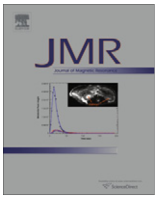
Journal of Magnetic Resonance 214 (2012) 296–301

Contents lists available at SciVerse ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr



Deterministic schedules for robust and reproducible non-uniform sampling in multidimensional NMR

Matthew T. Eddy [a,b], David Ruben [b], Robert G. Griffin [a,b], Judith Herzfeld [c,*]

[a] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Department of Chemistry, Brandeis University, Waltham, MA 02454, USA

ARTICLE INFO

Article history:
Received 9 September 2011
Revised 16 November 2011
Available online 10 December 2011

Keywords:
Non-uniform sampling
Non-linear sampling
Sparse sampling
Multi-dimensional NMR
Sampling indirect dimensions
Sampling schedules

ABSTRACT

We show that a simple, general, and easily reproducible method for generating non-uniform sampling (NUS) schedules preserves the benefits of random sampling, including inherently reduced sampling artifacts, while removing the pitfalls associated with choosing an arbitrary seed. Sampling schedules are generated from a discrete cumulative distribution function (CDF) that closely fits the continuous CDF of the desired probability density function. We compare random and deterministic sampling using a Gaussian probability density function applied to 2D HSQC spectra. Data are processed using the previously published method of Spectroscopy by Integration of Frequency and Time domain data (SIFT). NUS spectra from deterministic sampling schedules were found to be at least as good as those from random schedules at the SIFT critical sampling density, and significantly better at half that sampling density. The method can be applied to any probability density function and generalized to greater than two dimensions.

1. Introduction

Non-uniform sampling (NUS) is a powerful method for significantly reducing NMR data acquisition times and improving spectral resolution. This is especially true at stronger magnetic fields where simultaneously realizing the benefits of high resolution and desired bandwidth requires more extensive sampling. NUS has been successfully applied to a number of solution NMR problems that require high dimensionality and resolution, including assigning the signals and determining the structures of large soluble proteins [1] and solubilized membrane proteins [2], studying highly degenerate systems [3,4], measuring residual dipolar couplings [5], characterizing metabolic mixtures from cellular extracts [6], and studying transient systems, including proteins within whole cells [7]. Furthermore, NUS is spreading to solid state NMR where it has already proven useful for assignments in multidimensional experiments [8], unambiguous restraints in structure determination [9], multiple quantum magic angle spinning experiments on quadrupolar nuclei [10], and PISEMA experiments [11].

A problem with NUS is that it introduces a great deal of variability into experimental protocols, both with respect to sampling schedules and data processing algorithms. The simplest processing of NUS data is the discrete Fourier transform. Improved results can be obtained using maximum entropy and other model-dependent methods [10,12–23]. However, the price is variability due to modeling assumptions. Recently, it has been shown that modeling can be avoided by using knowledge of zeroes in the frequency domain to replace information missing in the time domain [24]. This process of Spectroscopy by Integration of Frequency and Time domain information (SIFT) is rapid and can be pursued with various, but well-defined, degrees of aggressiveness (in identifying frequency zeroes and in the ratio of time points dropped to frequency zeroes identified).

The situation is more complex with respect to sampling schedules. The idea has always been to sample more heavily at early times when the signal is strongest and only as much at long times as is necessary to resolve signals of interest. Early NUS used exponential sampling distributions, roughly paralleling the decay in the signal intensity [25]. However, it has recently been shown that Gaussian sampling provides better results [26]; apparently, the greater emphasis on sampling at early times need not entail undue sacrifice of sampling at late times.

Generating deterministic, and therefore easily reproducible, sampling schedules is straightforward, whether according to an exponential distribution [25] or any other distribution (see below). However, random sampling has become popular since it has been shown to decrease artifacts that arise from sampling below the Nyquist density [26,27] and minimize spectral aliasing. The difficulty is that random sampling creates two further problems: choosing a seed number and reproducibility. As Hyberts et al. have recently demonstrated in great detail [28], the choice of seed number used to generate a random schedule can yield widely varying

* Corresponding author. Fax: +1 781 736 2516.
E-mail addresses: meddy@mit.edu (M.T. Eddy), ruben@fbml-cmr.mit.edu (D. Ruben), rgg@mit.edu (R.G. Griffin), herzfeld@brandeis.edu (J. Herzfeld).

1090-7807/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2011.12.002

spectral quality. To ascertain the quality of spectra obtained with random sampling, one would need to collect spectra with many different seeds, defeating the time-saving benefits of NUS. Second, while it is difficult to define exactly what criteria should be used to judge the "best" schedule, a goal we do not attempt to pursue here, it is generally desirable that experimental results be reproducible. In principle this is not achievable with random sampling unless all the sampled points are specified for each spectrum (or the specific random number generator algorithm and precise seed are supplied). Thus, the benefits of random sampling must be weighed against the potential pitfalls of selecting a bad seed and the desire for straightforwardly reproducible results.

Here we show that a deterministic approach can preserve the benefits of random sampling (i.e., inherently reduced spectral artifacts) while avoiding the need for an arbitrary seed, and the attendant possibility of generating a poor sampling schedule. The method is a simple and entirely reproducible alternative to stochastic sampling. We compare the two approaches in the context of 2D HSQC experiments on the β1 domain of immunoglobulin binding protein G (GB1). Importantly, our approach is entirely reproducible and can be generalized to experiments beyond two dimensions and to sampling distributions other than Gaussian.

2. Methods

2.1. Sample preparation and NMR spectroscopy

U-^{15}N labeled GB1 (mutant T2Q) was prepared as described previously [24]. The solution HSQC data were recorded at 278 K with a gradient-enhanced scheme [29] at 591 MHz (^{1}H Larmor frequency) using a custom-built console and software and a Z-SPEC 5 mm triple-resonance IDTG590-5 probe (NALORAC Co., CA). Four scans were averaged at a recycle delay of 2 s. The ^{15}N bandwidth of 67 ppm (3984 Hz) was uniformly sampled with 128 points, and the ^{1}H bandwidth of 13.6 ppm (8013 Hz) was uniformly sampled with 1024 complex points. The total acquisition time was 34 min. The maximum evolution time in the full data set was 32 ms. The master spectrum corresponding to the full data set is shown in Fig. 1.

2.2. Generation of random NUS schedules

Random on-grid NUS schedules were generated with a Gaussian probability distribution, exp($-t^2/2\sigma^2$), by first generating corresponding off-grid optimized Gaussian sampling using the "timetab_gen" program available from the Warsaw NMR group: http://nmr700.chem.uw.edu.pl/. To conform to a grid, the evolution time of each point was increased just enough to coincide with the first unoccupied grid point.

2.3. Generation of deterministic NUS schedules

A unique sampling schedule is generated from the cumulative distribution function (CDF) of a given desired probability density function (PDF). The CDF is the integral of the PDF and can always be determined, if necessary by numerical methods. A sampling schedule which closely approximates the desired CDF will automatically closely approximate the desired PDF.

Here we provide details of the scheduling algorithm for the Gaussian PDF

$$\mathrm{pdf}(t) = \frac{\exp(-t^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

where t ranges from $-\infty$ to $+\infty$. The corresponding cumulative distribution function for positive t is

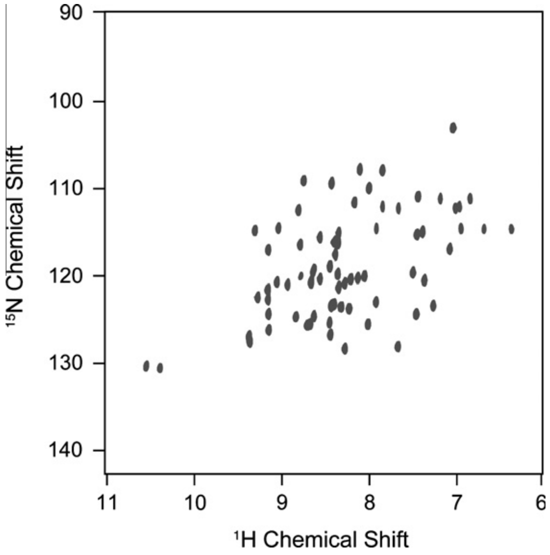


Fig. 1. HSQC spectrum of GB1 acquired with full uniform sampling of 128 t1 points.

$$\mathrm{cdf}(t) = \mathrm{erf}\left(t/\sqrt{2\sigma^2}\right)$$

where erf is the error function. We map the uniform sampling grid to the time domain region between 0.0 and 1.0 using $t = I_{GRID}/N_{GRID}$, where N_{GRID} is the total number of grid points and I_{GRID} ranges between 1 and N_{GRID}. This conventional mapping is completely general given the freedom to choose σ. The CDF is scaled to obtain N_{SCHED} acquisition points

$$N_{\mathrm{SCHED}}\left[\frac{\mathrm{cdf}(t)}{\mathrm{cdf}(1.0)}\right]$$

and discretized by rounding the value off to the nearest integer. This discretized CDF is calculated for each grid point and those grid points corresponding to the step increases in value are chosen as the points to be experimentally acquired. In order to facilitate later NMR processing, we move the first scheduled point to the first grid point, in the rare case that it is not already selected. This algorithm can be used with only trivial modifications for any probability distribution other than Gaussian.

2.4. Sift

SIFT, uses knowledge of zeros in the frequency domain to fill gaps in the time series without affecting the acquired points [24]. Here we use the most conservative form of SIFT which refrains from identifying dark regions between signals by either precedent or thresholding. Rather, dark regions are included at the fringes of the spectrum by expanding the bandwidth. This increased bandwidth decreases the dwell times proportionately. However, as shown in the original work, the freedom to sample non-uniformly nevertheless improves S/N. Furthermore, the results only degrade slowly when sampling is reduced beyond a simple 1:1 trade-off of time points for frequency points.

The SIFT cycle has been implemented in MATLAB (http://people.brandeis.edu/~herzfeld/SIFT), calling for input files that contain (1) the sampling schedule, (2) the corresponding time domain NUS data, and (3) specification of the dark frequency points. In the present application, signals occur only between 132 and 101 ppm in the ^{15}N dimension (see Fig. 1). All areas outside this range were defined as dark and amount to half of the frequency points. Therefore, for SIFT in this experiment, 50% NUS is critical and 25% NUS is subcritical.

2.5. Data processing

All time domain data (SIFTed or not) were processed in MATLAB by applying a cosine-squared function in t1 that reached zero at the maximum evolution time, and zero filling to 4096 points in the direct dimension and 512 points in the indirect dimension. Spectra were then exported to Sparky for viewing and plotting [30]. Contour levels were set to 10% of the highest peak intensity in each spectrum. SIFTed data were processed by first applying SIFT. The resulting time domain data were then processed as described above.

3. Results

To evaluate the deterministic approach we compared results for Gaussian NUS with a random schedule and our deterministic schedule. The initial comparison used 64 t1 points (50% of full uniform sampling density) and $\sigma = 0.5$. As shown in Fig. 2, the results are similar for the two schedules at this level of sampling. With SIFT processing, both faithfully reproduce the original HSQC spectrum.

Often more aggressive NUS is desired. But the sparser the sampling, the more room for mischief there is in random sampling. Fig. 3 shows the results of Gaussian NUS with $\sigma = 0.5$ at the 25% level (i.e., with only 32 t1 points). While the results are degraded for both the random and the deterministic sampling schedules, the effect is milder for the latter. Looking at the unSIFTed spectra at the top of the figure, we see that the deterministic schedule is significantly less noisy in the bright region (i.e., the area of interest between 132 and 101 ppm in the ^{15}N dimension and 10.9 and 6.1 ppm in the ^{1}H dimension). Since SIFT works by imposing zeroes in the known dark regions of the spectra, pushing noise outside of the bright region and into the dark regions produces a less noisy SIFTed spectrum. Deterministic sampling is presumably accomplishing this by avoiding large gaps between consecutive points.

On the other hand, the deterministically sampled spectra show some aliasing in the nitrogen dimension. This indicates excessive uniformity in the gaps between points, which corresponds to a probability distribution that is too flat for such a small number of points. Fig. 4 shows that choosing the 32 points with a tighter Gaussian ($\sigma = 0.3$, comparable to the fraction of grid points sampled) provides cleaner results, not only for deterministic sampling, but also for random sampling. The improved quality is at least partially due to increased sampling at earlier time points. At the same time, the line widths remain good in spite of reduced sampling at

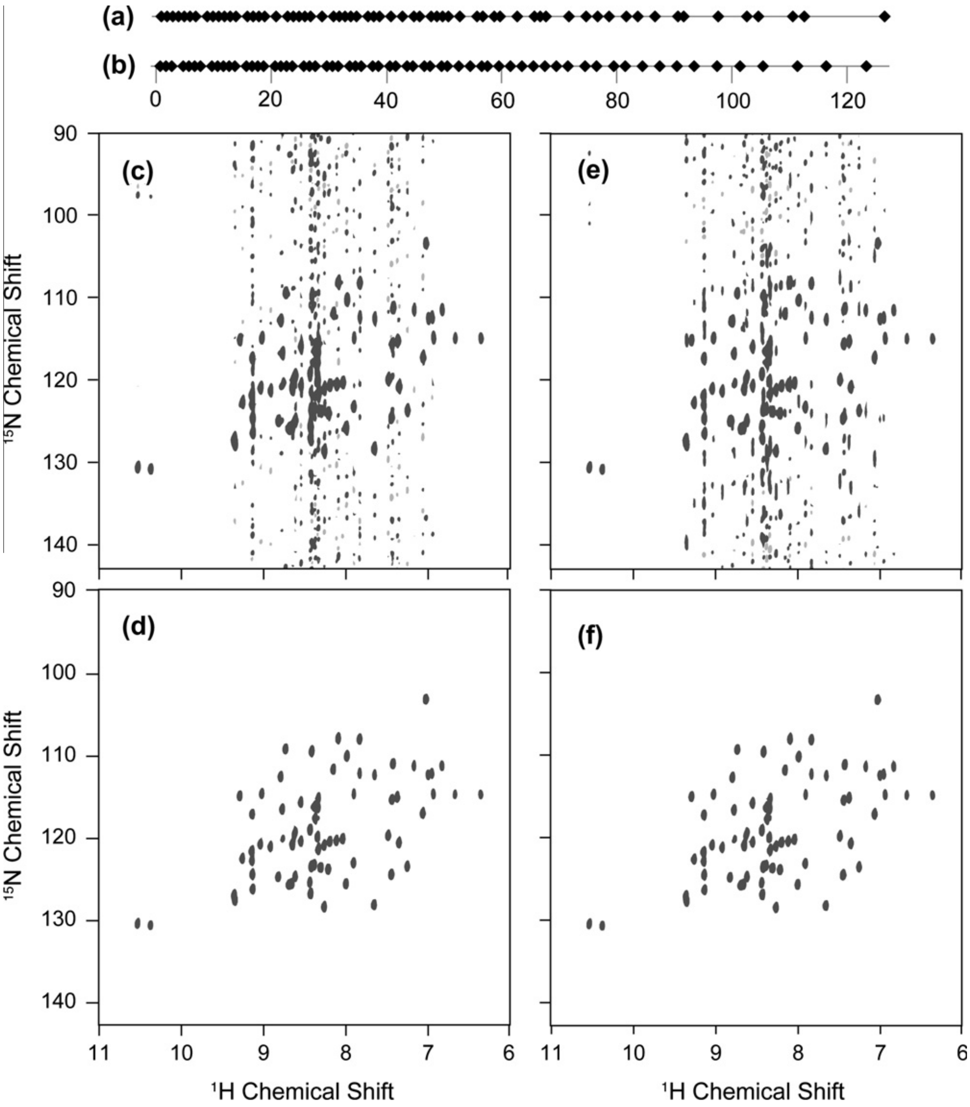


Fig. 2. Critical HSQC spectra of GB1 from 64 Gaussian distributed t1 points with sigma = 0.5: (a) the random sampling schedule, (b) the deterministic sampling schedule, (c) the spectrum resulting from the random schedule without SIFT processing, (d) the spectrum resulting from the random schedule with SIFT processing, (e) the spectrum resulting from the deterministic schedule without SIFT processing and (f) the spectrum resulting from the deterministic schedule with SIFT processing.

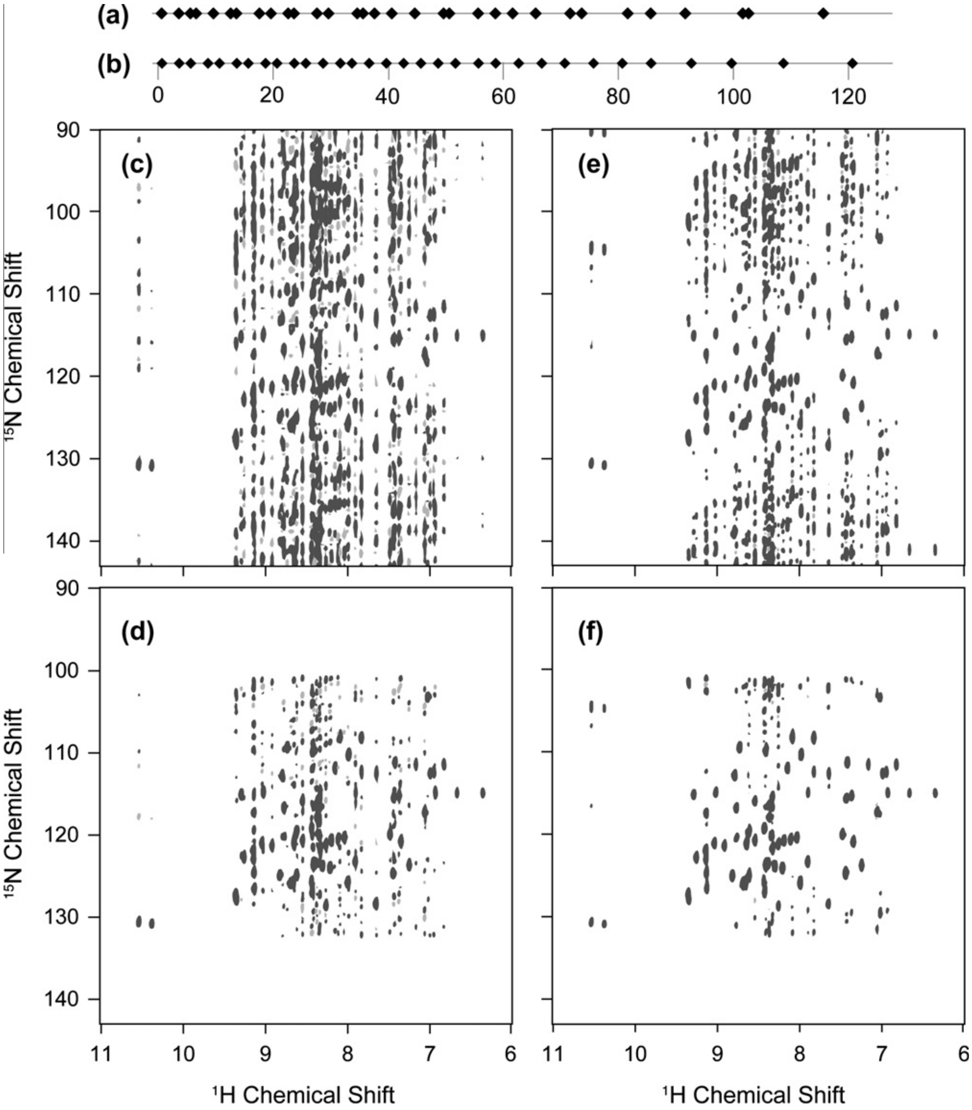


Fig. 3. Subcritical HSQC spectra of GB1 from 32 Gaussian distributed t1 points with $\sigma = 0.5$: (a–e) as in Fig. 2.

long evolution times. Overall, the deterministic sampling schedule generates fewer artifacts than the random schedule; it has less noise in the bright region and somewhat better post-SIFT results.

It should be noted that the comparisons made here all used random sampling based on the single seed embedded in the "timetab_gen" program (see Section 2). Of course, there may be better seeds than the one provided by this utility. But the good seeds are likely to vary from experiment to experiment (e.g., depending on the chosen level of sampling and value of σ) and finding them would involve more effort than full sampling to begin with, thereby totally defeating the benefits of NUS.

4. Discussion

Problems with random sampling have been previously reported, as noted above. Methods to overcome these problems have also been proposed, including jittered sampling [31], random sampling with constraints [32], and Poisson disk sampling [32]. These methods aim to minimize the clustering of samples in the NUS schedule, providing smoother sampling while maintaining the benefits of randomization. Our method produces the same sought after effects with a more straightforward and reproducible approach that does not require calculation of additional sampling parameters and appears to give results that are at least as good as a typical random method. Spectra recorded via Poisson disk sampling or random sampling with constraints have been reported to show an inhomogeneous distribution of sampling noise, with less noise observed near real peaks [32]. As we observe a similar effect with our deterministic approach, it seems that we have achieved similar favorable noise characteristics in a simpler fashion. Moreover, since SIFT fills in the FID by applying known frequency zeros in the dark regions, this type of noise shaping is especially suited to SIFT processing.

5. Conclusions

Deterministic NUS, using a CDF corresponding closely to the targeted probability distribution, appears to provide a reliable and effective alternative to random NUS. Smoothing the sampling function (i.e., avoiding unnecessarily large gaps between time points) produces less noise in bright areas, thereby leading to better results, especially with SIFT processing. Aliasing is easily avoided at lower sampling levels by choosing steeper probability distributions.

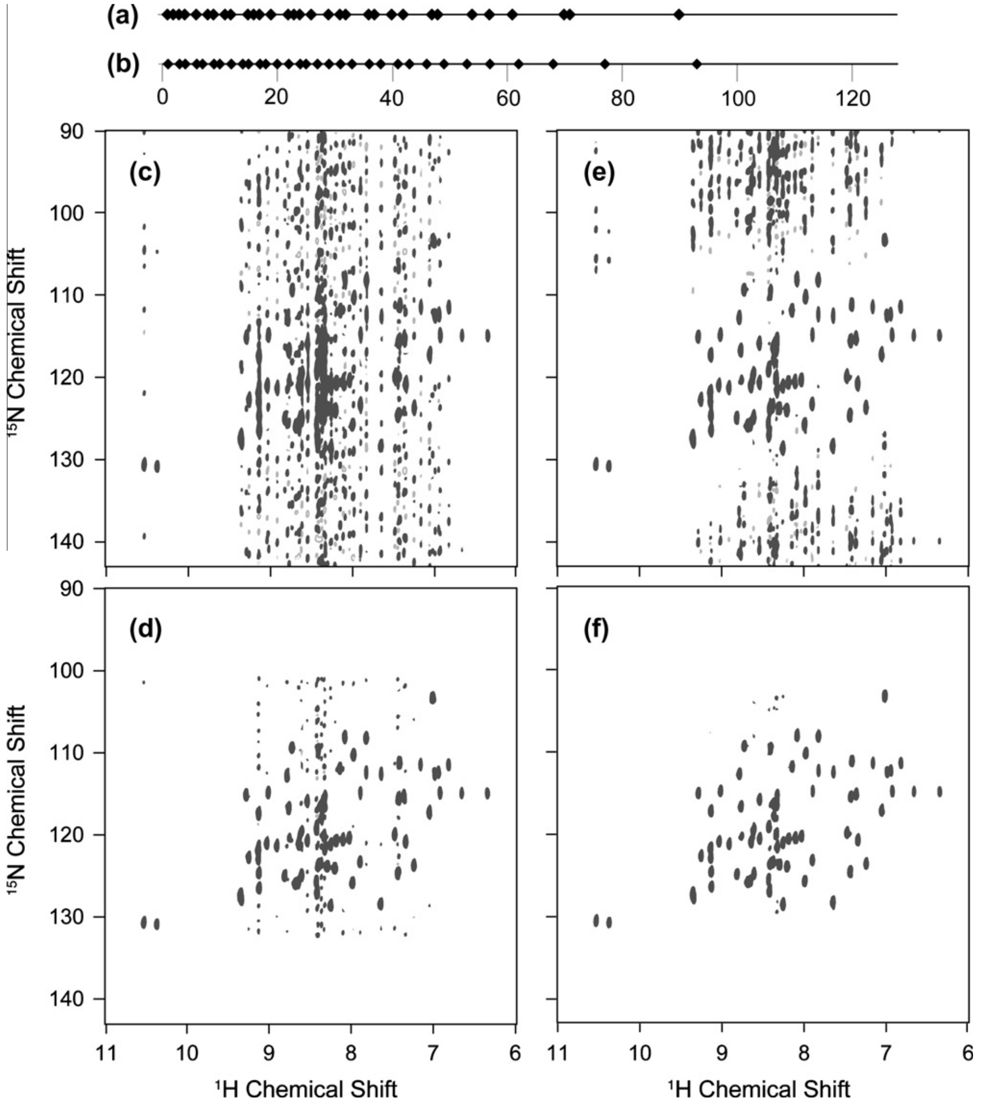


Fig. 4. As in Fig. 3, but with $\sigma = 0.3$.

Role of funding source

The study sponsors had no role in the design of this study; the collection, analysis, or interpretation of the data; the writing of the report; or the decision to submit the paper for publication.

Acknowledgments

The authors would like to thank Dr. David Rovnyak for helpful discussions, Dr. Chris Turner for his assistance in collecting the initial data, and Dr. Yoh Matsuki for assistance with the SIFT MATLAB code and helpful discussions. This work was supported by NIH grants EB001035, EB003151, EB001960, and EB002026.

References

[1] A. Zawadzka-Kazimierczuk, K. Kazimierczuk, W. Koźmiński, A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins, Journal of Magnetic Resonance 202 (2010) 109–116.

[2] S. Hiller, I. Ibraghimov, G. Wagner, V.Y. Orekhov, Coupled decomposition of four-dimensional NOESY spectra, Journal of the American Chemical Society 131 (2009) 12970–12978.

[3] N. Pannetier, K. Houben, L. Blanchard, D. Marion, Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins, Journal of Magnetic Resonance 186 (2007) 142–149.

[4] V.A. Jaravine, A.V. Zhuravleva, P. Permi, I. Ibraghimov, V.Y. Orekhov, Hyperdimensional NMR spectroscopy with nonlinear sampling, Journal of the American Chemical Society 130 (2008) 3927–3936.

[5] J.A. Kubat, J.J. Chou, D. Rovnyak, Nonuniform sampling and maximum entropy reconstruction applied to the accurate measurement of residual dipolar couplings, Journal of Magnetic Resonance 186 (2007) 201–211.

[6] 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, Ultrahigh-resolution (1)H–(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction, Journal of the American Chemical Society 129 (2007) 5108–5116.

[7] D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B.O. Smith, M. Shirakawa, P. Güntert, Y. Ito, Protein structure determination in living cells by in-cell NMR spectroscopy, Nature 458 (2009) 102–105.

[8] Y. Matsuki, M.T. Eddy, R.G. Griffin, J. Herzfeld, Rapid 3D MAS NMR spectroscopy at critical sensitivity, Angewandte Chemie (International ed. in English) (2010).

[9] M. Huber, S. Hiller, P. Schanda, M. Ernst, A. Böckmann, R. Verel, B.H. Meier, A. Proton-Detected, 4D solid-state NMR experiment for protein structure determination, ChemPhysChem 12 (2011) 915–918.

[10] D. Rovnyak, C. Filip, B. Itin, A.S. Stern, G. Wagner, R.G. Griffin, J.C. Hoch, Multiple-quantum magic-angle spinning spectroscopy using nonlinear sampling, Journal of Magnetic Resonance 161 (2003) 43–55.

[11] D.H. Jones, S.J. Opella, Application of maximum entropy reconstruction to PISEMA spectra, Journal of Magnetic Resonance 179 (2006) 105–113.

[12] D. Rovnyak, D.P. Frueh, M. Sastry, Z.-Y.J. Sun, A.S. Stern, J.C. Hoch, G. Wagner, Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction, Journal of Magnetic Resonance 170 (2004) 15–21.

[13] P. Schmieder, A.S. Stern, G. Wagner, J.C. Hoch, Application of nonlinear sampling schemes to COSY-type spectra, Journal of Biomolecular NMR 3 (1993) 569–576.

[14] A.S. Stern, K.-B. Li, J.C. Hoch, Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear-prediction extrapolation and maximum-entropy reconstruction, Journal of the American Chemical Society 124 (2002) 1982–1993.

[15] T. Luan, V. Jaravine, A. Yee, C.H. Arrowsmith, V.Y. Orekhov, Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition, Journal of Biomolecular NMR 33 (2005) 1–14.

[16] V. Tugarinov, L.E. Kay, I. Ibraghimov, V.Y. Orekhov, High-resolution four-dimensional H-1-C-13 NOE spectroscopy using methyl-TROSY, Sparse data acquisition, and multidimensional decomposition, Journal of the American Chemical Society 127 (2005) 2767–2775.

[17] F.L. Zhang, R. Bruschweiler, Indirect covariance NMR spectroscopy, Journal of the American Chemical Society 126 (2004) 13180–13181.

[18] N. Trbovic, S. Smirnov, F.L. Zhang, R. Bruschweiler, Covariance NMR spectroscopy by singular value decomposition, Journal of Magnetic Resonance 171 (2004) 277–283.

[19] E. Kupce, R. Freeman, Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy, Journal of the American Chemical Society 126 (2004) 6429–6440.

[20] E. Kupce, R. Freeman, Fast reconstruction of four-dimensional NMR spectra from plane projections, Journal of Biomolecular NMR 28 (2004) 391–395.

[21] H.S. Atreya, E. Garcia, Y. Shen, T. Szyperski, J-GFT NMR for precise measurement of mutually correlated nuclear spin-spin couplings, Journal of the American Chemical Society 129 (2007) 680–692.

[22] T. Szyperski, D.C. Yeh, D.K. Sukumaran, H.N.B. Moseley, G.T. Montelione, Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment, Proceedings of the National Academy of Sciences of the United States of America 99 (2002) 8009–8014.

[23] G. Bodenhausen, R.R. Ernst, The accordion experiment, a simple approach to 3-dimensional NMR-spectroscopy, Journal of Magnetic Resonance 45 (1981) 367–373.

[24] Y. Matsuki, M.T. Eddy, J. Herzfeld, Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra, Journal of the American Chemical Society 131 (2009) 4648–4656.

[25] J. Barna, E. Laue, M. Mayger, J. Skilling, S. Worrall, Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments, Journal of Magnetic Resonance 73 (1987) 69–77.

[26] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, Journal of Biomolecular NMR 36 (2006) 157–168.

[27] J.C. Hoch, M.W. Maciejewski, B. Filipovic, Randomization improves sparse sampling in multidimensional NMR, Journal of Magnetic Resonance 193 (2008) 317–320.

[28] S.G. Hyberts, K. Takeuchi, G. Wagner, Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data, Journal of the American Chemical Society 132 (2010) 2145–2147.

[29] L. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, Journal of the American Chemical Society 114 (1992) 10663–10665.

[30] T. D. Goddard, D. G. Kneller, SPARKY 3, University of California, San Francisco.

[31] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Lineshapes and artifacts in multidimensional Fourier transform of arbitrary sampled NMR data sets, Journal of Magnetic Resonance 188 (2007) 344–356.

[32] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, Optimization of random time domain sampling in multidimensional NMR, Journal of Magnetic Resonance 192 (2008) 123–130.