

Spectral reconstruction methods in fast NMR: Reduced dimensionality, random sampling and maximum entropy

Mehdi Mobli^a, Alan S. Stern^b, Jeffrey C. Hoch^{a,*}

^a Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-3305, USA

^b Rowland Institute at Harvard, 100 Edwin H. Land Boulevard, Cambridge, MA 02142, USA

Received 26 April 2006; revised 25 May 2006

Available online 11 July 2006

Abstract

The need to reduce data acquisition times of multidimensional NMR experiments has fostered considerable interest in novel data acquisition schemes. A recurring theme is that of reduced dimensionality experiments, in which time evolutions in the indirect dimensions are incremented together, rather than independently. Spectral analysis of such data is carried out using methods such as filtered back-projection, GFT, or parametric signal modeling. By using Maximum Entropy reconstruction of reduced-dimensionality data, we show that the artifacts that arise in reduced dimensionality experiments are intrinsic to the data sampling, and are not, in general, the result of the methods used to compute spectra. Our results illustrate that reduced dimensionality is a special case of non-uniform sampling in the time domain. We show that MaxEnt reconstruction yields more accurate spectra for reduced dimensionality data than back-projection reconstruction and that randomly choosing time increments based on an exponentially weighted distribution is more efficient, with fewer artifacts, than the systematic coupling of time increments used in most reduced dimensionality approaches.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Data processing; Spectrum analysis; Non-uniform data sampling; Maximum entropy; Backprojection reconstruction; Reduced dimensionality

1. Introduction

Recent advances in hardware and pulse sequence design have contributed to very active research in methods requiring less data for the reconstruction of multidimensional NMR experiments (for detail see reviews [1,2]). These methods are especially appropriate for the analysis of bio-molecules, where multiple multidimensional experiments are employed to enable sequence specific assignment and to obtain structural information, typically requiring days or weeks of instrument time. The time saved by using fewer data samples is not only important for improving the utilization of expensive spectrometers but is also crucial for studies of biomolecules that are weakly soluble or marginally stable. Furthermore, the introduction of higher field

magnets requires more frequent sampling (to avoid aliasing) and a concomitant increase in the number of samples needed to probe longer evolution periods.

A promising approach is that of reduced dimensionality experiments, which mainly rely on coupling evolution periods in the indirect dimensions [3–5]. Several methods have been introduced to facilitate the processing and analysis of data collected in this manner. The GFT method uses a G-matrix based on phase encoding of the linear combination of the coupled evolution periods prior to Fourier transformation in order to retrieve specific chemical shift information [6]. The back-projection reconstruction (BPR) method uses the relation between the time and frequency dimensions embodied in the Fourier transformation: a cross-section in the time domain through the origin with an angle α corresponds to a projection in the frequency domain onto an axis passing through the origin at the same angle. Various projections are reconstituted into a higher dimensional spectrum using BPR [7]. The

* Corresponding author. Fax: +1 860 679 3408.
E-mail address: hoch@uconn.edu (J.C. Hoch).

related filtered-BPR method applies an apodization step in the inverse Radon transform prior to the reconstruction of the spectrum [8]. The APSY method [9] analyzes each projection separately and then uses the information from all projections to determine peak frequencies using vector algebra. The EVOCOUP approach [10] similarly does not attempt reconstruction of the spectrum but uses singular value decomposition (SVD) to extract peak information from the known linear combination of chemical shifts. The HIFI method [11] uses Bayesian methods to analyze individual projections and utilizes prior knowledge of shift distributions to guide the choice of projection angle.

An alternative to reduced dimensionality is to treat the aforementioned cross-sections as incomplete multidimensional data. This approach requires methods for computing the spectrum that do not rely on uniformly sampled data. Suitable methods include maximum entropy (MaxEnt) reconstruction [12,13] maximum likelihood methods [14], multi-way decomposition [15,16], and Lagrange interpolation FT [17]. These methods have previously been employed to reduce experiment time by non-uniform sampling of the indirect dimensions using various strategies for selecting subsets of the data. Furthermore, these methods can also be used in conjunction with reduced dimensionality approaches to enhance their utility [18]. Despite the obvious connection between reduced dimensionality methods and non-uniform sampling, the two have so far not been systematically compared. In this paper, we demonstrate that MaxEnt reconstruction is a general platform for the reconstruction of non-uniformly sampled data regardless of sampling strategy. We explicitly use the term “reconstruction” in conjunction with “MaxEnt”, to emphasize that we are computing the spectrum. In contrast, parametric methods of spectrum analysis attempt to extract signal parameters, such as frequencies, amplitudes, linewidths and phases. Although these methods can be used to construct a spectrum, in general such a spectrum is not necessary, and the table of signal parameters can be used for further processing, such as assignment. A drawback of parametric methods, however, is that they make assumptions about the signal, and when these assumptions are not strictly met the derived signal parameters are subject to bias. MaxEnt reconstruction makes no assumptions about the nature of the signal, and for this reason is generally more robust for determining accurate frequencies of signal components than parametric methods [13].

Considering coupling of evolution periods as a non-uniform sampling method, MaxEnt will enable us to distinguish between artifacts introduced by the processing method and those that are intrinsic to the sampling method. We show that reduced dimensionality data, whether processed by MaxEnt reconstruction or BPR, lead to similar artifacts. These artifacts arise from the regular nature of the sampling, and not from the algorithm used to process the data. Destroying this regularity by adding a degree of randomness to the sampling diminishes the structure of the artifacts, greatly improving spectral quality.

2. Theory

The theory of maximum entropy and its implementation into the Rowland NMR Toolkit has been extensively reviewed [19,20] and therefore only a brief summary will be given here. Similarly the theory of back-projection reconstruction is well documented [4] and only an overview will be given here.

2.1. Maximum entropy

MaxEnt reconstruction solves the inverse problem in which the information content of a trial spectrum is minimized while ensuring that its inverse discrete Fourier transform (DFT) agrees with the observed data to within the experimental error of the data. The agreement of the reconstruction with the data is measured using an unweighted χ^2 statistic:

$$C(\mathbf{f}) = \sum_{k=0}^{N-1} |m_k - d_k|^2, \quad (1)$$

where m_k is the mock data given by the inverse DFT of the trial spectrum with elements f_i and d_k is the experimental data. The measure of the information content is based on an entropy functional [13] resembling:

$$S(f) = - \sum_{i=0}^{M-1} \frac{|f_i|}{def} \log \left(\frac{|f_i|}{def} \right) \quad (2)$$

where M is the number of points in the spectrum, and def is a scale factor. By maximizing $S(f)$ subject to the constraint $C(f) \approx C_0$, where C_0 is an estimate of the experimental error, MaxEnt reconstruction determines the spectrum containing the least information, consistent with the measured data.

2.2. Back projection reconstruction

BPR relies on the relation between the time and frequency domains through the projection–cross-section theorem [4]. For a two-dimensional spectrum \mathbf{F} , the projection $\mathbf{P}_\alpha(\omega)$ onto a line passing through the origin at an angle α is given by:

$$\mathbf{P}_\alpha(\omega) = \int_{-\infty}^{\infty} f(\omega \cos \alpha - s \sin \alpha, \omega \sin \alpha + s \cos \alpha) ds \quad (3)$$

The cross-section $\mathbf{S}_\alpha(t)$ at angle α through the time domain data \mathbf{D} is given by:

$$\mathbf{S}_\alpha(t) = \mathbf{D}(t \cos \alpha, t \sin \alpha). \quad (4)$$

The projection cross-section theorem says that if \mathbf{F} is the Fourier transform of \mathbf{D} , then \mathbf{P}_α is the one-dimensional Fourier transform of \mathbf{S}_α .

Back-projection algorithms use this relation. Starting with a cross-section \mathbf{S}_α , one applies a Fourier transform to obtain \mathbf{P}_α which is “back-projected” onto the two-dimensional frequency plane: the value of $\mathbf{P}_\alpha(\omega)$ is replicated

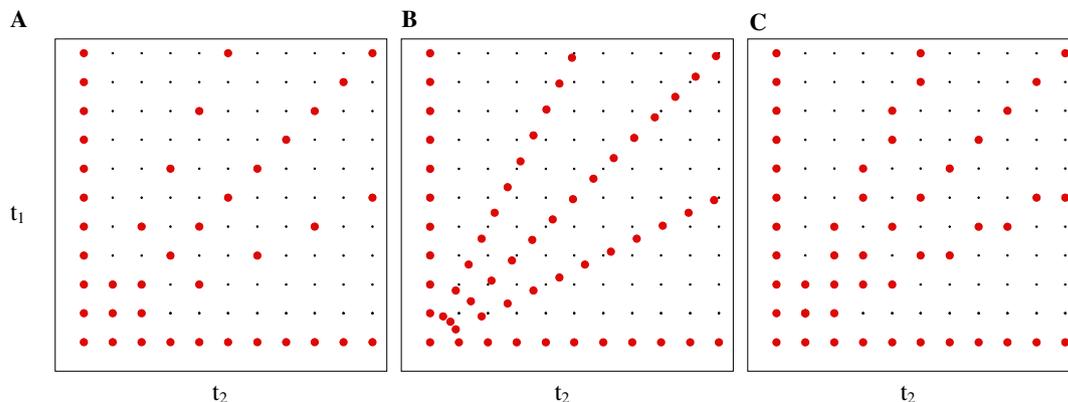


Fig. 1. Sampling the time domain with coupled evolution periods. The small dots define the Nyquist grid and the large dots represent sampled data points. (A) Radial on-grid sampling. (B) Radial off-grid sampling used in the BPR method. (C) An on-grid approximation of the sampling shown in (B), appropriate for processing with MaxEnt.

along the line perpendicular to the line through the origin with angle α and intersecting the line at the point ω . Back-projections corresponding to different cross-sections are combined to yield a spectrum. Various BPR algorithms differ in the way they combine these back-projections. Filtered back-projection algorithms apply a weighting function to \mathbf{S}_α prior to the computation of \mathbf{P}_α [8].

The sample points belonging to the cross-section $\mathbf{S}_\alpha(t)$ all lie on a ray emanating from the origin at an angle α . We refer to this as a radial sampling scheme. Note that sampling using coupled evolution periods is a form of radial sampling. The n th sample along a given ray then has evolution times:

$$t_1 = n\Delta t \sin \alpha, \quad (5)$$

$$t_2 = n\Delta t \cos \alpha, \quad (6)$$

where Δt is the inverse of the spectral width along the ray. In order to avoid aliasing it is generally necessary to choose Δt so that $1/\Delta t$ is larger than the spectral width in F_1 or F_2 . Normally when collecting multi-dimensional data the evolution times are all multiples of fixed dwell times Δt_1 , Δt_2 , etc.; such points determine a uniform grid in the time domain which we call the Nyquist grid. In radial sampling, however, most choices of α lead to evolution times that do not lie on this grid. We refer to this situation as “off-grid” sampling. One can approximate radial sampling using on-grid samples by adjusting the evolution times to the nearest grid point (Fig. 1).

3. Methods

3.1. Experimental data

Triple resonance data (HNCO) was collected using a $^{15}\text{N}/^{13}\text{C}$ labeled ubiquitin sample obtained commercially (AVR Protein). The number of data points collected was 512 in the proton, 128 in the carbon and 52 in the nitrogen dimensions. The spectral windows in both indirect dimensions were set to 1885.7 Hz. This was done to make the datasets more suitable for comparison with radially sampled

data. Subsets of the sampled data points in this master dataset were extracted to study non-uniform sampling. For BPR a 45° projection was also collected using the Varian projection data collection method incorporated in BIOPACK (ghn_co). Note that quadrature detection also yields the corresponding negative projection [4]. The Varian implementation determines the spectral width of the projection from that of the first indirect dimension (carbon in this case).

The ubiquitin sample was prepared in a 50 mM NaPO_4 buffer at pH 6.5 with added 1 mM NaN_3 solution. The ubiquitin sample concentration was ca. 1 mM, and the data was collected at 23 °C.

3.2. Synthetic data

A synthetic two-dimensional dataset was created, representing the two indirect dimensions of a 3D dataset. The linewidth of the peaks was set to 20 Hz in each dimension. The dwell times in the two dimensions were set to correspond to spectral widths of 2000 Hz. The maximum number of points in each dimension was set to 128. The number and position of synthetic peaks together with the amount of noise varied for the different applications. Dataset A has peaks at F_1 and F_2 frequencies of $(-170, 0)$ and $(200, -80)$ Hz, respectively, with amplitudes of 10. No noise was added to this dataset. The corresponding cross-section at 63° with respect to the F_2 axis, was also created for this dataset. Dataset B includes four additional peaks appearing at F_1 and F_2 coordinates of $(-650, -700)$, $(-685, -700)$, $(500, -700)$, $(360, 850)$ Hz. The amplitudes of all peaks were set to 10 except the peak at $(500, -700)$ Hz which was set to -10 . In this dataset Gaussian noise with an average of 0 and standard deviation of 2 was also added. Three instances of dataset B were created using different pseudo-random noise sequences.

3.3. Sampling schedules

The non-uniform sampling schedules were obtained using the *sampsched2d* application within the Rowland

NMR Toolkit. The sampling schedules for the synthetic dataset were calculated using 381 randomly sampled data points with exponentially weighted distributions corresponding to linewidths of 20, 10, and 0 Hz in both dimensions, the latter producing a completely random sampling schedule. The sampling schedule for the experimental data used 249 data points with exponentially weighted distributions corresponding to line widths of 15 and 25 Hz in the carbon and nitrogen dimensions, respectively. A more detailed account of this sampling method has previously been reported [21].

3.4. Data processing

All data processing was done using the Rowland NMR Toolkit [22]. The back-projection spectrum of the synthetic data was reconstructed using a program called *planemath* written for the Rowland NMR Toolkit. The *aim* parameter was used here in all MaxEnt reconstructions rather than the λ parameter [23]. *Aim* is related to C_0 by the expression $C_0 = N \times aim^2$. The reconstruction of synthetic dataset A used 0.1 for both *def* and *aim*, and the final output size of the spectrum was 128 in both dimensions. The MaxEnt reconstruction of synthetic dataset B had an output size of 512 in both dimensions and was reconstructed using values of 2.0 and 0.5 for *aim* and *def*, respectively.

For the experimental data the direct dimension of the HNCO experiment was processed by zero filling to 2048 (from 512) followed by DFT. In order to facilitate the comparison of different datasets, they were all extended to 512 data points in the indirect dimensions. The MaxEnt reconstruction of the experimental data employed values of 4.0 and 1.0 for *aim* and *def*, respectively.

The BPR algorithm implemented in *planemath* determines the spectral values at coordinates (ω_1, ω_2) by using linear interpolation to estimate the value of $\mathbf{P}_\alpha(\omega_1 \sin \alpha + \omega_2 \cos \alpha)$. The addition of the individual back-projected planes was also performed using *planemath*.

3.5. Data analysis

Processed spectra were peak-picked and assigned using a program written for the Rowland NMR Toolkit (*2dpeak*). A peak was defined as a maximum in the frequency domain with respect to all neighboring data points. The detected peaks were fitted to Lorentzian lineshapes using the *nonlin-LS* program from the NMRPipe software package [24]. The fitted frequencies and linewidths were used as quality measures. The root-mean-square (rms) error of the fitted frequency, compared to a reference value is reported. The reference values for the synthetic data were the known input frequencies, for the experimental data the fitted frequencies obtained using the master dataset were used, together with the average linewidth in each dimension. For the synthetic data a splitting parameter was used to characterize the separation of two nearly overlapping peaks. The splitting parameter is defined as: $\Delta = 1 - 2C/$

$(A + B)$, where A and B are the peak amplitudes of the two peaks and C is the lowest point between them. Another measure of quality used is the signal to noise ratio (S/N) which is the ratio of the fitted amplitude of the smallest real peak to the standard deviation of the noise. The noise region was defined as areas other than ± 30 Hz from known peaks. For the experimental data the weak peak was determined as the weakest detectable/assignable peak in the master dataset. The signal to artifact ratio, S/A , is calculated as the ratio of the weakest peak (signal amplitude) and the highest artifact peak.

4. Results

In Fig. 1, we show the differences between various approaches to radial sampling. In this section synthetic data will be used to study the impact of these methods on spectral reconstruction. Subsequently the impact of introducing a degree of randomness will be shown. Finally, experimental data will be used to directly compare MaxEnt reconstruction of irregular sampling with BPR of radially sampled data.

4.1. On vs. off-grid

A distinguishing characteristic of BPR spectra are the ridge artifacts, illustrated in Fig. 2 for the noiseless synthetic dataset A. The contour levels were chosen to clearly show the most significant ridge artifacts. These artifacts are due to ambiguities introduced by the sampling scheme and we see, by comparison of panels A and B that the BPR is nearly identical to MaxEnt reconstruction when the same data samples are used. Note however that the peaks are more clearly defined in the MaxEnt reconstruction; in the BPR spectrum the real peak at $F_1 = -170$ Hz, $F_2 = 0$ Hz, has similar amplitude to the artifact peak with same F_1 frequency at $F_2 = 80$ Hz, whereas in the MaxEnt spectrum the artifact peak is significantly weaker than the real peak.

In addition to the expected ridges passing through the peaks at angles perpendicular to the cross-section angle we also note the presence of folded ridges due to the small spectral window in the off-axis projection; the dwell time between two data points in this projection corresponds to a spectral width of 1788 Hz, which corresponds to spectral widths in F_1 and F_2 of about 800 and 1600 Hz, respectively. To maintain radial sampling with a larger spectral width in the off-axis projection we can no longer sample on-grid. In panel C we see that off-grid sampling has the desired effect on the spectrum. This mode of sampling cannot be reconstructed using the Rowland NMR Toolkit implementation of MaxEnt due to technical difficulties associated with processing off-grid data. This sampling scheme may, however, be approximated by an on-grid version where data points on the grid near the ray are collected (see Fig. 1C). Such a cross-section cannot be transformed using conventional processing schemes, but can be reconstructed using MaxEnt. Panel D shows that this sampling method similarly

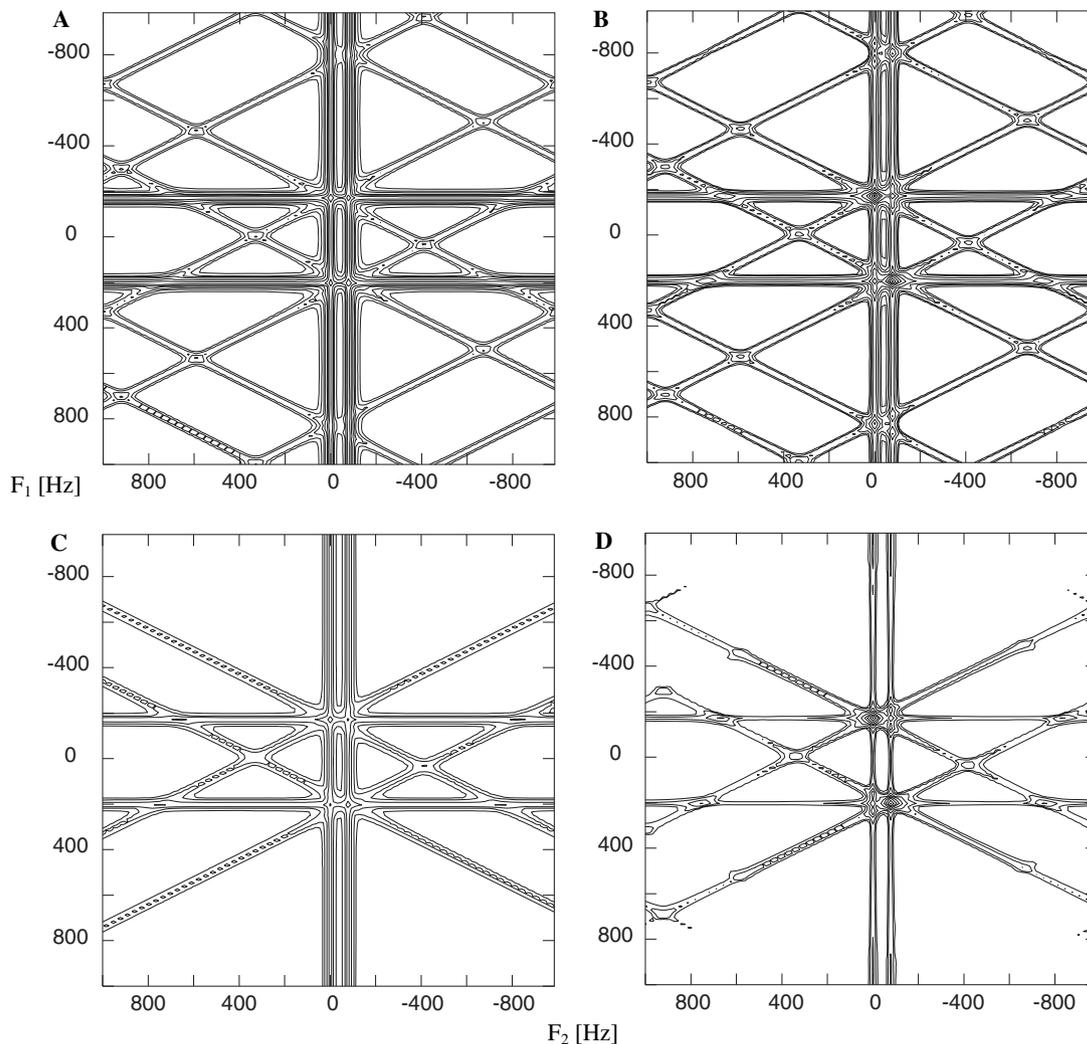


Fig. 2. Back-projection (A and C) and MaxEnt (B and D) reconstruction using three projection angles (0° , 90° and 63.43°). (A) Back-projection reconstruction of on-grid data, where the tilted projection is collected at integer increments of the Nyquist grid in t_1 and t_2 of 1 and 2, respectively. (B) MaxEnt reconstruction using the same sampling scheme as A. (C) Back-projection reconstruction using the same projection angles sampled off-grid. (D) MaxEnt reconstruction using the on-grid approximation of the samples used in C. (C and D) Larger spectral widths, compared to A and B, along the projection and therefore avoid aliasing artifacts.

suppresses folding artifacts and is a good approximation of off-grid sampling. The presence of folding artifacts at high frequencies in the on-grid method is due to the distance between two neighboring points along this cross-section never being as close as those in the off-grid sampling method. Note, however, that off-grid radial sampling requires more data samples to achieve the same resolution as the on-grid methods. Also we note again that the peak positions are better defined in the MaxEnt reconstruction.

4.2. Regular vs. random

It is clear that the source of ridge artifacts is closely related to the regular nature of radial sampling and the associated ambiguities introduced, irrespective of the data samples being collected on or off-grid or the processing method used. Destroying this regularity by introducing a degree of randomness should therefore diminish these arti-

facts and improve spectral quality. Fig. 3 shows the distribution of sampling artifacts for a non-decaying peak of 0 frequency, computed using the DFT. The figure illustrates the relatively high intensity of the ridge artifacts when data is sampled along specific angles, and how these are diminished by removing the regularity of the data sampling. Random samples, however, cannot be reconstructed using conventional processing methods. The MaxEnt method therefore appears to be a suitable platform for the comparison of these different modes of sampling.

Dataset B was used to systematically study the impact of sampling on spectral quality. Recovery of the known parameters of the synthetic peaks was used to test the spectral qualities of the reconstructions. Two of the peaks are placed close to each other to measure resolution. One negative peak is inserted along the same dimension as the two overlapping peaks to test the ability of the methods to deal with a few negative peaks in the presence of many positive

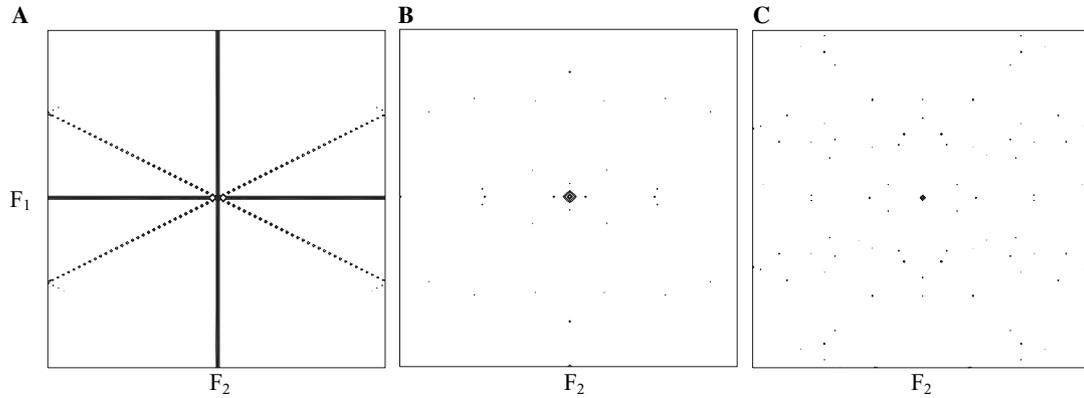


Fig. 3. Fourier transform of sampled data points according to on-grid radial sampling of three projections at 0° , 90° , 63.43° (A), exponentially weighted random sampling corresponding to a decay of 20 Hz (B) and random sampling (C), showing the artifact distribution of a single peak with 0 frequency and 0 linewidth. The same contour levels are used in all spectra. Spectral widths are from dataset A (see Section 3).

ones. Peaks close to the edges of the spectrum are also used to test for aliasing. Synthetic noise was added, as described in Section 3, to measure sensitivity. The sampling methods used in addition to the on-grid approximation to radial sampling are; random sampling with an exponentially weighted distribution corresponding to the line width of the peaks (20 Hz), half the line width of the peaks (weight-

ed less towards shorter evolution periods), and no weighting, giving completely random data samples. The same number of data samples was collected for each method. Note that all data samples are on-grid. The data were generated three times for each method using three different pseudo-random sequences of the added noise. Fig. 4 shows one of the resulting spectra, each having the same underly-

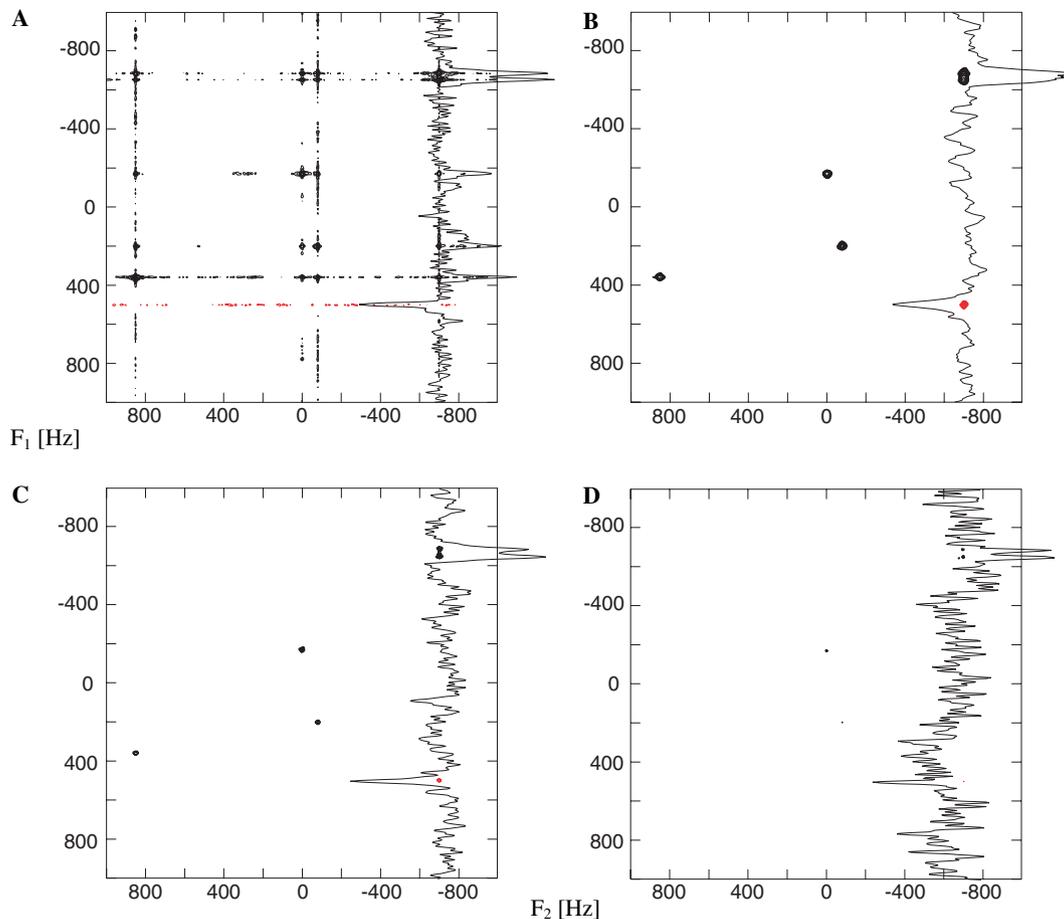


Fig. 4. MaxEnt reconstruction of synthetic data using 381 data points according to different sampling strategies. 1D trace at an F_2 frequency of 699 Hz is superimposed. (A) On-grid approximation of radial sampling. (B) Random sampling with exponentially weighted distribution corresponding to the LW of the synthetic peaks (20 Hz). (C) Random sampling with exponentially weighted distribution equal to half the LW of the synthetic peaks (10 Hz). (D) Randomly sampled data. The contour levels shown are five times the standard deviation of the noise and higher.

ing noise sequence (using the same random seed). The figure shows contour levels at 5 times the standard deviation of the noise in the spectrum and higher. We see the clear presence of ridge artifacts at these contour levels in panel A. The resolution/sensitivity trade-off of the different random distributions of sample points is also amply illustrated.

Table 1 lists the computed measures of spectral quality together with average values and the corresponding standard deviations. The table includes data for frequency accuracy reported as the average rms error of the frequency of each peak fitted to a Lorentzian lineshape (see Section 3) compared to that of the input frequency. Similarly average fitted values for line width are given. S/N and S/A are also reported as described in Section 3. Note that the noise distribution in the on-grid approximation of radially sampled data does not have a Gaussian distribution due to the ridge artifacts. We also note that the negative peak in the reconstruction of this sampling method is not identified as a peak as it is present only as a saddle point, being a maximum in one dimension and a minimum in the other. This may lead to overestimated values of the S/N and S/A ratios for this method. Values for the average evolution time for each sampling scheme are also reported. The results for a hybrid method are also included in the table but not shown in Fig. 4. This method involves sampling along 0° and 90° cross-sections as in the radial sampling method but the data samples corresponding to the off-axis cross-section is distributed according to an exponentially weighted random scheme with a decay corresponding to 20 Hz in both dimensions.

4.3. Comparison of MaxEnt reconstruction vs. BPR for experimental data

The on-grid approximation of cross-sections enabled us to compare radial sampling with random based sampling using the same reconstruction technique, eliminating possible biases arising from the specific reconstruction algorithm used. Practically, however, it is of interest to compare radially sampled data reconstructed using BPR with a non-regular sampling method reconstructed using MaxEnt, using experimental data. Here we use a triple resonance NMR experiment (HNCO) on a readily available protein (ubiquitin). An extensive uniformly sampled dataset was collected and is used as a reference for comparison with the radially and non-uniformly sampled datasets. A subset of this data was sampled non-uniformly, based on an exponentially decaying distribution, and the spectrum was reconstructed using MaxEnt (see Section 3). A separate off-grid projection was collected at 45° , using the spectral width of the first indirect dimension (^{13}C). These data were combined with the 0 and 90° cross-sections from the master dataset for BPR. Both fast methods (with respect to the master data set) contain the same number of sample points, requiring roughly the same experiment time. The uniformly sampled dataset in Fig. 5 consists of 6656 data points, whereas

the two fast methods use only 249 data points, which is 3.7% of the master dataset. The results are shown in Fig. 5 and Table 2. The cross-sections across the spectra in Fig. 5 are at the nitrogen frequency of the weakest peak found in the master dataset.

From the spectral reconstructions we see that BPR gives rise to several artifacts and fails to detect a weak peak in the presence of stronger peaks due to the base line artifacts (see cross-section in Fig. 5). Table 2 reveals an improvement on all measured spectral qualities, when using exponentially weighted random sample points reconstructed using MaxEnt compared to radial sampling reconstructed using BPR. We note that the comparatively low S/A ratio of the DFT of the linearly sampled data is due to the highest artifact being a truncation artifact.

5. Discussion

Reducing experiment time in NMR has recently been the focus of many investigations [2,6,9–11,17,25]. Coupling evolution times has emerged as a popular method to achieve this. This idea started with the accordion experiment [26], where the evolution period is coupled to a mixing period. This was formally extended to coupled evolution periods in the pseudo multiple-quantum spectroscopy experiment [27]. This experiment presaged the fundamental concepts behind RD experiments. These methods subsequently evolved into RD [5] and GFT [6] methods. The latest development has been the introduction of BPR [4], which in contrast to its predecessors attempts to reconstruct the full multidimensional spectrum from a few reduced dimensionality spectra. There are also recent developments involving methods which are not restricted by coupling of evolution times in the time domain [13,19,28], referred to here as non-uniform sampling methods.

In this study, we have shown that the two general approaches (coupled vs. non-uniform sampling) are closely related and differ only in the sampled time points. We have shown that radial sampling can be used in conjunction with either BPR or MaxEnt reconstruction. We find however that the spectral reconstruction using MaxEnt is superior to that using BPR. It is also clear from the results that regularities in the data samples lead to an abundance of artifacts which can make the identification of weak peaks difficult and lead to misrepresentation of peaks with different phases. Removing the regularities in the data sampling solves these problems and provides a degree of freedom in choosing data samples to enhance resolution or sensitivity. We see improved frequency accuracy in radial sampling compared to random sampling. We attribute this to the highly resolved 0° and 90° projections. This is confirmed when random sampling is used in conjunction with radial sampling, in the hybrid method. The average evolution period follows the hybrid properties confirming that the S/N and linewidths are closely related to this parameter, independent of frequency accuracy. The apparent relation-

Table 1
Spectral quality from synthetic data reconstructed with MaxEnt using 381 data points sampled according to different strategies

Sampling method	Peaks found	Frequency error (Hz)				Line width (Hz)				S/N	S/A	Splitting parameter	Average evolution time (ms)				
		F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2								
Radial	5	<i>0</i>	0.61	<i>0.25</i>	0.66	<i>0.13</i>	26.78	<i>5.09</i>	33.06	<i>1.84</i>	7.32^a	<i>0.59</i>	1.01	<i>0.10</i>	0.32	<i>0.05</i>	35.2
	5		0.53		0.51		28.48		30.93		7.72		1.05		0.36		
	5		0.90		0.71		21.05		34.12		6.64		0.90		0.27		
	5		0.41		0.77		30.80		34.12		7.59		1.09		0.34		
Exponentially weighted random, 20 Hz decay	6	<i>0</i>	0.85	<i>0.09</i>	1.07	<i>0.50</i>	37.08	<i>0.75</i>	38.14	<i>1.06</i>	10.74	<i>0.45</i>	2.46	<i>0.43</i>	0.14	<i>0.04</i>	25.2
	6		0.83		1.20		36.22		38.75		10.88		2.19		0.17		
	6		0.94		1.48		37.54		36.91		11.10		2.96		0.16		
	6		0.77		0.52		37.48		38.76		10.25		2.24		0.09		
Exponential weighted random, 10 Hz decay	6	<i>0</i>	1.18	<i>0.12</i>	1.07	<i>0.18</i>	26.59	<i>1.16</i>	30.83	<i>1.05</i>	7.49	<i>0.44</i>	1.69	<i>0.08</i>	0.42	<i>0.07</i>	36.7
	6		1.30		0.87		26.11		30.85		7.79		1.64		0.45		
	6		1.16		1.11		25.76		29.77		7.70		1.79		0.47		
	6		1.06		1.23		27.92		31.88		6.98		1.65		0.34		
Random	6	<i>0</i>	2.37	<i>1.50</i>	2.10	<i>0.60</i>	21.96	<i>1.76</i>	25.98	<i>2.72</i>	4.57	<i>0.04</i>	0.87	<i>0.06</i>	0.59	<i>0.06</i>	50.9
	6		2.23		2.77		20.57		24.73		4.57		0.85		0.65		
	6		0.93		1.89		21.39		29.10		4.61		0.83		0.60		
	6		3.93		1.63		23.93		24.11		4.53		0.93		0.53		
Hybrid, radial (0°, 90°) plus exponentially weighted random, 20 Hz decay	5	<i>0</i>	0.75	<i>0.50</i>	0.90	<i>0.07</i>	32.78	<i>1.75</i>	31.06	<i>1.38</i>	10.11	<i>0.20</i>	1.46	<i>0.05</i>	0.26	<i>0.03</i>	28.3
	5		0.35		0.98		32.76		29.64		10.12		1.44		0.28		
	5		1.31		0.89		34.53		31.15		10.30		1.43		0.23		
	5		0.60		0.84		31.04		32.40		9.90		1.51		0.27		

Data were measured for three reconstructions each with a different random noise sequence maintaining the same noise level. Radial refers to the on-grid approximation of off-grid radial sampling. Average values are given in bold and standard deviations are given in italics. The negative synthetic peak is not detected using radial sampling, as it appears as a saddle point.

^a Signal to noise value may be overestimated as one of the peaks is not detected.

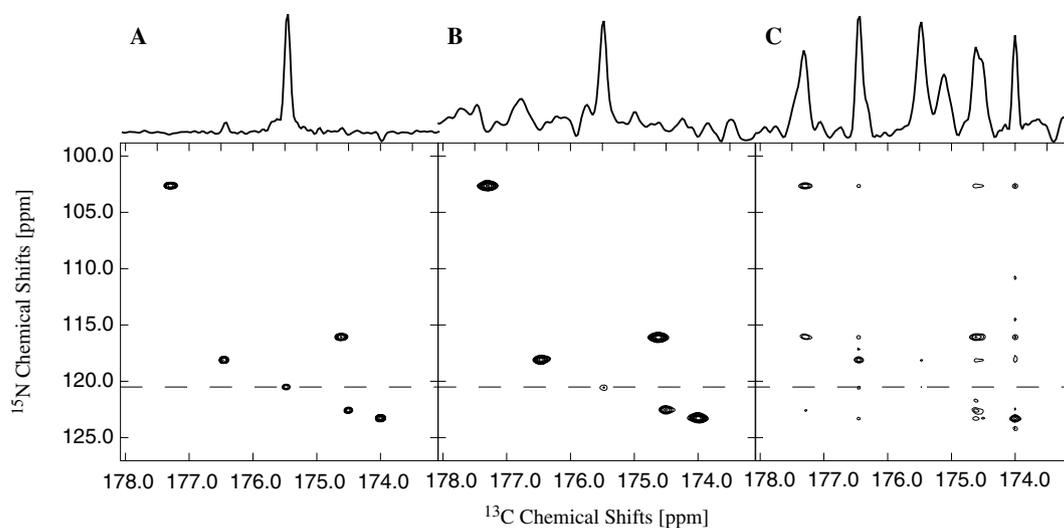


Fig. 5. Spectra obtained from experimental data (HNCO) for ubiquitin obtained using uniform, non-uniform and off-grid radial sampling. Figures show a nitrogen and carbon planar cross-section at a proton frequency of 8.14 ppm. The one-dimensional cross-sections through the plots are of the carbon row at the weakest peak (nitrogen frequency of 120.5 ppm). The contour levels were chosen to show the weakest peak. The cross-sections are scaled so that the highest and lowest amplitudes are aligned across the three spectra. (A) Using a full dataset, 6656 data points, processed using DFT (B) using 249 randomly sampled data points with an exponentially weighted distribution, reconstructed using MaxEnt (C) using 249 data radially sampled off-grid data points, at 0° , 90° and 45° projection angles, reconstructed using the back-projection method.

Table 2
Spectral quality from experimental data

Data samples	Sampling method	Reconstruction method	^{13}C LW (Hz)	^{15}N LW (Hz)	^{13}C Freq. error (Hz)	^{15}N Freq. error (Hz)	S/N	S/A
6656	Uniform	DFT	13.8	26.2	0.0	0.0	85.6	1.9
249	Non-uniform	MaxEnt	14.7	22.7	0.6	1.5	12.01	1.7
	Radial	BPR	27.5	39.2	5.1	1.0	6.30	0.7

Uniform refers to the full dataset (6656 data points). Non-uniform refers to sampling according to an exponentially weighted distribution (249 data points) corresponding to decay rates of 15 Hz for carbon and 25 Hz for nitrogen. Radial refers to sampling at angles 0° , 45° and 90° (249 data points).

ship between these properties raises concern for the use of off-grid sampling in BPR, which will bias the sampling towards shorter evolution periods (see Fig. 1). It is therefore reasonable to conclude there is a loss in resolution per sample point when sampling off-grid using this method. Indeed this problem will increase with the collection of additional projections. Considering that recent results suggest the need for very high number of projections for accurate reconstruction, the efficiency of using this method becomes questionable [8]. Methods such as the filtered back-projection may to some extent correct for the effective line broadening, but only with a loss in sensitivity per unit time.

The experimental results confirm those using the synthetic data but also highlight some of the additional problems with off-grid sampling in conjunction with BPR. The expected gain in frequency accuracy using radial samples is not reproduced in the experimental results. This can be explained by analyzing the experimental data, in Fig. 5 for example (panels A and B) we see two peaks at a carbon frequency of ca. 174.5 ppm. Since one of these peaks has a higher intensity, it is apparent that its ridge has moved the peak position of the weaker peak (Fig. 5C). This explains the larger error in the carbon dimension in Table 2. This

presents a possible pitfall, where the potential for high frequency accuracy of radially sampled data result in a falsely inflated confidence in the actual frequency accuracy. The synthetic data also shows that the linewidth and resolution is related to the average evolution time, but we note that this is not the case for the experimental results where the average evolution periods for the non-uniformly and radially sampled data are 22 and 25 ms, respectively (for the plane shown in Fig. 5). The discrepancy is most likely due to the same reasons as the discrepancy in frequency accuracy, i.e. overlap of ridges with real peaks. Another potential source of the discrepancy between the synthetic and experimental results may be the non-Lorentzian lineshape of the two-dimensional peaks, when reconstructed using BPR. It is true that each projection will maintain a Lorentzian lineshape, but this is not necessarily true for the back projection reconstruction.

Finally we note that in this work we have used a simple heuristic for generating suitable sampling schedules. While we have shown that this heuristic leads to more efficient data collection than BPR approaches, we have not shown that the schedules constructed are optimal in any sense. Preliminary evidence suggests that there is substantial room for improvement in sampling schedules, and thus

the improvements that we show here over BPR approaches are likely to be conservative. Clearly parameters such as sample concentration, molecular size (which impacts line-width), field strength, resonance frequencies and the nature of the experiment being performed are all factors that can influence the design of an optimal sampling schedule [29]. Exploration of these factors in the design of optimal sampling schedules is the subject of ongoing investigations.

6. Concluding remarks

Our comparison of MaxEnt and BPR spectra demonstrates that the artifacts commonly observed in the latter results from the regular nature of the data sampling. For equivalent data, however, MaxEnt reconstructions are generally superior to BPR. We also show that the time savings achieved by the sampling methods employing coupled evolution periods (i.e. projections) come at a price, either a loss in resolution or the introduction of artifacts, and frequently both. The artifacts are intimately tied to the regular nature of the sampling, and we demonstrate that introducing randomness into the selection of sampling points greatly diminishes the artifacts, enabling far more efficient data collection. Based on these findings we conclude that non-uniform sampling based on a random distribution combined with MaxEnt reconstruction is preferable to RD or BPR. Further improvements can be anticipated through the design of optimal sampling schedules.

Acknowledgments

We thank Dr. Ēriks Kupče for help with the Varian implementation of radial sampling and its processing, and Dr. Mark Maciejewski for data collection. Financial support from the National Institutes of Health, via Grants GM047467 (Gerhard Wagner, PI), GM072000 (Amos Ron, PI), and RR-020125 (Jeffrey Hoch, PI) is gratefully acknowledged. This is a contribution from the NMRA consortium.

References

- [1] D. Malmodin, M. Billeter, High-throughput analysis of protein NMR spectra, *Prog. Nucl. Mag. Res. Spectrosc.* 46 (2005) 109–129.
- [2] Ē. Kupče, R. Freeman, New methods for fast multidimensional NMR, *J. Biomol. NMR* 27 (2003) 101–113.
- [3] T. Szyperski, G. Wider, J.H. Bushweller, K. Wüthrich, Reduced dimensionality in triple resonance NMR experiments, *J. Am. Chem. Soc.* 115 (1993) 9307–9308.
- [4] Ē. Kupče, R. Freeman, Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy, *J. Am. Chem. Soc.* 126 (2004) 6429–6440.
- [5] K. Ding, A.M. Gronenborn, Novel 2D triple resonance NMR experiments for sequential resonance assignments of proteins, *J. Magn. Reson.* 156 (2002) 262–268.
- [6] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, *J. Am. Chem. Soc.* 125 (2003) 1385–1393.
- [7] E. Kupče, R. Freeman, The Radon Transform: a new scheme for fast multidimensional NMR, *Concepts Magn. Reson.* 22A (2004) 4–11.
- [8] B. Coggins, R.A. Venters, P. Zhou, Filtered backprojection for the reconstruction of a high-resolution (4,2)D CH₃-NH NOESY spectrum on a 29 kDa protein, *J. Am. Chem. Soc.* 127 (2005) 11562–11563.
- [9] S. Hiller, F. Fiorito, K. Wüthrich, Automated projection spectroscopy (APSY), *Proc. Natl. Acad. Sci. USA* 102 (2005) 10876–10881.
- [10] D. Malmodin, M. Billeter, Signal identification in NMR spectra with coupled evolution periods, *J. Magn. Reson.* 176 (2005) 47–53.
- [11] H.R. Eghbalnia, A. Bahrami, M. Tonelli, K. Hallenga, J.L. Markley, High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection, *J. Am. Chem. Soc.* 127 (2005) 12528–12536.
- [12] J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling, S.J.P. Worrall, Exponential sampling: an alternative method for sampling in two dimensional NMR experiments, *J. Magn. Reson.* 73 (1987) 69–77.
- [13] A.S. Stern, K. Li, J.C. Hoch, Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear prediction extrapolation and maximum-entropy reconstruction, *J. Am. Chem. Soc.* 124 (2002) 1982–1993.
- [14] R.A. Chylla, J.L. Markley, Theory and application of the maximum likelihood principle to NMR parameter estimation of multidimensional NMR data, *J. Biomol. NMR* 5 (1995) 245–258.
- [15] T. Luan, V.Y. Orekhov, A. Gutmanas, M. Billeter, Accuracy and robustness of three-way decomposition applied to NMR data, *J. Magn. Reson.* 174 (2005) 188–199.
- [16] V. Tugarinov, L.E. Kay, I. Ibraghimov, V.Y. Orekhov, High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition and multidimensional decomposition, *J. Am. Chem. Soc.* 127 (2004) 2767–2775.
- [17] D. Marion, Fast acquisition of NMR spectra using Fourier transform of non-equispaced data, *J. Biomol. NMR* 32 (2005) 141–150.
- [18] Z.J. Sun, S.G. Hyberts, D. Rovnyak, S. Park, A.S. Stern, J.C. Hoch, G. Wagner, High-resolution aliphatic side-chain assignments in 3D HCcoNH experiments with joint H–C evolution and non-uniform sampling, *J. Biomol. NMR* 32 (2005) 55–60.
- [19] S. Sibisi, J. Skilling, R.G. Brereton, E.D. Laue, J. Staunton, Maximum entropy signal processing in practical NMR spectroscopy, *Nature* 311 (1984) 446–447.
- [20] J.C. Hoch, A.S. Stern, *NMR Data Processing*, Wiley-Liss, New York, 1996.
- [21] P. Schmieder, A.S. Stern, G. Wagner, J.C. Hoch, Application of nonlinear sampling schemes to COSY-type spectra, *J. Biomol. NMR* 3 (1993) 569–576.
- [22] J.C. Hoch, A.S. Stern, *RNMR Toolkit, Version 3*, 2005.
- [23] P. Schmieder, A.S. Stern, G. Wagner, J.C. Hoch, Quantification of Maximum Entropy Spectrum Reconstructions, *J. Magn. Reson.* 125 (1997) 332–339.
- [24] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277–293.
- [25] L. Frydman, T. Scherf, A. Lupulescu, The acquisition of multidimensional NMR spectra within a single scan, *Proc. Natl. Acad. Sci. USA* 99 (2002) 15858–15862.
- [26] G. Bodnhausen, R.R. Ernst, The accordion experiment, a simple approach to three dimensional NMR spectroscopy, *J. Magn. Reson.* 45 (1981) 367–373.
- [27] O.W. Sørensen, R.R. Ernst, Remote heteronuclear correlation via pseudo multiple-quantum spectroscopy, *J. Magn. Reson.* 55 (1983) 338–343.
- [28] V.Y. Orekhov, I.V. Ibraghimov, M. Billeter, MUNIN: a new approach to multi-dimensional NMR spectra interpretation, *J. Biomol. NMR* 20 (2001) 49–60.
- [29] D. Rovnyak, D.P. Frueh, M. Sastry, Z. 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, *J. Magn. Reson.* 170 (2004) 15–21.