VIP Solid-State NMR Spectroscopy

DOI: 10.1002/anie.201003329

Rapid Three-Dimensional MAS NMR Spectroscopy at Critical Sensitivity**

*Yoh Matsuki, Matthew T. Eddy, Robert G. Griffin, and Judith Herzfeld**

Solid-state magic angle spinning (MAS) NMR spectroscopy is maturing rapidly. Progress is indicated by recent examples that demonstrate its capability to yield site-specific assignments and atomic resolution structural information on fibrillar,[1–3] membrane-associated,[4–6] and non-crystalline proteins.[7–9] Furthermore, applications to systems of ever-increasing molecular size are limited only by the signal-to-noise ratio (S/N) in the multidimensional spectra required for adequate resolution, rather than by more fundamental limitations from spin relaxation. Therefore, the most pressing need in MAS NMR is arguably for more efficient data acquisition methods.

Acquisition problems in MAS NMR, relative to solution NMR, are twofold. First, ^{13}C detection is necessary to obtain the narrow linewidths required for site-specific assignments and structure determination; however, ^{13}C detection is inherently less sensitive than ^{1}H detection. Second, slower relaxation and the need for high-power ^{1}H decoupling in solids necessitate longer recycle delays. For these reasons, employing three or more dimensions in MAS NMR experiments has not yet become common practice.[7,10–14]

In solution NMR, more efficient acquisition has relied on non-uniform sampling (NUS), which has been successfully applied to multidimensional experiments on larger systems.[15–19] Although extending NUS to MAS NMR would be of enormous practical importance, the application of conventional NUS methods to MAS NMR has been limited by the specific problem of accurately modeling weak signals in noisy spectra,[20–22] in addition to the general problems of quantitative spectral reconstruction and slow computation.[23] The lower sensitivity in MAS NMR experiments requires an unprecedented robustness of any NUS method to minimize artifacts.

Herein, we address these challenges with SIFT (spectroscopy by integration of frequency and time domain information), a rapid and model-free method for computing a NMR spectrum from a NUS time-domain dataset.[24] SIFT works by replacing missing information in the time domain with a priori knowledge of "dark" regions in the frequency domain; that is, those regions known to contain no NMR signals. The frequency domain information, assimilated by a very rapid computational process, obviates some time-domain sampling with no sacrifice in resolution and no modeling bias.

We previously used SIFT to process 2D NUS ^{15}N-HSQC solution data, where dark regions created by over-sampling were utilized to replace up to 75% of the uniform time-domain data points.[24] Herein we demonstrate the effectiveness of the SIFT method in solids, using dark regions resulting from the need for rotor-synchronized sampling in the indirect dimensions. Unlike other NUS data processing methods that actively model signals to reconstruct a spectrum, SIFT suppresses the sampling noise by using only definitive information from the dark spectral areas. Thus, SIFT avoids bias from subjective discrimination between weak signals and noise, and reconstructs missing time data points with high fidelity, as if they had been actually recorded. These favorable properties make SIFT uniquely suited for processing NUS data in the sensitivity-limited regime.

To demonstrate the application of SIFT to NUS MAS NMR, we recorded a 3D NCOCX spectrum (Figure 1a) of a

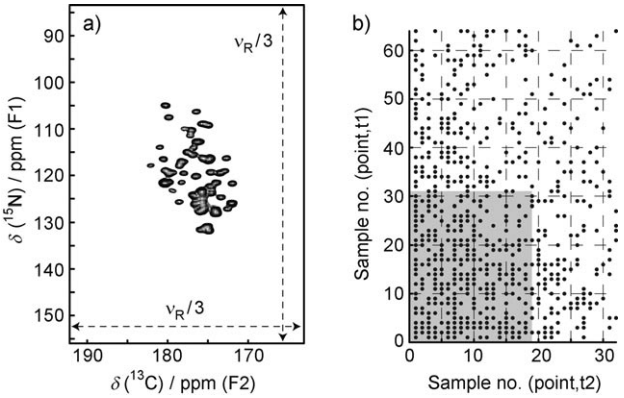

Figure 1. a) The projection down the direct acquisition axis (F3) of the NCOCX spectrum of GB1. The full spectral range is shown for both the ^{15}N (F1) and ^{13}C (F2) dimensions. b) The dots represent the NUS schedule comprising 608 quasi-random on-grid points with exponentially decaying probability density. The shaded area corresponds to a truncated set of 608 (32×19) uniform points. A full dataset with 2048 (64×32) samples was acquired as a reference and time samples were omitted according to the schedule to form the NUS dataset. Forty scans were averaged per t1/t2 sample. The NUS schedule was generated by the sampling scheduler available at http://sbtools.uchc.edu/nmr/sample_scheduler/. The dwell time was 270 μs for t1 and t2 at a MAS frequency of 11.1 kHz.

[*] Dr. Y. Matsuki, Prof. Dr. J. Herzfeld
Department of Chemistry, Brandeis University
Waltham MA 02454 (USA)
Fax: (+1) 781-736-2538
E-mail: herzfeld@brandeis.edu
Dr. Y. Matsuki, M. T. Eddy, Prof. Dr. R. G. Griffin
Department of Chemistry and Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology
Cambridge MA 02139 (USA)

[**] This research was supported by NIH grants EB001035, EB003151 and EB002026. Y.M. acknowledges partial financial support from the Naito Foundation.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201003329.

microcrystalline, uniformly [^{15}N,^{13}C]-labeled sample of the β1 domain of protein G (GB1) at high digital resolution (1.1 ppm for F1, 0.7 ppm for F2, before zero-filling). For both t1 and t2, the dwell time was synchronized to three times the rotor period, $3/\nu_R$ (bandwidth equal to $\nu_R/3$ Hz), in order to fold the spinning sidebands onto the corresponding centerbands. The spinning frequency ν_R was chosen to avoid rotational resonance owing to overlap of the sidebands of carbonyl ^{13}C signals with the aromatic and aliphatic signals in the acquisition dimension (F3). Because of these constraints, the bandwidth in the indirect acquisition dimensions (F1, F2) left spectral regions known to be devoid of signals. Whereas those "dark" regions are conventionally neglected, SIFT actively uses them.

To model critical sensitivity, we intentionally underpacked the sample rotor such that only about 50% of the active volume was filled. With a molecular weight of about 6.4 kDa for GB1, the NMR sensitivity for this half-packed sample would correspond to that of a fully packed circa 13 kDa microcrystalline protein at the same density. Signal averaging took about 7 min per t1/t2 sample for an acceptable S/N ratio, whilst extended signal evolution was also required for suitable linewidths.

The NUS schedule employed (Figure 1b) omits more than 70% of the time domain samples in the full, uniform grid of $64(t1)\times32(t2)$ time points. The input for SIFT processing comprises the NUS schedule, the acquired time-domain data, and specification of the dark spectral regions. The processing, which involves no user intervention or parameter tuning, took about 2 min on a single processor. The SIFTed time domain data may be transformed and phased as though directly acquired.

Figure 2 shows that the S/N degradation owing to NUS (d) is almost perfectly reversed after SIFT-processing (e), leading to S/N that is nearly the same as that of the fully acquired spectrum with more than three times as many acquired points (a). This effect is significantly better than $(608/2048)^{1/2}=0.54$ relative S/N that is expected for the shortened acquisition time and obtained with other major processing methods, such as multidimensional decomposition (MDD),[19] that rely almost exclusively on information in the time domain. The nearly identical S/N achieved in less than one third of the acquisition time means that the SIFTed spectrum is more than 1.7-fold more sensitive per unit time relative to the conventionally acquired spectrum.

The improved S/N provided by SIFT processing reduced the number of lost signals by a factor of about ten compared with the NUS spectrum not processed by SIFT. Of the 123 correlation signals observed in the fully sampled data, the weaker 62 had S/N ratios ranging from 3.1 to 7.7. At this critical sensitivity, classical discrete Fourier transform of the NUS data (Figure 2d) resulted in 18 (29%) of the 62 weaker peaks being lost (that is, intensity below the threshold of 3σ, where σ is the average noise standard deviation measured in each spectrum). On the other hand, the SIFTed spectrum lost only 2 (3.2%) of these peaks (see the Supporting Information).

With NUS and SIFT combined, we recorded a 3D NCOCX spectrum at about 1 ppm digital resolution in 2.8 days, instead of 9.5 days with the standard method. A sample with twice the molecular weight (or a fully-packed circa 26 kDa protein), would require about $7\times2^2=28$ min per t1/t2 sample for the same sensitivity. This translates into an 11 day experiment with NUS versus a 38 day conventional ^{13}C detected experiment.

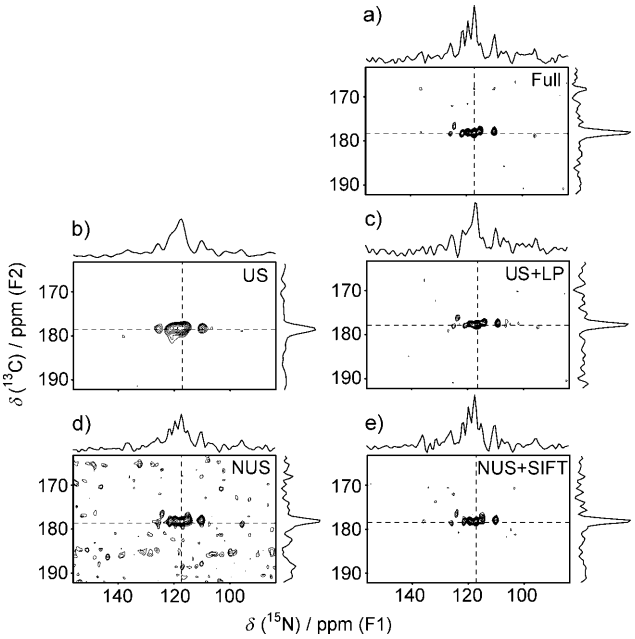

Figure 2. A two-dimensional slice taken at F3 = 178.4 ppm from the NCOCX spectrum using a) the full, uniform 2048 t1/t2-samples, b) the truncated 608 uniform samples corresponding to the shaded area in Figure 1b with zero-filling , c) the data from (b) with linear prediction, d) the 608 NUS samples corresponding to the dots in Figure 1b without SIFT processing , and e) the data from (d) with SIFT processing. The spectrum without SIFT processing ((d), where FFT was used after the missing time points were filled with zeroes) corresponds to the classical discrete Fourier transform (DFT) of the NUS data.[17] The minimum contour line is at 18% of the tallest peak in each panel. 1D traces at dashed lines are also shown, for which the vertical scales are the same for all panels. The slice shown contains signals with about 1–2 times the median peak intensity and a S/N ratio of between 8.4 and 15.6. Slices with weaker signals are shown in the Supporting Information, Figure S1.

In addition to excellent S/N, SIFT offers high fidelity. Line-shape distortion of the type seen in MaxEnt-processed spectra[21,22] is not observed in the SIFTed spectrum, despite the much lower intrinsic S/N of the present data. We also found that SIFT yields more accurate peak intensities and positions than those obtained by linear prediction (LP) of a truncated uniformly sampled (US) dataset. For example, Figure 2c shows that LP does not accurately restore crowded spectral regions even with a reasonable number of coefficients (8). This effect is especially clear in the F1 slice. Further instances are illustrated by the additional F1 slices in Figure 3. In all the spectral regions where LP failed, the SIFT spectrum reproduced the full data very well. Overall, SIFT rendered peak frequencies with RMS errors of 0.13, 0.11, and 0.09 ppm in F1, F2, and F3, respectively (see the Supporting Information, Figure S2).

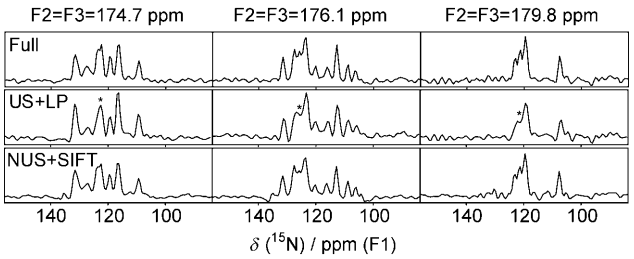

Figure 3. Representative F1 slices through crowded spectral regions. Each column compares the spectrum of the full data (top), LP of truncated US data (middle), and SIFT-processed NUS data (bottom). From the left to right, slices were taken at F2 = F3 = 174.7, 176.1, and 179.8 ppm. The vertical scale is the same throughout each column. The asterisks mark the unresolved peaks in the LP spectra.

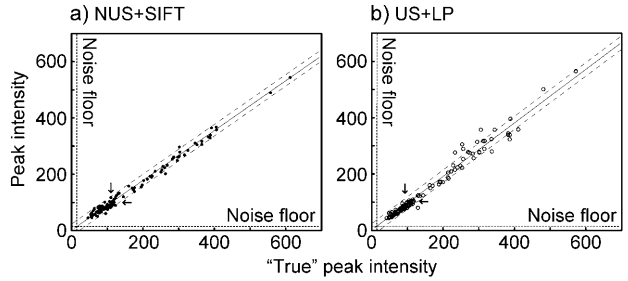

Figure 4. Signal intensities observed in the spectrum of a) FT-processed NUS data and b) LP-processed US data versus fully sampled reference data. The linear regression shown with a solid line is slightly different between (a) and (b). Dashed lines flanking the regression line show the intrinsic noise width in the reference spectrum. Dashed lines along the *x* and *y* axes show the floor of noise standard deviation. The medians of all the observed signal intensities in the reference, SIFT-processed, and linear-predicted spectra (marked by arrows) were about 110, 100, and 95, respectively.

To assess the accuracies of peak intensities, the intensities in the SIFTed spectrum and the linear predicted spectrum versus the "true" peak intensities obtained from the fully acquired dataset were plotted (Figure 4). The excellent linearity obtained with SIFT (Figure 4a) demonstrates accurate relative peak intensities. The correlation coefficient was 0.995 overall and 0.877 for the "weak" signals. The dynamic range was about 13 in this example, but the superb linearity of SIFT signal intensities over a dynamic range of about 100 has been shown previously.[24] The variation of the observed peak intensities was mostly within the intrinsic noise-width of the dataset (ca. ± 25), which is indicated by the dashed lines flanking the regression line. This observation indicates the absence of intensity bias in SIFT processing. The accuracy remained high for the weakest signals, thus confirming the robustness of the SIFT process in the presence of formidable noise. In contrast, LP tends to reproduce large signals with less accuracy (Figure 4b).

In conclusion, by using the noise-tolerant SIFT process,[24] we have extended the applicability of high-dimensional NUS-NMR methodology to data with the marginal sensitivity that is typical of MAS NMR of biological macromolecules. Quick SIFT processing (ca. 2 min) of NUS data yielded a high-quality 3D spectrum without any calibration or parameter optimization. After SIFT processing, the reduced number of time samples in NUS did not appreciably decrease the S/N relative to that for uniformly acquired reference data. Meanwhile, the measurement time was reduced by a factor of about 3.4. These results suggest that a 3D NCOCX-type MAS experiment can be recorded at sufficient sensitivity to resolve single nuclear sites for circa 25 kDa proteins in a reasonable period of time. The approach demonstrated herein requires no special hardware and will expedite experiments similarly on all FT spectrometers. For example, if a tenfold sensitivity gain were available by dynamic nuclear polarization (DNP),[25] the above high-resolution 3D experiment would be possible for 250 kDa proteins. Moreover, the exquisite accuracy of SIFT signal frequencies and intensities paves the way for quantitative structural and dynamical investigations in noisy systems that have frustrated all other reported NUS processing methods. Thus, SIFT will significantly expand opportunities for high-dimensional MAS NMR experiments in studies of large molecules and molecular assemblies of biological and medical importance.

Experimental Section

Uniformly ^{13}C and ^{15}N labeled GB1 was prepared according to previously published procedures[26,27] as described in the Supporting Information. The NMR experiment was performed on a custom-built 500 MHz (^1H frequency) spectrometer equipped with a solenoid-coil 3.2 mm MAS system (Revolution NMR, Fort Collins, CO). The sampling schedule is converted into a text-based list that is read by the pulse program and control macro to set respective delays (courtesy of Dr. P. van der Wel, University of Pittsburgh). Details on the NMR parameters are given in the Supporting Information.

MATLAB scripts for SIFT processing are available at http://www.brandeis.edu/~herzfeld/SIFT. The signal-containing "bright" region was $\delta = 102.5$–133.2 ppm (F1, ^{15}N) and $\delta = 169.3$–183.9 ppm (F2, $^{13}C'$), which was established from 1D "scouting" experiments for the corresponding nuclei. The number of SIFT cycles was 10, which took about 2 min. Further details on processing parameters and spectral analysis are given in the Supporting Information.

Received: June 1, 2010
Published online: October 18, 2010

Keywords: data acquisition methods · NMR spectroscopy · nonuniform sampling · protein structures

[1] A. K. Paravastu, R. D. Leapman, W.-M. Yau, R. Tycko, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18349.
[2] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, *Science* **2008**, *319*, 1523.
[3] K. Iwata, T. Fujiwara, Y. Matsuki, H. Akutsu, S. Takahashi, H. Naiki, Y. Goto, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18119.
[4] M. Etzkorn, S. Martell, O. C. Andronesi, K. Seidel, M. Engelhard, M. Baldus, *Angew. Chem.* **2007**, *119*, 463; *Angew. Chem. Int. Ed.* **2007**, *46*, 459.
[5] V. S. Bajaj, M. L. Mak-Jurkauskas, M. Belenky, J. Herzfeld, R. G. Griffin, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9244.
[6] C. Ader, R. Schneider, S. Hornig, P. Velisetty, E. M. Wilson, A. Lange, K. Giller, I. Ohmert, M. F. Martin-Eauclaire, D. Trauner,

S. Becker, O. Pongs, M. Baldus, *Nat. Struct. Mol. Biol.* **2008**, *15*, 605.
[7] A. Goldbourt, B. J. Gross, L. A. Day, A. E. McDermott, *J. Am. Chem. Soc.* **2007**, *129*, 2338.
[8] A. Egawa, T. Fujiwara, T. Mizoguchi, Y. Kakitani, Y. Koyama, H. Akutsu, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 790.
[9] A. C. Sivertsen, M. J. Bayro, M. Belenky, R. G. Griffin, J. Herzfeld, *J. Mol. Biol.* **2009**, *387*, 1032.
[10] C. M. Rienstra, M. Hohwy, M. Hong, R. G. Griffin, *J. Am. Chem. Soc.* **2000**, *122*, 10979.
[11] B. Q. Sun, C. M. Rienstra, P. R. Costa, J. R. Williamson, R. G. Griffin, *J. Am. Chem. Soc.* **1997**, *119*, 8540.
[12] M. Hong, *J. Biomol. NMR* **1999**, *15*, 1.
[13] W. T. Franks, K. D. Kloepper, B. J. Wylie, C. M. Rienstra, *J. Biomol. NMR* **2007**, *39*, 107.
[14] Y. Li, D. A. Berthold, H. L. Frericks, R. B. Gennis, C. M. Rienstra, *ChemBioChem* **2007**, *8*, 434.
[15] S. Kim, T. Szyperski, *J. Am. Chem. Soc.* **2003**, *125*, 1385.
[16] E. Kupče, R. Freeman, *J. Am. Chem. Soc.* **2004**, *126*, 6429.
[17] K. Kazimierczuk, W. Kozminski, I. Zhukov, *J. Magn. Reson.* **2006**, *179*, 323.
[18] M. Mobli, M. W. Maciejewski, M. R. Gryk, J. C. Hoch, *Nat. Methods* **2007**, *4*, 467.
[19] V. Jaravine, I. Ibraghimov, V. Y. Orekhov, *Nat. Methods* **2006**, *3*, 605.
[20] C. D. Ridge, V. A. Mandelshtam, *J. Biomol. NMR* **2009**, *43*, 151.
[21] D. Rovnyak, C. Filip, B. Itin, A. S. Stern, G. Wagner, R. G. Griffin, J. C. Hoch, *J. Magn. Reson.* **2003**, *161*, 43.
[22] D. H. Jones, S. J. Opella, *J. Magn. Reson.* **2006**, *179*, 105.
[23] 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.
[24] Y. Matsuki, M. T. Eddy, J. Herzfeld, *J. Am. Chem. Soc.* **2009**, *131*, 4648.
[25] T. Maly, G. T. Debelouchina, V. S. Bajaj, K. N. Hu, C. G. Joo, M. L. Mak-Jurkauskas, J. R. Sirigiri, P. C. A. van der Wel, J. Herzfeld, R. J. Temkin, R. G. Griffin, *J. Chem. Phys.* **2008**, *128*, 052211.
[26] W. Franks, D. Zhou, B. Wylie, B. Money, D. Graesser, H. Frericks, G. Sahota, C. Rienstra, *J. Am. Chem. Soc.* **2005**, *127*, 12291.
[27] H. L. F. Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson, C. M. Rienstra, *J. Phys. Chem. B* **2007**, *111*, 14362.

