Published in final edited form as:

J Biomol NMR. 2017 June ; 68(2): 139–154. doi:10.1007/s10858-017-0103-z.

Interpolating and extrapolating with hmsIST - seeking a $t_{\max}$ for optimal sensitivity, resolution and frequency accuracy

Sven G. Hyberts, Scott A. Robson, and Gerhard Wagner*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Abstract

Non-Uniform Sampling (NUS) has the potential to exploit the optimal resolution of high-field NMR instruments. This is not possible in 3D and 4D NMR experiments when using traditional uniform sampling due to the long overall measurement time. Nominally, uniformly sampled (US) time domain data acquired to a maximum evolution time $t_{\max}$ can be extended to high resolution via a virtual maximum evolution time $t_{\max}^*$ while extrapolating with linear prediction or iterative soft thresholding (IST). At the high resolution obtainable with extrapolation of US data, however, the accuracy of peak positions is compromised as observed when comparing inter- and intra-residue peaks in a 3D HNCA experiment. However, the accuracy of peak positions is largely improved by spreading the same number of acquired time domain data points non-uniformly over a larger evolution time to an optimal $t_{\max}$ followed by extrapolation to a total $t_{\max}^*$ and processing the data with an appropriate reconstruction method, such as hmsIST. To explore the optimum value of experimentally measured $t_{\max}$ to be reached non-uniformly with a given number of sampling points we have created test situations of time-equivalent experiments and evaluate sensitivity and accuracy of peak positions. Here we use signal-to-maximum-noise ratio as the decisive measure of sensitivity. We find that both sensitivity and resolution are optimal when PoissonGap sampling to a $t_{\max}$ of about $\frac{1}{2}T_2^*$. Digital resolution is further enhanced by extrapolating the range of acquired time domain data to $2T_2^*$ but without measuring experimental points beyond $\frac{1}{2}T_2^*$.

Keywords

non-uniform sampling; NUS; protein backbone chemical shift assignments; maximum entropy reconstruction; iterative soft threshold reconstruction; reduced time multidimensional NMR spectroscopy

Introduction

Non-uniform sampling (NUS) is now frequently used for multi dimensional NMR spectroscopy. When sufficient intrinsic sensitivity is present, and there is a need for both a

* gerhard_wagner@hms.harvard.edu, phone 617-432-3213, fax 617-432-4383.

Compliance with Ethical Standards

Conflict of Interest: GW receives a honorarium from Springer as Editor in Chief of JBNMR. The authors declare that they have no further conflicts of interest

large sweep width and high digital resolution to be obtained with many consecutive indirect increments at uniform sampling, NUS can provide effective time saving compared to traditional uniform sampling (US). Consequently, for a given amount of measuring time, NUS enables covering larger sweep widths, higher digital resolution, or a combination of both. Programs written to reconstruct NUS data utilize a number of different algorithms. These algorithms are either based on Claude Shannon's work (Shannon 1948) on Maximum Entropy reconstruction (MaxEnt) (Balsgart and Vosegaard 2012; Delsuc and Mallavin 1998; Donoho et al. 1992; Hoch and Stern 1996; Hyberts et al. 2007; Laue et al. 1985; Paramasivam et al. 2012; Worley 2016), on Jan Högbom's work (Högbom 1974) in the CLEAN algorithm (Coggins and Zhou 2008; Kupce and Freeman 2005; Matsuki et al. 2009), work by (Candes and Walkin 2008; Donoho 2006; Drori 2007) on Compressed Sensing (Holland et al. 2011; Hyberts et al. 2012; Kazimierczuk and Orekhov 2011; Stern et al. 2007; Sun et al. 2015), or, by signal or vector decomposition (Mandelshtam et al. 1998; Orekhov et al. 2001; Orekhov and Jaravine 2011). Alternatively the NUS obtained spectra may be used with transform only, either by radial transform (Coggins and Zhou 2006; Kim and Szyperski 2003; Kupce and Freeman 2008) or by discrete FT (DFT) (Kazimierczuk et al. 2010).

Whether it is possible to gain sensitivity by skewing NUS to regions where the content of the FID is highest has been discussed in the literature (Barna et al. 1987; Hyberts et al. 2013; Rovnyak et al. 2011). Of course, any comparison of sensitivity or frequency accuracy needs to compare situations with time-equivalent experiments. For example, with an overall NUS sampling density of 25%, fourfold more scans per sampled point are collected with time-equivalent NUS, which has been shown to indeed increase the probability of detecting weak peaks above the noise (Hyberts et al. 2013). However, less has been reported on the effect of NUS on peak resolution, and how this is related to digital resolution. It has previously been suggested that the optimal sampling evolution time is πT_2 when sensitivity is no issue and $0.4\pi T_2$ ($= 1.256 T_2$) when sensitivity has to be concerned (Rovnyak et al. 2004). As for sensitivity discussed in this reference, it was shown that sampling beyond $1.256 T_2$ will add more noise than signal and thus decrease the signal-to-noise ratio (SNR). This analysis was based on uniform sampling, and did not address the question of how to best spread a given number of sampling points non-uniformly over a NUS range to obtain maximal sensitivity.

One of the realizations during this investigation concerning the sensitivity issue is that the distribution of the noise is different in a reconstructed NUS spectrum relative to that of traditional uniform sampling. More recently, Rovnyak and coworkers (Palmer MR et al. 2015) extended their theoretical analysis to distributing the acquired points according to NUS sampling profiles using a weighted sampling scheme, which needs to see the required implementation procedure. That is, the data point sampled at $1.256 T_2$ in an exponential matched sampling is sampled with a weight of 0.2846 of the initial point, not a probability of 0.2846 the initial point. The intrinsic SNR (iSNR) as defined by Palmer, Rovnyak et al. (Palmer MR et al. 2015) turns out to be very close to the traditionally measured SNR ($SN_{rms}R$) of a reconstructed NUS simulated time-equivalent practical implementation (see below). This is not to state that the iSNR and $SN_{rms}R$ are of the same nature as the rationale behind the two concepts is entirely different. The iSNR is theoretically predicted in the time domain while the simulations of the traditionally NUS spectra measure $SN_{rms}R$ values in the

frequency domain. Also the proposed schedulings are different. The iSNR is obtained from weighted acquisition of fractions of an integer, and the SN_{rmsR} is obtained from NUS points that are either acquired or not, and the frequency of non-acquired data points increases for later positions in the time domain.

As far as how NUS sampling strategies affect the resolution in terms of line width and frequency accuracy of the resulting peaks, we have found little published data. Celik & Shaka (Celik and Shaka 2010) comment that line widths and accuracy of frequency depend on the amount of noise (essentially the S/N) but do not evaluate this with respect to optimal sampling. On the other hand, as this is the case in the direct axis using uniform sampling, it is only hypothesized that the peak accuracy and the sharpness of line widths depend on the time position of the last sampled point and that this point is supposed to have as large a t_{max} as possible. t_{max} is the maximum evolution time, i.e. the time point of the last acquired data point in the FID, acquired either uniformly or non-uniformly.

In order to gather a clear answer to these posed questions we pursued two aspects. (i) We used extensive simulations to explore the effect of uniform and non-uniform sampling strategies for a given number of sampling points (time-equivalent NUS) on the accuracy of peak positions and sensitivity using various evolution times, under high and low S/N ratios. (ii) We used experimental data to again evaluate parameters for various acquisition situations. We find that sampling to an evolution time of $t_{max} \cong \frac{1}{2}T_2^*$ and extrapolating to about $2T_2^*$ is optimal with respect to both sensitivity and peak accuracy. In the following, we will make no difference between T_2^* and T_2 , as the two will have the same value in simulations where no inhomogeneity is accounted for, and more importantly we want to avoid confusion as we will make use of the asterisc (*) to differentiate between t_{max} as the actual time point last obtained and t_{max}^* as the last time point used for reconstruction. Extrapolation by means of NUS type reconstruction algorithms have briefly been approached with respect to sensitivity by (Hyberts et al. 2013; Stern et al. 2007) and for purpose of resolution enhancement (Lee et al. 2017).

Results

Extrapolation of US data yields narrow line shapes but lacks peak accuracy

When examining the effect of extrapolating US time domain data points with hmsIST on HNCA spectra we realized that the intra and inter-residue cross peaks did not line up at the high digital resolution obtainable (Fig. 1a). This inaccuracy of peak positions can however, largely be reduced by spreading the same number of data points non-uniformly over a large evolution time range (Fig. 1b). This motivated us to further explore what is the optimum range of NUS acquired data points (t_{max}) followed by extrapolation to t_{max}^* .

Alternate ways of measuring SNR

Measuring and defining SNR of NUS acquired spectra requires careful consideration primarily because the underlying noise is sampled non-uniformly. Obviously, certain time data points will be absent. Transformed into the frequency domain, the noise will appear non-Gaussian after IST reconstruction. The non-Gaussian character can be described by

calculating the kurtosis value, the fourth standardized statistical moment. In (Palmer MR et al. 2015), the group extend the work of (Rovnyak et al. 2004), now to include NUS sampling profiles. Both studies calculate iSNR formulae in the time-domain. The 2015 work assumes that data can be acquired at each time point with a fractional weight. Fractional weight means here an acquisition of a fraction of an integer; such as 1.3 acquisitions, or 0.28 acquisitions and so on. Fractional weights are not commonly considered when acquiring NUS spectra. For integer weighted acquisition and fixed $t_{\max}$, please refer to (Waudby and Christodoulou 2012). In the majority of the now used sampling schedule implementations, the sampling includes non-acquired data points, where typically the frequency of non-acquired data points increases for larger acquisition time ranges.

Even though the iSNR formulae are calculated in the time-domain, it should be possible to visualize the approximate forms in the frequency domain as well. Indeed, by using matched exponentially-weighted random sampling, we constructed the 512 data-point time domain out of an increasingly longer evolution time period. That is 512 points are sampled initially out of the first 512 points (uniform sampling), then sparsely out of 640 time points, etc, thus increasing the largest experimentally sampled time domain by steps of 128 points until 512 points have been sampled out of 4096 points. These sampling schedules are applied to an exponential decaying signal to a length of $2 \cdot T_2$ reached at 4096 points. To enable a meaningful statistical analysis 100 different sets of simulated Gaussian noise were added at each point prior to time-equivalent reduction. As seed values are required for practical implementation of the exponential random weighting, ten different seed values were used, creating a sampling set of 1000 different NUS like spectra for each evaluation point representing a particular $t_{\max}$. Reconstruction to convergence was done with hmsIST. An apodization by a cosine window was applied to reach zero at the last collected data point. This was, followed by zero filling to 4096 points and Fourier transformation was applied.

Both $SN_{\text{rms}}R$ (or $SN(\ell_2)R$), and $SN_{\text{max}}R$ (or $SN(\ell_\infty)R$) were measured (black and blue symbols in Figure 2). Also included in the figure is a scaled theoretical iSNR curve (dashed line) calculated with the formula from Palmer et al. (Palmer MR et al. 2015). The differences between the calculated iSNR and measured $SN_{\text{rms}}R$ curves can be explained by (a) the initial sampling in the measuring procedure explained above was done uniformly random, and hence with a uniform rather than an exponential profile, and (b) the proportion of obtained data points relative to the total number of reconstructed points is decreasing, creating greater gaps and making it harder to reconstruct the original spectrum. Even with these discrepancies, the similarity between the calculated curve and the measured $SN_{\text{rms}}R$ are striking.

Most importantly, however, there is a dramatic difference between $SN_{\text{max}}R$ and $SN_{\text{rms}}R$. The $SN_{\text{max}}R$ values are decreasing after a maximum at $t = \frac{1}{2} \cdot T_2$. As we have shown that $SN_{\text{max}}R$ values are more relevant to spectral analysis and peak picking than $SN_{\text{rms}}R$ (Hyberts et al. 2013), we will henceforth consider $SN_{\text{max}}R$ a more relevant measurement of sensitivity, the ability to distinguish real peaks from maximum noise.

Apart from directly measured SNR values, kurtosis and the similar measurement of $SN_{\text{rms}}R / SN_{\text{max}}R$ clearly indicate that there is an increasing deviation from Gaussian white

noise (Figure S1). Figures S2, S3 and S4 further explore this point by the following procedures: (a) We calculate the precision of measuring the signal heights ($SN_{top}R$). Here the value of $SN_{top}R$ is calculated as the average signal height over the standard deviation of the signal height. This measurement can only be done over an ensemble. Obviously, the precision of the peak height decreases with spreading the same number of time domain data to longer evolution times. (b) We explore if the profiles of the $SN_{rms}R$ and $SN_{max}R$ values changes over different time scales of $1 \cdot T_2$ or $2 \cdot T_2$; they do not (Figure S2). Please note that the obtained sparsity at $t_{max}^* = \frac{1}{2} \cdot T_2$ is different in the two panels of Figure S2: 512 of 1024 at the left panel (50%) and 512 of 2048 (25%) at the right panel. (c) We test whether apodization prior to Fourier transform drastically changes the outcome; it does not (Figure S3). (d) To examine whether the deviations between the three S/N characterizations are a result of reconstructing data with intermitted noise distribution we apply a straight DFT transformation of the NUS data and observe no deviation (Figure S4, left).

Figure S5 explores the option of an extremely heavy weighting towards the early sampling points by a weighting function of e^{-15t/T_2} . Even though not explicitly specified by (Palmer MR et al. 2015), improved sensitivity (iSNR) occurs at heavy early weighting (bias) and measurements where the t_{max} is as large as possible. At a first glance, this figure seem to contradict our finding of a sensitivity ($SN_{max}R$) maximum at a value for t_{max} at approximately $\frac{1}{2} \cdot T_2$. However, by closer analysis, no sampling schedule had the actual t_{max} beyond $0.65 \cdot T_2$. When using ten separate random sampling schedules with same exponential weighting, we can also measure the average latest t_{max} as $0.56 \cdot T_2$. This illustrates the difference in the time t'_{max} used for creating the schedule with a weighted probability function and actual t_{max} in the sampling schedule. Noteworthy is the persistent similarity between the calculated iSNR and the measured SNR ($SN_{rms}R$).

An experimental spectrum as test case for optimal sampling strategies

Figure 3a presents a HSQC spectrum of GB1 with high S/N (251:1) obtained on a Bruker 500 MHz spectrometer with a room temperature probe. 512 complex increments were recorded in the indirect nitrogen dimension with Echo-Anti-Echo acquisition (Cavanagh et al. 1991; Kay et al. 1992). The sweep width was set to 1824.651 Hz, which provides a dwell time of 0.55 msec. The t_{max} is therefore 280 msec, which is approximately $2 \cdot T_2$ as the majority of the nitrogen transverse relaxation times for this sample are between 140 and 150 msec. 75 peaks are present in the spectrum, including both backbone and side chain H-N pairs. The 75 peaks are picked in this uniformly sampled data set and used as standards of peak position. Several of the side-chain residues relax slower than the majority of the backbone H-N pairs. On the other hand, a few of the backbone residues relax faster as well. All in all, this presents a very practical test case for simulations to explore optimal lengths of experimentally covered evolution times.

In the right panel of Figure 3 artificial Gaussian noise was added to the time domain data of the GB1 HSQC experiment for two reasons: first, to adjust the spectrum to a S/N level reflecting more typical protein samples, and second, by virtue of adding differently generated noise, we can produce many spectra composed of the same signals, but with different noise to allow for statistical analysis. Two different levels of noise were simulated

with the resultant S/N in the spectra being 250:1 and 21:1 and analyzed in Figures 3a and 3b, respectively. The simulation with the higher SNR closely mimics the original data in noise level, but again this added noise provides multiple spectra to be generated for statistical analysis of frequency accuracy. The simulation with the lower SNR emulates more practical conditions, for example when studying higher molecular weight proteins or samples with lower concentration.

For both conditions where noise was added, 50 data sets with differently seeded noise were produced. Once these 50 time domain 2D HSQC spectra were produced with 512 data point in the indirect dimension, each of them was reduced to 64 data points in the indirect dimension. These 64 data points were selected using the following strategy variants (see Figure 4a)

1. Bottom row: The first 64 out of the 512 experimental data points were selected uniformly, and the subsequent 448 points were reconstructed, i.e., a total of 512 points are reconstructed, with subsequent processing to the frequency domain. The last obtained data point is the 64th increment, or $t_{\max} = 35.2 \text{ msec}$ ($\sim 0.25 \cdot T_2$).
2. Row two from bottom: Here, 64 points were selected from the $64+16=80$ first data points using PoissonGap sampling (sinusoidal weight 2) (Hyberts et al. 2010). $t_{\max}$ is therefore the 80th data point because this is the last possible acquired point. Hence, $t_{\max} = 44.0 \text{ msec}$. The 16 internal not-measured points together with the 432 subsequent data points are reconstructed and, together with the 64 obtained points, yield again a total of 512 points that are reconstructed. These are then processed to the frequency domain as above.
3. This procedure is repeated, adding another 16 data points to the group from which the 64 data points are selected. This is continued for every row above in Figure 4a, adding another 8.8 msec to $t_{\max}$. Thus, Figure 4a describes six example schedules for selecting the 64 experimental data points from 64, 128, 192, 256, 384 and 512 points, where the first schedule is uniformly sampled. These represent $t_{\max}$ values of 33.0, 70.4, 107.8, 140.8, 211.2 and 280 msec respectively. Please note that for each row in the **Figure**, the same number of experimental data points are selected but spread out differently and covering 512 grid points of equivalent time points to be reconstructed. The filled circles of the bottom row of Fig. 4a indicate $t_{\max}$ of the full 29 schedules analyzed in the subsequent figures.
4. The last data point in Figure 4a represents non-extrapolated NUS, i.e., the 64 data points selected are sampled using PoissonGap sampling from the entire 512 data points of the provided data. The $t_{\max}$ is therefore 280 msec ($\sim 2 \cdot T_2$). This is visualized in the top row of the figure.

Reconstruction of missing points below $t_{\max}$ and above was performed with hmsIST. Specifically, all of these data sets were reconstructed to 512 data points in the frequency domain. This means that all data sets have identical lengths of time domains $t^*_{\max} = 280 \text{ msec}$, (in contrast to $t_{\max}$ = the longest measured

evolution time), were apodized with a cosine window, zero filled once and Fourier transformed. Figure 4b shows a corresponding strategy for selecting 512 experimental points and spread over 4096 grid points.

Please note that the overall sparsity used for reconstruction is constant choosing 512 points of 4096 (or 12.5%). The acquired sparsity however varies from 100% when acquired uniformly, to 80% at the second point and finally to 12.5% with strict NUS for the last point.

The NMRDraw (Delaglio et al. 1995) peak picker 'pk.tcl' was used to pick the same 75 peaks in each spectrum. The final average of peak height and the line width as well as the standard deviation of the peak positions in the indirect ^{15}N dimension were averaged over the 50 different spectra and the 75 resonances seen with various relaxation parameters. Considering that each HSQC contains 75 pickable peaks, the total peak sampling base consists of 3750 (50×75) peaks.

In order to avoid mistakes in peak picking, a couple of cutoff criteria were introduced. First, the intrinsic cutoff level of the peak picker is adjusted to avoid picking noise. This means for every peak present in the linear set, we picked only one peak in our reconstructed data, if a peak could be found at all. Second, the peak may not be more than 5 Hz off position relatively to the standard peak position in the nitrogen dimension. This is a relatively loose constraint as it turns out that it is more than 5σ of the variation of the peak positions, and hence the constraint should not affect the statistics more than preventing loss from possible outliers. Third and finally, if a particular resonance reports a line width larger than 20 Hz, this peak is rejected. The final success rate was equal to or greater than 90% for each sampling method used. These criteria ensured stable statistical analysis.

Results from this analysis are summarized in Figure 5: addressing sensitivity, Figure 6: addressing line width and Figure 7: addressing accuracy of peak position. In Figures 5a, 6a and 7a, black circles mark the results from the original data (no noise added), while results for data with low level noise added are marked with red diamonds (S/N 250:1) and higher level noise with blue squares (S/N 21:1). Optimal sensitivity is obtained at a $t_{\max}$ of about $\frac{1}{2}T_2$, while extending the time domain data with hmsIST to $2T_2$. Spending instrument time on measuring experimental points beyond $\frac{1}{2}T_2$ leads to lower SNR (Figure 5a), does not further decrease the line width (Figure 6a) and deteriorates accuracy of peak positions (Figure 7a).

Figure S6 provides a visual verification of the effects of these sampling strategies with a representative peak from the HSQC spectrum of GB1. The four panels illustrate the results of using a $t_{\max}$ of 35 msec ($0.25T_2$, panel A), 70 msec ($0.5T_2$, panel B), 140 msec ($1T_2$, panel C) and 280 msec ($2T_2$, panel D). The line broadening and intensity variations are apparent. The position variation on the other hand is hard to see in this particular visualization. The visually-best reconstruction is obtained for a $t_{\max}$ of 70 msec, which corresponds to an estimated $\frac{1}{2}T_2$, which is consistent with the maximum in Figure 5a.

To supplement the HSQC data above, we performed simulations that allowed us to extend data to $4T_2$ (Figure 4b). The simulations emulate a signal with a T_2 of 150 msec, on the Nyquist grid and where different noise has been added 1000 times (Figure 4b). The peaks

were picked in the same manner as with the GB1 HSQC data, but unlike the experimental situation the success rate for each simulation point was 100%. Two simulation sets were produced that differ in the amount of white noise added to the time domain; the amount of noise added for the simulation indicated with the blue squares is twice the amount added in the simulation indicated with white triangles. (Figures 5b, 6b & 7b.)

The observed sensitivity (Figure 5), is calculated based on $\text{SN}(\ell_{\infty})R$ ($\text{SN}_{\max}R$). That is, sensitivity as a value is considered to be the signal height over the maximum absolute value of the noise, and averaged over the 50 simulations and all 70 HSQC peaks analyzed in Figure 5a, or the 1000 simulations on a single Lorentzian line of Figure 5b. In both the experimental and in the simulated case, the observed sensitivity varies dependent on the distribution of sampling. For data with low intrinsic sensitivity, (blue squares in both cases), the apparently best sampling technique is not to interpolate but to only extrapolate. On the other hand, in the case of greater intrinsic sensitivity (i.e., less noise, black disks, red diamonds and white triangles), maximum sensitivity is achieved around $t_{\max} = \frac{1}{2}T_2$. This variation is based on the fact that reconstruction with hmsIST is a fundamentally non-linear process; even though it mimics rather well a linear process, especially when the signals have larger intrinsic sensitivity.

For the case of line width measurements (Figures 6a and 6b) a different picture is presented. The measured line widths in the experimental situation are larger than what is found in the simulation of the single Lorentzian line, and vary throughout the curve. A minimum is present for the experimental data at a time $t_{\max}$ slightly greater than $\frac{1}{2}T_2$. On the other hand, the two cases with simulated lines (Figure 7b) exhibit essentially a flat response, and show narrower line widths than the ones found in the experimental curve. This could be due to the following issues: (i) The 1D simulations are performed with an extension from $2T_2$ to $4T_2$. This could yield slightly narrower lines. Also, (ii) the simulations are using a 4k FID whereas the HSQC was acquired with 512 complex data points in the indirect dimension. (iii) The simulations are carried out with the signals directly on the Nyquist grid, where this is not the case in the experimental HSQC spectra. (iv) The HSQC spectrum contains more than one signal per indirect FID, which may result in the variations. The first three points are investigated in Figure S7. Here we restrict the simulations to 512 complex data points in the FID (matching the HSQC data and addressing point (ii)); without the extension (point (i)) and simulations on and off the Nyquist grid, respectively (point (iii)), and with different line shapes (Lorentzian and Gaussian), all yield equally flat curves. As there is no doubt that the HSQC spectrum provides varied line-shapes (see Figure S6), we are left with hypothesis (iv) which states that the increase of line-widths for $t_{\max}$ beyond $\frac{1}{2}T_2$ in the experimental spectra may be due to the fact that more than one signal is present per indirect FID.

Finally, frequency accuracy (Figure 7) shows a picture resembling that of both observed sensitivity and line widths. The experimental curve with more noise added (blue squares in Figure 7a) practically mimics that presented in the line width case. Thus, we obtain a curve with a minimum at $t_{\max}$ slightly greater than $\frac{1}{2}T_2$. On the other hand, at higher intrinsic sensitivity, the frequency variations are reduced with increasing $t_{\max}^*$. The simulated curves show that the intrinsic sensitivity is clearly related to frequency accuracy. About twice the

variation of the frequency is observed at twice the amount of added noise. Larger variations are also found at $t_{\max}$ values shorter than $\frac{1}{2}T_2$.

Figures 5, 6 and 7 suggest that it is prudent to obtain indirect dimensions with an acquisition of $t_{\max}$ equal to approximately $\frac{1}{2}T_2$ or slightly more. The observed sensitivity may be higher at slightly smaller $t_{\max}$; however the best resolution and frequency accuracy are not reached until a $t_{\max}$ of $\frac{1}{2}T_2$. Additionally, it appears desirable to further extrapolate the data up to four fold beyond $\frac{1}{2}T_2$ with hmsIST or another reconstruction algorithm to increase the digital resolution.

Application to a 3D HNCA spectrum

With the analysis described above we return to the HNCA observations displayed in Figure 1, which indeed initiated this current study. Here, three time equivalent HNCA spectra were obtained. Each of these was obtained with the same pulse program using semi constant time evolution in the nitrogen dimension. The spectra were obtained with GB1 at 750 MHz, using 8 scans, and sweep widths of 10504, 2431 and 6032 Hz in the ^1H , ^{15}N and ^{13}C dimensions, respectively. The total time of each acquisition was about 10 hours for each spectrum. 900 acquisition points were taken in the $^{15}\text{N}/^{13}\text{C}$ plane in 3 fashions, namely; (i) uniformly 30×30 ($\rightarrow t_{\max}[^{15}\text{N}] = 12 \text{ msec} \cong T_2/12$), (ii) 900 points in a non-uniform 183×30 grid, 16%, sampled with PoissonGap sampling ($\rightarrow t_{\max}[^{15}\text{N}] = 75 \text{ msec} \cong \frac{1}{2}T_2$) and (iii) 900 points in a non-uniform 365×30 grid, 8%, sampled with PoissonGap sampling ($\rightarrow t_{\max}[^{15}\text{N}] = 150 \text{ msec} \cong T_2$). Acquisition in the carbon dimension was the same for all spectra.

As the spectra were obtained with high intrinsic sensitivity, Gaussian distributed white noise was added in order to simulate typical protein sample conditions. Three differently generated sets of noise with identical RMS were created, one each for every spectrum. The generated noise sets were added to the time domain of each spectrum. Where applicable, interpolation was performed with hmsIST, and extrapolation, was achieved with Zero filling (ZF), Linear prediction (LP) and by hmsIST. When both interpolation and extrapolation with hmsIST were applied, this was performed at the same time during a single reconstruction of the data.

Figure 8 presents a visual overview of the interpolation and extrapolation data in our HNCA example. Table 1 provides a more thorough evaluation with the averages for each spectrum and averaging values over 58 pairs of intra and inter-residual HNCA signals. Average line widths for both major and minor signals in the pair, and are denoted $lw(\text{maj})$ and $lw(\text{min})$ respectively. Zero filling (ZF) the uniformly obtained data obviously does not prevent the widening of signals; a broadening that is due to truncated sampling. Only by using reconstruction of time equivalent non-uniform sampling of the signals it is possible to obtain these signals with little or no line broadening. For acquisition time equivalency, the narrowest line widths are provided when using hmsIST extrapolation of the NUS spectrum with 150 msec $t_{\max}$ in the nitrogen dimension ($\sim 1T_2$) to a $t_{\max}^*$ of 600 msec. The line profiles are not perfectly Lorentzian, but have large shoulders at the base. In either case, the measured line widths are only a few Hz wider than the natural line width of 2.1 to 2.3 Hz

($\nu_{1/2} = 1/\pi T_2$, $T_2 = 140\text{--}150$ msec). Should the spectrum be very crowded, extrapolation of the data to $4 \cdot T_2$ may be beneficial.

The best observed sensitivity $SN(\ell_{\infty})R$ ($SN_{\max}R$), (columns $sen(maj)$ and $sen(min)$), respectively) is found for the hmsIST reconstruction to $t_{\max}^* = 150$ msec, whether this is extrapolated from the uniformly sampled spectrum, or the interpolated NUS spectrum where $t_{\max} = 150$ msec. Extrapolation to $t_{\max}^* = 600$ msec yields slightly lower values, although the hmsIST reconstruction $t_{\max} = 75$ msec $\rightarrow$ $t_{\max}^* = 600$ msec yield sensitivity values on par with most reconstructions to $t_{\max}^* = 150$ msec. The single worst observed sensitivity is found when using linear prediction (LP) from the uniformly sampled spectrum.

The average frequency deviation of each method was measured as the average absolute deviation of the peak centers of the two HNCA signals within the pair, in their nitrogen frequency. It should be noted that individual deviations might easily be three times the average. All the various applications of interpolating NUS spectra with hmsIST are found with the least deviation, together with applying zero-filling (ZF) after hmsIST reconstruction of the $t_{\max} = 75$ msec spectrum.

Figure 9 demonstrates the effect of noise on frequency deviation. The leftmost panel shows the same system as selected in Figure 8 ($t_{\max} = 75$ msec, $t_{\max}^* = 150$ msec), this time without any simulated noise added. The centers of these two peaks are essentially identical (0.03 Hz; Table 2). However, with increasing amount of noise, the deviation increases. The fundamental reason for this is seen in the underlying noise, which represents noise added without spectral data (rightmost panel, Figure 9). The more noise, the more deviation, although it is not possible a-priori to know the amount or even the direction of the shift of peak maxima. It is however easy to see that in spectra with very low intrinsic sensitivity, the frequency deviations can be similar in value to the line widths seen when reconstructing a NUS spectrum to $2 \cdot T_2$.

A slightly different evaluation is used on a separate pair of semi-constant-time-obtained HNCA spectra on GB1 shown in Figures 1 and S8. The first of these HNCA spectra (Figure 1a) emulate a uniformly sampled spectrum extended with hmsIST. 8 scans were obtained for 900 hyper-complex increments in the $^{15}\text{N}/^{13}\text{C}$ plane. 30 increments were collected in both indirect dimensions. The sweep width was set to 2190 Hz in the nitrogen dimension, yielding 73Hz per data point. The $t_{\max}$ is hence 14 ms; reflecting one-tenth of the T_2 around 140 ms.

The second of these HNCA spectra (Figure 1b) is obtained in a true non-uniform fashion, still using 8 scans and 900 hyper-complex increments in the $^{15}\text{N}/^{13}\text{C}$ plane. The data points were distributed by two-dimensional Poisson Gap sampling (sinusoidal weight 2), at a maximum of 30 increments in the carbon dimension but 150 increments in the nitrogen dimension. The $t_{\max}$ is hence fivefold; i.e. 72 ms or approximately half of the T_2 for the systems in this spectrum. The peak separation in the ^{15}N dimension is significantly reduced and the peaks are nearly perfectly aligned.

Picking all peaks in the both HNCA spectra using `pk.tcl`, the frequency differences between paired inter- and intra-cross peaks are shown in Figure S8. The frequency variations are

purely random, but the expectation values of the absolute variation for the two spectra are different. As a rule, the accuracy of peak positions in the non-uniformly sampled spectrum is 4 to 5 times better than in the spectrum that was extended after being uniformly sampled.

Discussion

Extending time domain data in the indirect dimensions is a common practice in multi-dimensional NMR experiments; however, it is not clear what is the best procedure for extrapolation and how far one should go with respect to the transverse relaxation time within a given overall measurement time. Here we compare simple zero filling, linear prediction and NUS for extending time domain data. For reconstruction of NUS data we used hmsIST. Other reconstruction methods are expected to yield similar results, however this will need further investigations, which is beyond what can be reported in this paper.

We suggest that the most meaningful measure of sensitivity in NUS spectra is to use the ratio of signal height over maximum noise ($SR_{\max}N$) because it best represents the ability to distinguish signals from noise. To measure the accuracy of peak positions we compare intra- and inter-residue HNCA peaks, which we know must align.

The search for optimal $t_{\max}$ using data from a HSQC spectrum and extensive simulation show that (a) for greatest observed sensitivity, $t_{\max}$ should be set to approximately $\frac{1}{2}T_2$, (b) line widths are insensitive to $t_{\max}$, but are strongly dependent on $t_{\max}^*$; longer $t_{\max}^*$ yields narrower lines, and (c) for greatest observed frequency accuracy (least frequency deviation), $t_{\max}$ should be at least $\frac{1}{2}T_2$, but accuracy does not increase dramatically with larger $t_{\max}$ values. A best overall value of $t_{\max}$ is hence around $\frac{1}{2}T_2$. Furthermore, the observed intensity and frequency accuracy deteriorate with more noise whereas the line width measurements are independent of the noise as long as the values can be reasonably measured.

Of the extrapolation methods, zero filling (ZF) a uniformly sampled spectrum, yields the broadest of all line-widths. This is expected since the apodization window does not extend to $t_{\max}^*$ but to $t_{\max}$ only. On the other hand, ZF yields a fair amount of observed sensitivity and frequency accuracy. As this is strictly not a reconstruction (no information has been added), and the profile of the baseline noise is Gaussian distributed ($kurt-3 = 0.0$).

On the other hand, of the methods extending the uniformly sampled spectrum, linear prediction (LP) provides the narrowest lines, but the worst observed sensitivity and worst frequency accuracy. It should also be noted that the reconstructed baseline noise tends to be “spiky” as the kurtosis value is large.

When applying hmsIST entirely to extrapolate a uniformly sampled spectrum, yields narrow line widths, the best observed sensitivity, and good frequency accuracy. To be noted are the identical values of kurtoses, and even if they are greater than zero, they adopt relatively small values. This means that even though the noise is not Gaussian, its character is not as “spiky” as that of linear prediction.

When sampling the spectrum non-uniformly to $\frac{1}{2}T_2$ and tested with ZF and LP, the data are first reconstructed by hmsIST. Thereafter the data are extrapolated to T_2 . Interestingly, the values of line widths, observed sensitivity, frequency accuracy, and even the kurtosis value are now quite similar. Hence we can draw the conclusion that although it is technically possible to linear prediction time domain data after reconstruction with hmsIST, it does not yield a better result than just zero filling.

When doing both interpolation and extrapolation with hmsIST, the reconstruction to $t_{\max}^*$ is done in one step rather than two. The results are overall better than using the two-step procedure, both with respect to sensitivity and frequency accuracy.

If the spectrum has very crowded areas, it may help to further extrapolate the signal. When the signals are extrapolated to 4 times T_2 , the resulting line widths approach their natural line widths. There is however a slight loss in observed sensitivity, and the spectra become more “spiky”. Even though the reported frequency accuracy is independent on whether the signals are extrapolated to one or four times T_2 , individual values of the frequency difference may be on the order of the observed line widths. This may cause confusion when doing assignment e.g. in a HNCA spectrum. Additionally, the lines seem to have a wide feature in the lower part of the peak.

When using values of $t_{\max} = T_2$ rather than $t_{\max} = \frac{1}{2}T_2$, the result yields slightly lower sensitivity values and slightly better values for line width and for frequency accuracy. In the end, it depends on the particular situation whether to interpolate or to do a combination of interpolation and extrapolation.

In summary, we have shown that hmsIST may be used to extrapolate as well as interpolate with high fidelity. We have also found that, optimally, data should be obtained to at least $\frac{1}{2}T_2$, for sensitivity, resolution and frequency considerations, with diminishing returns for points exceeding $\frac{1}{2}T_2$. Extrapolation to approximately $2T_2$ with hmsIST yields sharper lines. Furthermore, extension and subsequent application of window functions does not significantly scale down measured time points, which most likely enhances the accuracy of the resulting spectra. Given NUS’s ability to sample to large evolution times like $\frac{1}{2}T_2$ in reasonable acquisition times, we strongly suggest that this be done in practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Gregory Heffron for the assistance with NMR instruments and Dr. David Morgan for help with the computers used for this study.

Funding: This study was supported by the National Institute of Health (Grant GM047467 and EB002026).

References

- Balsgart NM, Vosegaard T. Fast Forward Maximum entropy reconstruction of sparsely sampled data. *Journal of magnetic resonance*. 2012; 223:164–169. DOI: 10.1016/j.jmr.2012.07.002 [PubMed: 22975245]
- Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of magnetic resonance*. 1987; 73:69–77.
- Candes JC, Walkin MB. An Intorduction to Compressive Sensing. *IEEE Signal Processing Magazine*. 2008 Mar.:21–30.
- Cavanagh J III AGP, Wright PE, Rance M. Sensitivity Improvement in Proton-Detected Two-Dimensional Heteronuclear Relay Spectroscopy. *Journal of magnetic resonance*. 1991:91.
- Celik H, Shaka AJ. Filter diagonalization using a "sensitivity-enhanced basis": improved performance for noisy NMR spectra. *Journal of magnetic resonance*. 2010; 207:17–23. DOI: 10.1016/j.jmr.2010.07.019 [PubMed: 20801065]
- Coggins BE, Zhou P. Polar Fourier transforms of radially sampled NMR data. *Journal of magnetic resonance*. 2006; 182:84–95. DOI: 10.1016/j.jmr.2006.06.016 [PubMed: 16820311]
- Coggins BE, Zhou P. High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. *Journal of biomolecular NMR*. 2008; 42:225–239. DOI: 10.1007/s10858-008-9275-x [PubMed: 18853260]
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995; 6:277–293. [PubMed: 8520220]
- Delsuc MA, Mallavin TE. Maximum Entropy Processing of DOSY NMR Spectra. *Anal Chem*. 1998; 70:2146–2146.
- Donoho DL. For most large underdetermined systems of linear equations the minimal [MATHEMATICAL SCRIPT SMALL L]1-norm solution is also the sparsest solution. *Comm Pure Appl Math*. 2006; 59:797–829. DOI: 10.1002/cpa.20132
- Donoho DL, Johnstone IM, Hoch JC, Stern AS. Maximum Entropy and the Nearly Black Object. *J R Statist Soc B*. 1992; 54:41–81.
- Drori I. Fast Minimization by Iterative Thresholding for Multidimensional NMR Spectroscopy EURASIP. *Journal on Advances in Signal Processing*. 2007; 2007:020248.doi: 10.1155/2007/20248
- Hoch, JC., Stern, AS. *NMR Data Processing* Wiley-Liss. New York: 1996.
- Hogbom JA. Aperature Synthesis with a Non+Regular Distribution of Interferometer. *Baseline Astron Astrophysi Suppl*. 1974; 15:417–426.
- Holland DJ, Bostock MJ, Gladden LF, Nietlispach D. Fast multidimensional NMR spectroscopy using compressed sensing. *Angewandte Chemie*. 2011; 50:6548–6551. DOI: 10.1002/anie.201100440 [PubMed: 21648028]
- Hyberts SG, et al. Ultrahigh-Resolution 1H-13C HSQC Spectra of Metabolite Mixtures Using Nonlinear Sampling and Forward Maximum Entropy Reconstruction. *J Am Chem Soc*. 2007; 129:5108–5116. [PubMed: 17388596]
- Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *Journal of biomolecular NMR*. 2012; 52:315–327. DOI: 10.1007/s10858-012-9611-z [PubMed: 22331404]
- Hyberts SG, Robson SA, Wagner G. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *Journal of biomolecular NMR*. 2013; 55:167–178. DOI: 10.1007/s10858-012-9698-2 [PubMed: 23274692]
- Hyberts SG, Takeuchi K, Wagner G. Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data. *J Am Chem Soc*. 2010; 132:2145–2147. [PubMed: 20121194]
- Kay LE, Keifer P, Saarinen T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J Am Chem Soc*. 1992; 114:10663–10665.

- Kazimierczuk K, Orekhov VY. Accelerated NMR spectroscopy by using compressed sensing. *Angewandte Chemie*. 2011; 50:5556–5559. DOI: 10.1002/anie.201100370 [PubMed: 21538743]
- Kazimierczuk K, Zawadzka-Kazimierczuk A, Kozminski W. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. *Journal of magnetic resonance*. 2010; 205:286–292. DOI: 10.1016/j.jmr.2010.05.012 [PubMed: 20547466]
- Kim S, Szyperski T. GFT NMR, a New Approach To Rapidly Obtain Precise High-Dimensional NMR Spectral Information. *J Am Chem Soc*. 2003; 125:1385–1393. [PubMed: 12553842]
- Kupce E, Freeman R. Fast multidimensional NMR: radial sampling of evolution space. *Journal of magnetic resonance*. 2005; 173:317–321. DOI: 10.1016/j.jmr.2004.12.004 [PubMed: 15780924]
- Kupce E, Freeman R. Fast multi-dimensional NMR by minimal sampling. *Journal of magnetic resonance*. 2008; 191:164–168. DOI: 10.1016/j.jmr.2007.12.013 [PubMed: 18191598]
- Laue ED, Skilling J, Staunton J. Maximum Entropy Reconstruction of Spectra Containing Antiphase Peaks. *Journal of magnetic resonance*. 1985; 63:418–424.
- Lee S, Wen H, An YJ, Cha JW, Ko YJ, Hyberts SG, Park S. Carbon Isotopomer Analysis with Non-Uniform Sampling HSQC NMR for Cell Extract and Live Cell Metabolomics Studies. *Anal Chem*. 2017; 89:1078–1085. DOI: 10.1021/acs.analchem.6b02107 [PubMed: 28029042]
- Mandelstam VA, Taylor HS, Shaka AJ. Application of the Filter Diagonalization Method to One- and Two-Dimensional NMR Spectra. *Journal of magnetic resonance*. 1998; 133:304–312. [PubMed: 9716473]
- Matsuki Y, Eddy MT, Herzfeld J. Spectroscopy by Integration of Frequency and Time Domain Information for Fast Acquisition of High-Resolution Dark Spectra. *J Am Chem Soc*. 2009; 131:4648–4658. [PubMed: 19284727]
- Orekhov VY, Ibraghimov IV, Billeter M. MUNIN: A new approach to multi-dimensional NMR spectra interpretation. *Journal of biomolecular NMR*. 2001; 20:49–60. [PubMed: 11430755]
- Orekhov VY, Jaravine VA. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Progress in nuclear magnetic resonance spectroscopy*. 2011; 59:271–292. DOI: 10.1016/j.pnmrs.2011.02.002 [PubMed: 21920222]
- Palmer MR, Suiter CL, Henry GE, Rovnyak J, Hoch JC, Polenova T, Rovnyak D. Sensitivity of nonuniform sampling NMR. *The journal of physical chemistry B*. 2015; 119:6502–6515. DOI: 10.1021/jp5126415 [PubMed: 25901905]
- Paramasivam S, et al. Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *The journal of physical chemistry B*. 2012; 116:7416–7427. DOI: 10.1021/jp3032786 [PubMed: 22667827]
- Rovnyak D, Hoch JC, Stern AS, Wagner G. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *Journal of biomolecular NMR*. 2004; 30:1–10. [PubMed: 15452430]
- Rovnyak D, Sarcone M, Jiang Z. Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. *Magnetic resonance in chemistry : MRC*. 2011; doi: 10.1002/mrc.2775
- Shannon CE. A Mathematical Theory of Communication. *The Bell System Technical Journal*. 1948; 27:379–423. 623–656.
- Stern AS, Donoho DL, Hoch JC. NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction. *Journal of magnetic resonance*. 2007; 188:295–300. DOI: 10.1016/j.jmr.2007.07.008 [PubMed: 17723313]
- Sun S, Gill M, Li Y, Huang M, Byrd RA. Efficient and generalized processing of multidimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. *Journal of biomolecular NMR*. 2015; 62:105–117. DOI: 10.1007/s10858-015-9923-x [PubMed: 25808220]
- Waudby CA, Christodoulou J. An analysis of NMR sensitivity enhancements obtained using non-uniform weighted sampling, and the application to protein NMR. *Journal of magnetic resonance*. 2012; 219:46–52. DOI: 10.1016/j.jmr.2012.04.013 [PubMed: 22609525]
- Worley B. Convex accelerated maximum entropy reconstruction. *Journal of magnetic resonance*. 2016; 265:90–98. DOI: 10.1016/j.jmr.2016.02.003 [PubMed: 26894476]

Figure 1.

Obtained peak accuracy by comparing intra and inter-residue peaks in HNCA spectra. **(a)** The leftmost panel: A slice presents a typical situation for uniformly sample spectrum where the number of complex points in the nitrogen dimension is 30. The corresponding 14 msec of evolution yield a digital resolution of 36 Hz in the nitrogen dimension after the data have been zero filled once. Middle panel: The time domain has been extrapolated to a total of 150 complex data points in the nitrogen dimension using hmsIST. This represents a virtual maximum acquisition time of 70 ms, which is approximately half that of the T_2 value. However, when picking the peaks they do not quite line up in the nitrogen dimension although belonging to the same nitrogen. Rightmost panel: To illustrate the misalignment, the line width has been additionally narrowed by 15Hz by exponential multiplication in the time domain. The value of separation is the same as that in the middle panel, yet here more clearly visible. **(b)** Lefthand panel: Shows the correspondence of the middle panel of (a) but obtained by hmsIST interpolation of a 20% time equivalent semi-constant time HNCA.. The 20% were sampled in then carbon nitrogen plane, though the number of collected complex data points in the carbon dimension is identical to the spectra shown in a. Righthand panel: The appearances of the peaks have been enhanced with a 15Hz line narrowing in the nitrogen dimension. Still, there is some misalignment between inter- and intra-residual peaks in the HNCA, but the misalignment is much smaller than in the linear case.

Figure 2.

Two arts of measuring SNR of non-uniformly sampled spectra, $SN_{\text{rms}}R$ (black diamonds) and $SN_{\text{max}}R$ (blue circles), measured on NUS reconstructed, simulated exponentially decaying signal. The sampling strategy is of an exponentially weighted random profile, emulating the matched exponentially sampled situation in (Palmer MR et al. 2015). The dashed black curve is theoretically calculated iSNR values and is arbitrary calibrated to that of the $SN_{\text{rms}}R$ data. For further information, see text.

Figure 3. Experimental 2D HSQC spectrum for testing extrapolation procedures. **Left:** HSQC spectrum of GB1 with high S/N (251:1) obtained on a Bruker 500 MHz spectrometer with room temperature probe. **Right:** Gaussian noise was subsequently added to the same spectrum for the purpose of modeling a more realistic situation with larger or less concentrated systems and to enable statistical analysis. The noise was added to the time domain of the GB1 HSQC experiment and resulted in a S/N of 21:1.

Figure 4.

Graphical explanation of the process used for sampling point selection in figures 5, 6 and 7. (a) The experimental HSQC spectrum was recorded with 512 complex data points. 29 separate reductions to 64 complex data points were done, each representing one of the black dots at the bottom. The first reduction was to select the 64 first complex data points, the second reduction 64 data points of the first 80 data points, and so on, selecting 64 data points from a number of data points incremented by 16 data points for each black dot. The process of reductions is repeated until 64 complex data points are selected from the full set of 512 complex data points. The hmsIST reconstructions were independently of type of reduction to

512 complex data points. This means in each instance the sparsity was 12.5% ($1/8^{\text{th}}$). An additional horizontal axis is inserted to visualize parts of T_2 when T_2 is set to 140 msec. **(b)** The simulation of a single FID with 4096 complex data points relaxed to $2 * T_2$ at the end of the FID, and with added simulated noise, was reduced in a similar function as the experimental HSQC spectrum. Here the first black dot symbolizes a selection of the 512 initial data points from the FID. An increment of 64 data points is done for each black dot until the 512 data points are selected from the 4096 initial complex data points. The hmsIST reconstruction at these data sets were done to 8192 complex data points, extending the data to $4 * T_2$. For further explanation see text.

Figure 5.

Graphs of observed sensitivity measured from the HSQC spectrum of GB1 (a) and from simulation (b). (a) The measures are derived from the average peak height over the maximum value pixel in the first row in the spectrum, where only noise is presented. The value hence represents $SN(\ell_\infty)R$. Please see the text regarding explanation of the data points. Note that the y-values of the curve with the blue boxes are to be read from the right axis. The figure indicates that data with high intrinsic sensitivity has a maximum around $t_{\max} = 60$ msec, which is somewhat less than $\frac{1}{2}T_2$ if the T_2 is in the range of 140 to 150 msec. Figure 5b: As in figure a) the values represents $SN(\ell_\infty)R$. The synthetic data contains one single

Lorentzian signal centered on a Nyquist frequency, and noise added. The value for the noise is derived from a portion of the spectrum away from the signal. The values of peak height and noise of the curve with blue boxes was tuned to mimic that found from the HSQC spectrum. By reducing the noise by half the curve with open triangles, the shape alters to that found for situations with higher intrinsic sensitivity for the HSQC spectrum, and a maximum appears around $\frac{1}{2}T_2$.

Figure 6.

Graphs of observed line widths measured from the HSQC spectrum of GB1 **(a)** and from simulation **(b)**. **(a)**: The first observation yields that the intrinsic sensitivity does not influence the ability of measuring the line width. The narrowest lines (and hence the best resolution) seems to be achieved at $t_{\max} = 100$ msec, $2/3 \cdot T_2$. **(b)** Contrary to the graphs extracted from experimental data, the simulated results give a mostly flat response. See text and figure S7 for discussion.

Figure 7.

Graphs of peak position variation from the HSQC spectrum of GB1 **(a)** and from simulation **(b)**. **(a)** The resulting curves show a somewhat different result; the maximum frequency accuracy (i.e. the minimal peak position variation) where the intrinsic sensitivity is less, show a minimum around $t_{\max} = 100$ msec, $2/3 \cdot T_2$, i.e similar to that of minimal line width measurement. However when the intrinsic sensitivity is higher, this minimum in the peak position variation is not seen, and suggests longer $t_{\max}$ for maximum frequency accuracy. **(b)**: The graph derived from simulation shows that short $t_{\max} < 3/4 \cdot T_2$ is not advisable. Otherwise the simulation seems to agree in general with the experimental situation.

Figure 8.

The figure compares different methods of extrapolating the nitrogen dimension of 3 HNCA spectra obtained time-equivalent but with different $t_{\max}$. The spectrum obtained with $t_{\max} = 12\text{ msec}$ is obtained uniformly; whereas the two others are obtained non-uniformly. The sampling was tuned to $t_{\max} = 75\text{ msec}$ and $t_{\max} = 150\text{ msec}$ representing $\frac{1}{2}T_2$ and $1T_2$ respectively. The uniformly sampled spectrum was extended in the time domain to $t^*_{\max} = 150\text{ msec}$ using Zero filling, Linear prediction and hmsIST. The spectrum sampled to $t_{\max} = 150\text{ msec}$ is simply reconstructed with hmsIST, whereas the spectrum sampled to $t_{\max} = 75\text{ msec}$ is first reconstructed to $t^*_{\max} = 75\text{ msec}$ and then extended by Zero filling and Linear

prediction in the processing portion. When hmsIST is used both for interpolating reconstruction and extrapolating extension, this is done in one step however. In addition both NUS acquired spectra are extended to $t_{\max}^* = 600$ msec or approximately $4 \cdot T_2$. The results are presented in Table 1 and discussed in the text.

Figure 9.

Illustration of the relation between intrinsic sensitivity and frequency accuracy. The strength of the noise shifts the measurable center of the signal. The sign and amplitude of the shift is hence random. Numerical results are found in Table 2.

Comparison of line widths, sensitivity and accuracy of peak positions in a 3D HNCA of GB1 obtained with three different NUS schemes and extrapolated with zero filling, Linear prediction or hmslST

Results from the HNCA study. The three first columns describe the acquisition length of the nitrogen dimension, the final length of the nitrogen dimension and the method of how the data was extended. (The reconstruction of the NUS data was always done by hmslST.) The two following columns report the measured line width in Herz for the major (inter-residual) signal and the minor (intra-residual) signal. Following these two columns are measurements of the sensitivity base in SN(ℓ_{∞})R for the major and minor signal respectively. The resulting frequency difference between the major and minor signals is reported in Hertz. The three best results for each of the above-mentioned columns are entered in bold and italics. The success rate of the measurements is presented in the second to last column. The last column indicates the nature of the resulting noise in excess kurtosis: (*kurt* – 3)

$[X] = \frac{\mu_4}{\sigma^4} - 3, 0 = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2} - 3, 0$. The practice of reducing the kurtosis value by 3 forces the equation to adopt a value of 0 (zero) for Gaussian distribution. Positive values indicate a more “spiky” distribution than a Gaussian distribution. Increasing values indicates an even “spikier” distribution.

t _{max}	t* _{max}	method	lw(maj)	lw(min)	sen(maj)	sen(min)	freq diff	kurt-3
12ms	150ms	ZF	85.2	73.3	7.7:1	3.5:1	3.00	0.0
12ms	150ms	LP	14.9	23.9	5.1:1	1.6:1	4.42	27.4
12ms	150ms	hmslST	17.2	21.2	9.8:1	3.4:1	2.78	3.4
75ms	150ms	ZF	15.8	18.7	7.6:1	2.3:1	0.80	8.0
75ms	150ms	LP	13.1	16.6	7.3:1	2.2:1	1.65	10.0
75ms	150ms	hmslST	8.7	9.3	9.7:1	3.3:1	0.79	6.2
75ms	600ms	hmslST	4.1	5.1	8.7:1	3.0:1	0.85	24.6
150ms	150ms	hmslST	9.0	9.4	8.7:1	2.9:1	0.71	9.2
150ms	600ms	hmslST	3.5	3.3	6.2:1	2.1:1	0.70	32.7

Table 2

Table shows the results added noise. The three first columns describe the acquisition length over the final length of the nitrogen dimension, units of noise added and the method of how the data was extended. The two following columns report the measured line width in Hertz for the major (inter-residual) signal and the minor (intra-residual) signal. Following these two columns are measurements of the sensitivity base in $SN(\ell_\infty)R$ for the major and minor signal respectively. The resulting frequency difference between the major and minor signals is reported in Hertz. The last column shows the excess kurtosis.

$t_{\max}/t_{\max}^*$	noise unit add	method	lw(maj)	lw(min)	sen(maj)	sen(min)	freq diff	kurt-3
75ms/150ms	0	hmsIST	11.4	10.9	113.6:1	41.2:1	0.03	16.5
75ms/150ms	1	hmsIST	11.4	10.8	44.3:1	15.1:1	0.71	2.5
75ms/150ms	2	hmsIST	11.5	10.5	22.0:1	7.3:1	1.26	3.9
75ms/150ms	5	hmsIST	11.6	11.3	8.8:1	2.5:1	2.52	4.3
75ms/150ms	10	hmsIST	10.0	13.1	4.5:1	1.0:1	4.69	4.3
75ms/150ms	20	hmsIST	12.7	12.4	2.4:1	0.5:1	6.05	4.2