomol. NMR* **2005**, *32*, 141 – 150.
[21] V. Y. Orekhov, I. V. Ibraghimov, M. Billeter, *J. Biomol. NMR* **2001**, *20*, 49 – 60.
[22] V. A. Mandelshtam, *J. Magn. Reson.* **2000**, *144*, 343 – 356.
[23] L. Frydman, T. Scherf, A. Lupulescu, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15858 – 15862.
[24] E. Kupče, T. Nishida, R. Freeman, *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *42*, 95 – 122.
[25] H. S. Atreya, T. Szyperski, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9642 – 9647.
[26] E. Kupče, R. Freeman, *J. Am. Chem. Soc.* **2004**, *126*, 6429 – 6440.
[27] R. Freeman, E. Kupče, *J. Biomol. NMR* **2003**, *27*, 101 – 114.
[28] D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, *Nature* **2008**, *454*, 903 – 906.
[29] D. J. Holland, D. M. Malioutov, A. Blake, A. J. Sederman, L. F. Gladden, *J. Magn. Reson.* **2010**, *203*, 236 – 246.
[30] M. Lustig, D. L. Donoho, J. M. Pauly, *Magn. Reson. Med.* **2007**, *58*, 1182 – 1195.
[31] R. Otazo, D. Kim, L. Axel, D. K. Sodickson, *Magn. Reson. Med.* **2010**, *64*, 767 – 776.
[32] I. Drori, *EURASIP J. Adv. Sig. Pr.* **2007**, 20248.
[33] A. Gautier, J. P. Kirkpatrick, D. Nietlispach, *Angew. Chem.* **2008**, *120*, 7407 – 7410; *Angew. Chem. Int. Ed.* **2008**, *47*, 7297 – 7300.
[34] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014 – 8015.

