Journal of Biomolecular NMR (2005) 32: 141–150
DOI 10.1007/s10858-005-5977-5

Article

Fast acquisition of NMR spectra using Fourier transform of non-equispaced data

Dominique Marion*
Institut de Biologie Structurale ''Jean-Pierre Ebel'', CNRS-CEA-UJF, 41, Rue Jules Horowitz, 38027, Grenoble Cedex 1, France

Received 25 January 2005; Accepted 29 March 2005

Key words: digital resolution, Fourier transformation, non-linear sampling, Lagrange interpolation, signal/noise ratio

Abstract

Rapid acquisition of high-resolution 2D and 3D NMR spectra is essential for studying biological macromolecules. In order to minimize the experimental time, a non-linear sampling scheme is proposed for the indirect dimensions of multidimensional experiments. These data can be processed using the algorithm proposed by Dutt and Rokhlin (*Appl. Comp. Harm. Anal.* 1995, **2**, 85–100) for fast Fourier transforms of non equispaced data. Examples of $^{1}H-^{15}N$ HSQC spectra are shown, where crowded correlation peaks can be resolved using non-linear acquisition. Simulated data have been used to analyze the artefacts produced by the Lagrange interpolation. As compared to non-linear processing methods, this algorithm is simple and highly robust since no parameters need to be adjusted by the user.

Abbreviations: LP – linear prediction; MEM – maximum entropy method.

Introduction

Multidimensional NMR is nowadays recognized as a powerful technique for structural studies on biological macromolecules. Resonance assignment remains a prerequisite which involves the collection of a dozen of 3D or 4D triple-resonance experiments (Tugarinov et al., 2002). In order to resolve ambiguities due to accidental spectral overlaps in the case of large proteins, two key objectives can be identified: the digital resolution along each dimension should be maximized as well as the dimensionality of the spectra. As a result of their joint realization, the protracted duration of data-gathering process grows inexorably, while the sample stability and the interspectral variation of the chemical shifts may become an issue. In order to retain high spectral resolution using short measurement times, reduced dimensionality experiments (Szyperski et al., 1993a, 1993b; Simorre et al., 1994; Brutscher et al., 1994), where the frequency of two nuclei – or more – is identified along a single dimension, have been proposed and processed with projection-reconstruction methods (Kupce and Freeman, 2004). The last step towards this direction, if sensitivity permits, relies on single-scan multidimensional experiments (Frydman et al., 2002) using pulsed field gradients.

While some NMR experiments exhibit *per se* a limited sensitivity due to the multiple coherence transfers (this is referred as the ''sensitivity limited'' data acquisition regime) (Szyperski et al., 2002), the overall length of the other ones is mainly dictated by the sampling of indirect (t_1 ...) dimensions (''sampling limited'' regime). Achieving high resolution for a wide spectral range is a time-demanding constraint as a result of the

*To whom correspondence should be addressed. E-mail: Dominique.Marion@ibs.fr

Nyquist sampling theorem: it states that a signal must be sampled at a rate $\nu \geq 2\nu_{max}$, where ν_{max} is the Nyquist frequency, i.e. the highest frequency detected in the spectrum. In contrast, the digital resolution of the transformed spectrum is determined by the overall acquisition length in the time-domain. A wide NMR spectrum will thus be well resolved if a *large* number of points are collected at a *high* rate. Besides, in a non-constant time experiment, the effective S/N ratio is higher for the initial points than for the last ones, however at the same cost in terms of measuring time. Rovnyak et al. (2004) have recently shown that the optimal resolution in an indirect dimension would require sampling up to $\sim 3\ R_2^{-1}$ while the maximum signal/noise ratio is obtained for $\sim 1.2\ R_2^{-1}$, where R_2 refers to the relaxation rate of the nuclei in the transverse plane. These rules are demanding at higher fields because the spectral widths increase while relaxation often exhibits little change.

In order to relax these two antinomic requirements, one can either numerically extend the acquired signal or modify the acquisition scheme. The first method called LP extrapolation (Tang and Morris, 1988) is very routinely used in processing 3D NMR spectra to reduce truncation artefact, despite the fact that it may introduce false peaks and frequency shifts (Stern et al., 2002). In non-linear sampling schemes, a faster rate can be used to record the first data points and a reduced rate for the last ones, in an attempt to compromise S/N ratio and resolution. For sake of clarity, we will distinguish two types of non-linear sampling: in the first one, the sampling grid of the linear scheme is kept ($t_i = k\Delta t$ with $k \in \mathbb{N}$), but a increasing number of points are omitted towards the end of the signal (this method will be later referred as the *incomplete linear grid*); a more general non-linear scheme uses a *continuously varying pitch*, where $t_i = \alpha$ with $\alpha \in \mathbb{R}$. In both cases, the standard discrete Fourier transform (DFT) algorithm may no longer be used and alternative processing methods are required. In the early days of NMR maximum entropy methods (MEM) have been used by Barna et al. (1987) to take advantage of a non-linear (exponential) sampling scheme. More recently, Hoch and coworkers (Hoch and Stern, 2001; Rovnyak et al., 2003, 2004) have analyzed in detail MEM processing of data sampled on an incomplete linear grid. The three way decomposition method [as implemented in the program MUNIN (Orekhov et al., 2001)] could also be considered for such data. In contrast, it should be mentioned that other substitute to regular FT such as Burg MEM algorithm, parametric linear prediction (LP) (Gesmar and Led, 1989) or Filter Diagonalization Methods (FDM) (Hu et al., 2000), does not seem – to our knowledge – suitable for non-uniform sampling in their current implementation.

The present paper would like to address the following question: can NMR spectra acquired in a non linear manner be "FT processed" as regular ones? This question follows from the observation that most spectroscopists strike a balance between efficiency and complexity and are prone to favor a less powerful method if remarkably easier to handle. As an example, MEM is less widely used than LP interpolation (Stern et al., 2002), although the former has been reported as more reliable than the latter. By nature, FT boils down to computing a definite integral over a time variable, or numerically to integrating a function over steps, which could be chosen of arbitrary length. Generalized FT algorithms are now available (see Potts et al., 1998 for a review) to *exploit* or *produce* non-equispaced data, without any iterative procedure. In this paper, we investigate the potentialities and limitations of NMR data sampled in a non-linear manner (continuously varying pitch) and processed with the FT algorithm of Dutt and Rokhlin (1995).

Materials and methods

The outline of this generalized FT algorithm is the following: a Lagrange interpolation is used to recast the irregularly sampled FID into a regularly sampled one which is then Fourier transformed in a conventional manner. From a set of n-points, the Lagrange method uses a interpolating polynomial of degree $n - 1$ that passes through the n data points to compute the value of the function at any intermediate value. Although described by Lagrange at the end of the 18th century, the formula seems to have been published earlier by Waring and Euler.

Let us consider a time-domain NMR signal sampled up to time $t = t^{max}$. In NMR, no signal exists prior to excitation and the origin of the time scale is defined by convention at the beginning of the detection ($t \geq 0$). The signal (f_j, $j = 1,\ldots, N$) is sampled at N *non-equispaced* time points $\{x_1,\ldots,$

$x_N\}$. From this data set, we intend to compute a regularly sampled FID (g_i, $i = 1,\ldots, M$) made of M points corresponding to time points $\{y_1,\ldots,y_M\}$, with $y_i = i \times \Delta y + t_0$. For simplicity, we will generally assume that $y_1 = x_1$ and $y_M = x_N$, but no further requirement on the x_i values has to be met (this corresponds to the variable pitch acquisition defined above). Data are interpolated over the interval $[y_1, y_M]$, but the algorithm could be slightly modified to extrapolate data *outside* this interval. Note that both $\{f_j\}$ and $\{g_i\}$ are complex signals. The frequency domain spectrum is obtained by regular fast FT from $\{g_i\}$ and phased using only a constant phase correction if $t_0 = y_1 = 0$.

As a result of the periodic properties of discrete FT, two conventions can be used for the input and transformed data which result in indices running over [0, $2k - 1$] or [$-k+1$, k]. Let us normalize the time variables over the interval [$-\pi$,π] rather than [0, $t^{\max}$] by defining $\tilde{x}_j$ and $\tilde{y}_i$ as:

$$\tilde{x}_j = \frac{2\pi}{t^{\max}} \times x_j + \pi \quad \text{and} \quad \tilde{y}_i = \frac{2\pi}{t^{\max}} \times y_i + \pi$$

Dutt and Rokhlin (1985) have shown that an equispaced signal (g_i, $i = 1,\ldots, M$) can be computed form the original signal (f_j, $j = 1,\ldots,N$) using the following expression:

$$g_i = \frac{(-1)^i}{N} \cdot \sum_{j=1}^{N} f_j \cdot \sin\left(\frac{N\tilde{x}_j}{2}\right) \cdot \left(\frac{1}{\tan\left((\tilde{x}_j - \tilde{y}_i)/2\right)} - i\right) \tag{1}$$

where *only* the last symbol 'i' refers to the square root of -1. Readers should refer to the original publication for the derivation of Equation (1). In terms of numerical complexity, inspection of Equation (1) shows that:

- all elements (except f_j) only depend on the two sampling grids. When a series of time domain data are processed, these constant coefficients can be evaluated only once.
- the computation of g_i involves three arithmetic operations: (1) a first scalar scaling of f_j, (2) a multiplication by a $N \times M$ matrix and (3) a final scalar scaling.

An important feature of the algorithm is the fact that M may differ from N. This means that the dwell time of the generated FID can be adjusted during the processing and is not irrevocably defined at the acquisition. This important feature, which permits to choose the spectral width after acquisition, will be discussed later in the context of the Nyquist sampling theorem.

The continuous Fourier transform is defined as an integral:

$$G(f) = \int_{-\infty}^{\infty} F(t) \exp(2\pi i \cdot ft) dt. \tag{2}$$

When it is replaced by a discrete Fourier transform, the integral is substituted by a sum,

$$G(f_k) = \sum_{j=0}^{N-1} F(t_j) \exp(2\pi i \cdot f_k t_j) \Delta t_j. \tag{3}$$

One can legitimately discard the various Δt_j coefficients, provided that they are equal, a statement which is only valid for regularly sampled data. In a more general case, the Δt_j coefficients should be preserved. Before being processed through the above mentioned Lagrange interpolation, the non-linear sampled data were scaled according to the actual length of the corresponding dwell time. This preprocessing is analogous to the first data point scaling by 0.5 in regular FT, as required when acquisition starts at t_1 or $t_2 = 0$ (Otting et al., 1986).

This generalized FT method processes 1D vectors of multidimensional NMR spectra and with special benefits for indirect dimensions. Furthermore, as the artefacts described below are proportional to the signals, this method is less suited for homonuclear correlation spectra such as NOESY or TOCSY, because of strong diagonal peaks. As a test case, ^{15}N–^{1}H HSQC spectra on a ^{15}N labeled 124-residue protein were recorded on a Varian INOVA 600 spectrometer. Among possible non-linear acquisition schemes, we have chosen a parabolic one: $x_j = (j-1)^2 \Delta t'$. The lack of any transcendental function makes it easy to implement on any spectrometer. Spectra were processed using nmrPipe software (Delaglio et al., 1995). The generalized FT involves two subroutines implemented as stand-alone C-programs that can be inserted in the processing using the UNIX filter philosophy of nmrPipe: the scaling of the data (in order to take into account the variable dwell time) and the Lagrange interpolation. The software is available upon request from the author for various UNIX platforms (SGI, MacOS, SUN, Linux).

Results and discussion

Six HSQC experiments were recorded, which only differ in the way the indirect dimension delay (t_1) is incremented. Table 1 summarizes the acquisition and processing parameters for these experiments labeled A–E and identified by means of the same symbols (○, ●, ◇, □, ■, ▩) throughout the rest of this paper. The timing of these acquisition schemes are shown in Figure 1: (A, ○), (B, ●) and (C, ◇) correspond to a standard scheme with respectively 256, 64 and 64 data points and (D, □), (E, ■) and (F, ▩) to non linear (more specifically parabolic) sampling with 256, 64 and 64 points. It should be noticed that (*i*) scheme (E) is not a subset of (D) in contrast to (B) with respect to (A) and (*ii*) the curves for (D) and (E) intersect those associated with linear sampling (respectively here at points 64 and 8). In other words, such a general scheme leaves the spectroscopist more or less free to choose when to sample the data points, provided that enough points at the very beginning (1/4 of the FID, in the present case) are sampled *faster* than required by the Nyquist theorem. If one switches from time- to frequency-domain, the correct location of the peaks in the spectrum (i.e. the lack for aliasing) is secured by the fast sampled points at the beginning and the resolution of the individual signal (peak sharpness) by the less frequent points at the end. The following discussion will contain three parts: (i) the spectra obtained by this new method will be first compared with standard techniques, (ii) the artefacts associated with the data interpolation will be analyzed on simulated data and (iii) a new feature provided with non-linear sampling (the possible adjustment of the spectral width when spectra are processed) will be discussed.

Comparison with other techniques

The length of the acquired signal t_1^{max} is 0.121, 0.030, 0.483 and 0.236 s, respectively for schemes (A), (B), (D), (E), but the global experimental time of (A) and (D) is four times longer than (B) and (E). The data sets have been processed as shown in Table 1, (A, ○), (B, ●) using regular FT, (C, ◇) using 2D-maximum entropy reconstruction (MEM) without deconvolution and (D, □), (E, ■) and (F, ▩) using the Lagrange interpolation along t_1. Sets D, E and F cannot be processed using MEM reconstruction which generally involves an inverse DFT of the data (Stern et al., 2002) and are only suited for data sampled on an incomplete linear grid. The spectral region corresponding to the side-chain NH_2 are shown in Figure 2, together with an expansion of two overlapping signals. In the upper part of Figure 2 are displayed the "linear" spectra (a, b and c) and in the lower one the non-linear data sets (d, e and f). All spectra were plotted with the same first contour level ($\pm 10^6$) except d and e (at $\pm 0.2 \times 10^6$).

The spectra obtained with these different techniques are graded in terms of resolution (line-width and peak position accuracy) and sensitivity (peak amplitude, noise level). The contour plots shown in Figure 2 evidence that spectrum d (256 points – non-linear sampling) is the best resolved followed by spectrum e (64 points – non-linear sampling). These spectra correspond to $t_1^{max} = 0.483$ s and 0.236 s, respectively. The two data sets recorded in a standard manner with the same number of data points (Figure 2a and b) exhibit markedly lower resolution. When processed by MEM, some resolution can be partially recovered for the latter data

Table 1. Summary of the experiments discussed in the text.

Experiment[a]		Acquisition	Processing	Data points[b]	Scans[c]	Total spectrometer time[d]
A	○	Linear	FT	256	1×	1
B	●	Linear		64	1×	1/4
C	◇	Linear	2D-MEM	64	1×	1/4
D	□	Non-linear	Lagrange interpolation + FT	256	1×	1
E	■	Non-linear		64	1×	1/4
F	▩	Non-linear		64	4×	1

[a]The same symbol (or letter) is used for an experiment throughout this paper, i.e. in the definition of the acquisition scheme (cf. Figure 1), for the contour plots shown in Figure 2 and for the quantification of the spectral parameters (cf. Figure 3).
[b]Number of complex pairs (as required for quadrature detection).
[c]This indicates how many times the minimal phase cycling has been repeated for a given t_1 increment.
[d]All overall experimental times are arbitrarily scaled relative to experiment A.

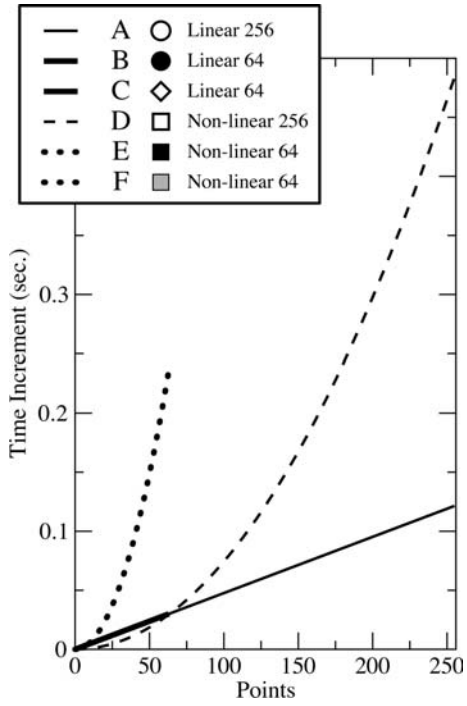

Figure 1. Sampling schemes (t_1 increment value vs. experimental *pairs* of data points) used for the indirect dimensions of ^1H–^{15}N HSQC: Schemes a and b show a standard linear acquisition [$t_j = (j-1)\Delta t$] and schemes d and e a non-linear one [$t_i = (j-1)^2\Delta t'$]. Note that the later schemes correspond to continuously varying pitches, as defined in the text. See Table 1 for more details and the definition of the various symbols (●, ○, ■, ▩, □, ◇). For each t_1 increment, two data points are sampled to lead to a complex signal and permit quadrature detection.

set (cf. Figure 2c vs. b). From this survey, one concludes that resolution is determined by the t_1^{max} value rather than the actual number of sampled data points, if these two parameters can be disconnected. The resolution enhancement provided by non-linear sampling is not associated with peak distortions as commonly seen with strong apodization. One-dimensional traces show that the negative tails on either edge of the peaks are less pronounced, since the apodization window is shifted towards longer t_1 increments.

The nmrDraw program was used for analyzing these spectra (Delaglio et al., 1995): the noise levels are given in Figure 2 caption and the position, line-width and height of 18 peaks reported in Figure 3(b, c and d). Let us first discuss the *position* of the peaks (chemical shift). For convenience, the position in each spectrum has been compared to the best resolved set (256 non-linear □) and reported in Figure 3(a). The position derived from the non-linear schemes are clearly not biased with respect to standard FT technique. The variations shown for peaks #5 and #6 can be easily rationalized by the fact that the peak picking routine is not able to distinguish them for low resolution spectra and thus report a common average value. Note that position fluctuations for the MEM processing are much larger, but this point is clearly beyond the scope of this paper.

We now turn to the *noise levels* in the various spectra. Different processing may introduce inconsistent scaling of the data (the coefficients involved in the definition of FT vary from one author to another) and the noise level is a sound basis for comparison. The noise ratio between spectra 2a and 2b is close to 2.0, as expected from the experimental time and the number of increments (4×). In contrast, spectrum 2d recorded in the same experimental time as 2a contains about 30% more noise: because noise contributes equally to all points of a free induction decay (even if the signal varies), acquiring data points at longer t_1 values, as done in the non-linear scheme, does not modify the *absolute* noise level. Consequently this additional noise is introduced by the processing due to the errors introduced by the Lagrange interpolation. Because the damped sinusoids are approximated as polynomials, part of the signal is displayed at an incorrect frequency and turns into noise.

The *peak heights* (Figure 3d) and *line-widths* (Figure 3c) are now dissected. The term "line-width" does not refer to the natural spectral line-width of the peak but to the digital one, which is usually larger in indirect dimensions. A single lorentzian line is characterized by its integral, which is proportional to the product of the peak height and its line-width. Because it is related to the first data point in the time-domain ($t_1 = 0$), the integral remains constant in the Fourier transformed spectrum, when the experimental number of points varies. Consequently, the peak height and line-width vary in opposite manner depending upon the digital resolution, provided that the FT implementation does not contain *hidden* scaling factors. Such an anticorrelated variation is observed for the linear acquisition (cf. ●, ○ in Figure 3) where the peak heights increase by a factor of 3 when 256 points instead of 64 are acquired, while the line-widths decrease by the same amount. When comparing linear with non-linear data acquired in the same amount of experimental time (256 linear ○ vs. 64

non-linear (4× scans) ▩), a similar line-width is obtained associated with a slight decrease of the peak height for the non-linear experiment (compare the corresponding spectra in Figure 2a and e). Correlation plot of the products (height × line-width) of the peaks have been drawn (data not shown) to compare the various schemes. If the peaks which cannot be discriminated by the peak picking routines are discarded (peaks #5, #6, #14 and #16 in Figure 3a), a correlation coefficient greater than 0.98 is found; compared to the 256–linear data set (○), the products (height × line-width) are scaled by 0.22 and 0.70 for the 256–non-linear (□) and the 64 non-linear (4× scans) (▩) respectively, although the acquisition of the 3 sets requires the same amount of spectrometer time.

Two conclusions can be drawn from this analysis: (i) the good correlation coefficient shows that the non-linear processing does not introduce any bias in the evaluation of peak heights and line-widths (apart from a global scaling) and is thus suitable for quantitative NMR. (ii) The price to pay in terms of sensitivity for using a non-linear sampling remains moderate, with a 30% increase of the noise and a similar decrease of the peak height. If the intrinsic sensitivity of the experiment permits, a much higher resolution (Figure 3c) can be obtained with non-linear sampling (compare ■ and ●, □ and ○) while sacrifying intensity (Figure 3d) for a constant spectrometer time. Depending upon the goal in terms of resolution and sensitivity, the non-linear schemes can be

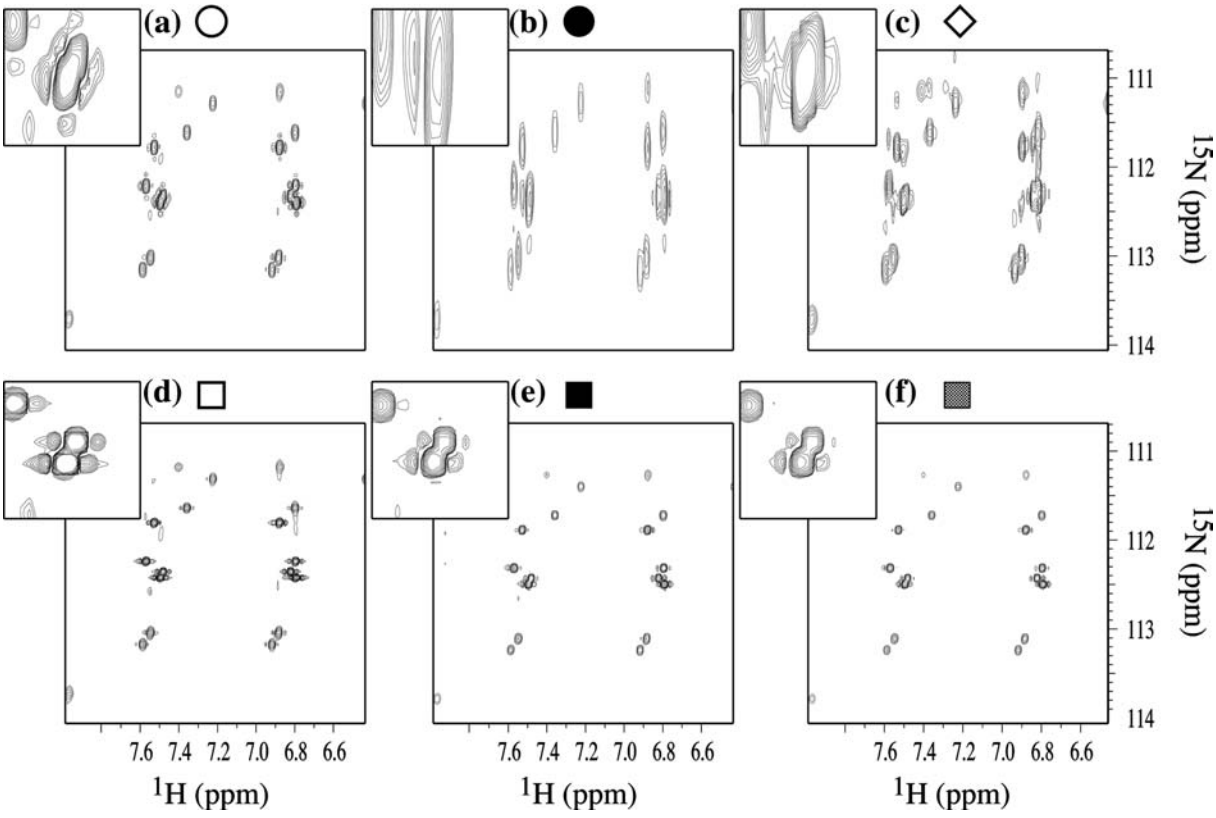

Figure 2. $^{1}H-^{15}N$ HSQC spectra of protein SufA from *E. coli* (Ollagnier-de Choudens et al., 2003), a protein involved in Fe–S cluster biosynthesis. Spectra shown in insets a, b and c have been obtained with corresponding linear sampling schemes a, b and c (cf. Figure 1) (F_1 spectral width = 2100 Hz) and spectra in insets d, e and f with non-linear schemes d, e and e. The same number of scans per increment is used for all spectra, except for f where it was multiplied by 4: the global experimental time is thus identical for spectra a, d and f as well as for b, c and e. Both positive and negative contour levels are plotted, starting at 10^6 for a, b, c, f and 0.2×10^6 for d and e. For reference, the noise level is respectively 13.7, 6.4, 17.6 and 19.7×10^3 for spectra a, b, d and e (evaluating the noise is not relevant for MEM spectra). For each spectrum, an expansion of the partially overlapping cross-peaks at $(^{1}H,^{15}N) \approx$ (7.5 ppm, 112.4 ppm) is shown. Spectra a and b have been processed with regular FFT (a $0.2 \times \pi$ shifted sine-bell was used to avoid truncation effect), c using 2D MEM reconstruction (as implemented in nmrPipe) and d, e, f using non-linear FT (this algorithm) along F_1. Data were acquired on a INOVA-600 MHz using Varian BioPack pulse sequence without any modification.

optimized as shown from the comparison of the 256- and 64 non-linear sets.

Artefacts due to interpolation

Because no processing technique is actually devoid of artefacts, it is essential to characterize them for lack of being able to suppress them completely. For instance, it is common practice to tolerate some weak "sinc wiggles" in FT spectra, in order to preserve the information contained in the last data points of the FID. In order to characterize the artefacts introduced by the Lagrange interpolation, time-domain data containing a single decaying frequency have been generated and processed with our technique (cf. Figure 4). Here the decay rate and the frequency of the signal have been modified while the acquisition scheme is kept unchanged (Scheme D, as

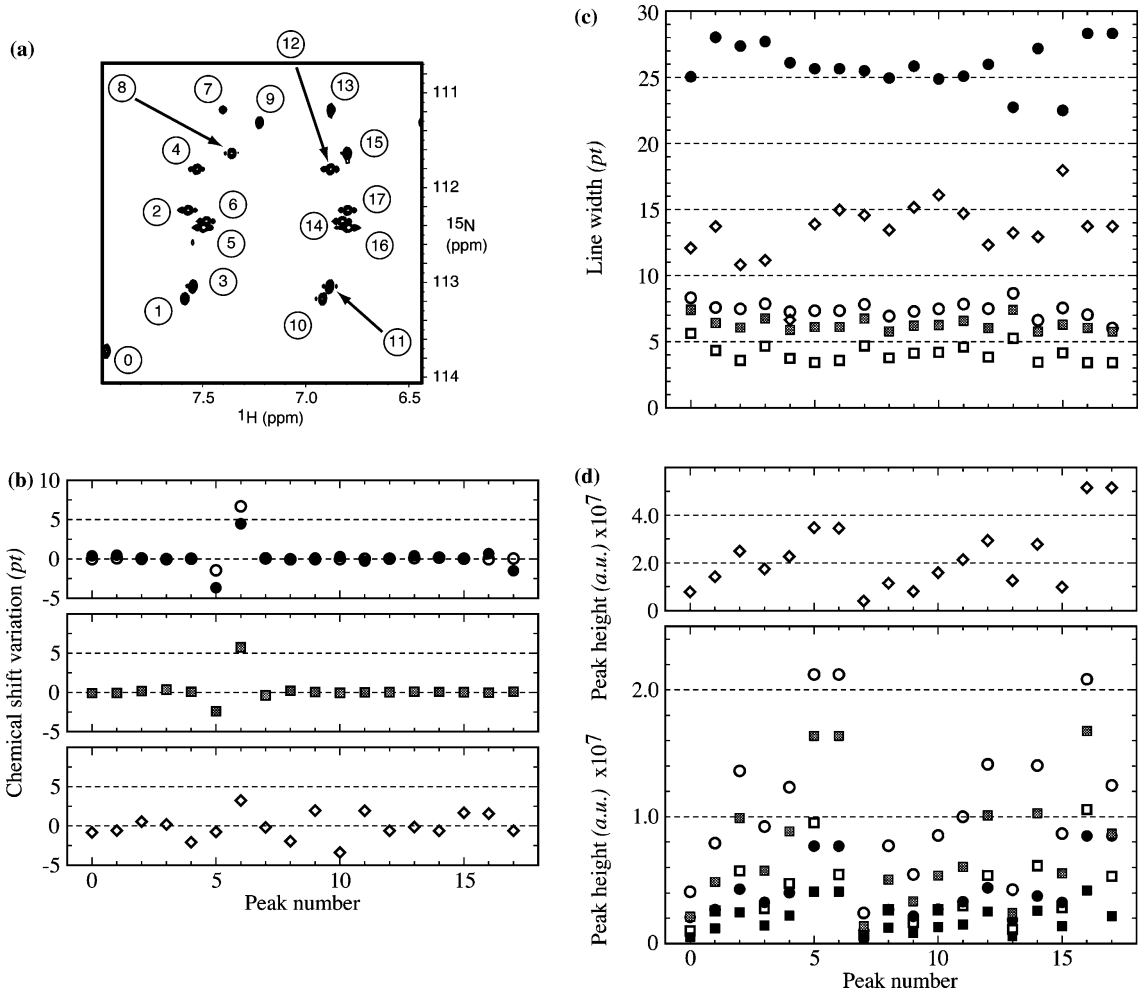

Figure 3. Comparison of the spectral parameters for 18 cross-peaks in the various $^{1}H-^{15}N$ HSQC spectra shown in Figure 2. The peaks are numbered in inset (a) on the same spectrum as Figure 2 d. All spectra were zero-filled to 2048 complex points in F_1 (spectral width = 2100 Hz) to yield to the same digital resolution (points/Hz) and thus permit a direct comparison. These spectral parameters have been obtained using the peak picking routine implemented in the program nmrDraw (Delaglio et al., 1995) without any editing. Circles (●, ○) correspond to regular FT, squares (□, ▒, ■) to non-linear FT and diamond (◇) to MEM processing, as defined in Table 1. The chemical shift variations (in data points), the line widths (in data points) as well as the peak heights are plotted in Figure 3(b, c and d), respectively. The peaks heights (Figure 3) and the contour levels (Figure 2) are reported in the same arbitrary units. The chemical shift variations (inset b) are expressed as a difference with respect to the best resolved spectrum (i.e. non-linear acquisition of 256 points). The large chemical shift fluctuation for peaks #5 and #6 results from their partial overlap, which prevents the peak-picking routine from correctly discriminating them. The peak heights obtained from MEM processing are apparently scaled by a constant factor with respect to the other spectra.

defined in Table 1). With this acquisition scheme, more data points have to be interpolated at the end of the FID on the basis of fewer experimental samples: it is therefore anticipated that broad and narrow signals may be affected differently, as well as low- and high-frequency peaks.

Spectra shown in Figure 4(a, b, c and d) illustrate how the interpolation performs when the signal becomes broader [for sake of clarity, a vertical expansion of spectrum (c) is reproduced in (c′)]. These 4 signals have the same integral and thus their amplitude decreases as the line-width increases. One should first notice that the base-line in the direct neighborhood of the signal remains perfectly flat and that the artefacts are clustered further away from the peak. Though these signals exhibit the same integral, the artefact *amplitude* is clearly larger for a narrow signal (spectrum a). However, the amplitude ratio (artefact/peak) remains constant (about 10%) for a wide range of line-width. The larger artefact amplitude in a vs. d is clearly related to the fact that the interpolation process is more demanding for slowly vanishing signals. Let us discuss the artefact pattern by comparing spectra c′ and a: when the signal is still reasonably broad, the Lagrange interpolation leads to a simple oscillation (in the frequency domain) which amplitude smoothly increases as one moves away from the peak. For narrow signal, a more complex oscillation (i.e. with higher frequencies) is observed and the transition becomes sharper. If one now focuses on signals at higher frequency (spectrum f vs. c), the artefact retains the same amplitude for a given line-width, but the oscillation pattern contains more components for high-frequency signal. Its maximum amplitude occurs roughly at $\nu \pm \Delta\nu/2$, where ν is the signal frequency and $\Delta\nu$ the reconstructed spectral width. When white noise is added to the time-domain data (Figure 4e), it adds to the oscillation artefact in a uniform manner, i.e. the standard deviation is the same at any location in the spectrum.

In conclusion, these simulations show that the Lagrange interpolation introduces some artefacts, because approximating a sinusoid function by a polynomial – even of high order – remains a crude approximation. The artefacts never occur in the direct vicinity of the true peak; they are made of more or less complex baseline oscillations with a amplitude less than 10% of the main peak, whatever its line-width is. The experimental noise adds to these artefact without introducing any bias. Although their presence remains annoying, these artefacts cannot be mixed up with real peaks because of their very specific shape. A more detailed artefact analysis for different acquisition schemes are currently in progress. These simulations show that this method is more suitable for heteronuclear correlation spectra (which contain peaks of similar amplitude but no diagonal peaks) than for homonuclear NOESY spectra (with a high dynamic range).

Adjustment of the spectral width

The Lagrange interpolation applied to non-linearly sampled signals allows one to modify the spectral

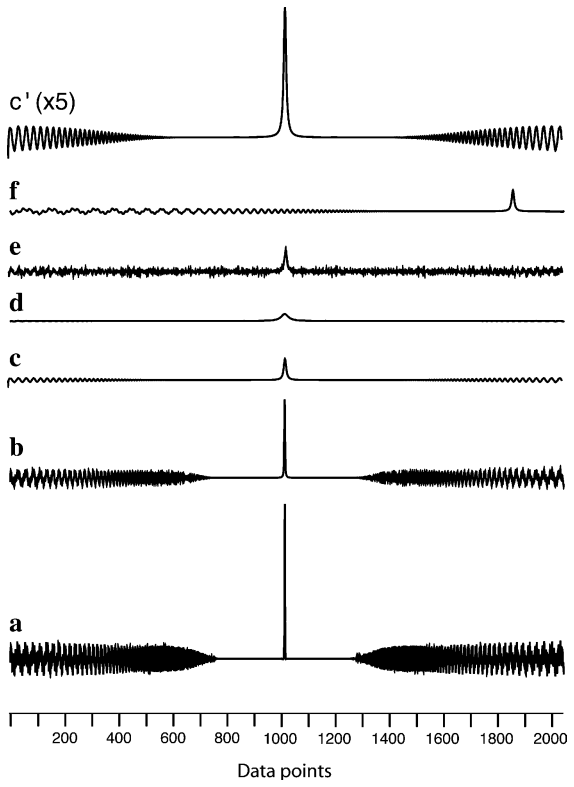

Figure 4. Evaluation of artefacts on simulated data. Time domain data containing a single decaying frequency $\mathbf{S}(t) = \exp(i2\nu t) \times \exp(-t/T_2)$ were generated using sampling scheme (a) (cf. Table 1). The time-domain decay rate corresponds to a signal line-width of 2 points (spectrum a), 3.5 points (b), 11 points (c, e and f) and 33 points (d). In all spectra (except F), the signal is centered on the carrier frequency, while it is shifted to the edge of the spectrum in (f). Spectrum (e) is similar to spectrum (c), except for the white noise added to the simulated data (the amplitude of the noise is 10% of the magnitude of the first data point in the time-domain). Spectrum (c′) shows a vertical expansion of spectrum (c) (5×).

width *during* processing and adjust it to fit the spectral requirements. In the previous statement, we only refer to the scheme called the continuously varying pitch. If one recalls the Nyquist sampling theorem quoted earlier, it might seem at first glance that the spectral width is irreversibly encoded by the dwell time chosen at the acquisition. This golden rule is known by NMR spectroscopists who have all experienced one day or another the effect of signal aliasing: in such situations, the peaks which should fall outside of the chosen spectral width are "folded back" into the spectrum and generally identified by their phase distortion, if the acquisition timing has been slightly delayed.

Equation (1) shows that the number of complex points (M) computed using the Lagrange interpolation can be chosen freely and is not related to the amount of sampled points (N). A data set acquired using the timing shown in Figure 1(d) (■) has been processed in 3 different manners and contour plots with the *entire* spectral width in F_1 (^{15}N) are displayed in Figure 5. The data were interpolated using respectively 256 pairs (inset c), 512 (b) and 1024 (a). In contrast to linear sampling where they are folded back at an apparently incorrect frequency, the signal at higher frequencies are spread over the entire spectrum and averaged out. This smearing is very analogous to that observed in NOESY spectra, where a random variation of the length of the mixing time is used to suppress the zero-quantum contribution (Macura et al., 1981).

Why the peaks at higher frequencies behave differently in linear and non-linear sampling can be rationalized in the following manner. The aliasing observed with classical FT is due to the lack of any information in the FID, which can provide the authentic frequency of the folded signal: *all* the digitized samples for two signals oscillating at ν_0 and $\nu_0 + \nu_{max}$ (where ν_{max} is the Nyquist frequency) *exactly* coincide. In contrast, in the present method, the pitch (i.e. the spacing) between the digitized samples varies continuously: if two arbitrary frequencies are considered, ν_1 and ν_2, the amplitude of a signal of normalized intensity may be accidentally identical for a few data points, but *not* for all of them. This intuitively explains the key difference between the two techniques, as far as the behavior of higher frequency signal is concerned. This statement deserves two additional comments: (i) the lack of periodicity between any frequency and the sampling grid is only true for the continuously varying pitch, ($t_i \in \mathbb{R}$) but not for the incomplete linear grid method ($t_i = k\Delta t$ with $k \in \mathbb{N}$). (ii) one should not trust outward appearances that the Nyquist theorem might be violated in the present case: the first data points have to be sampled faster than stated by this theorem for the largest frequency of interest.

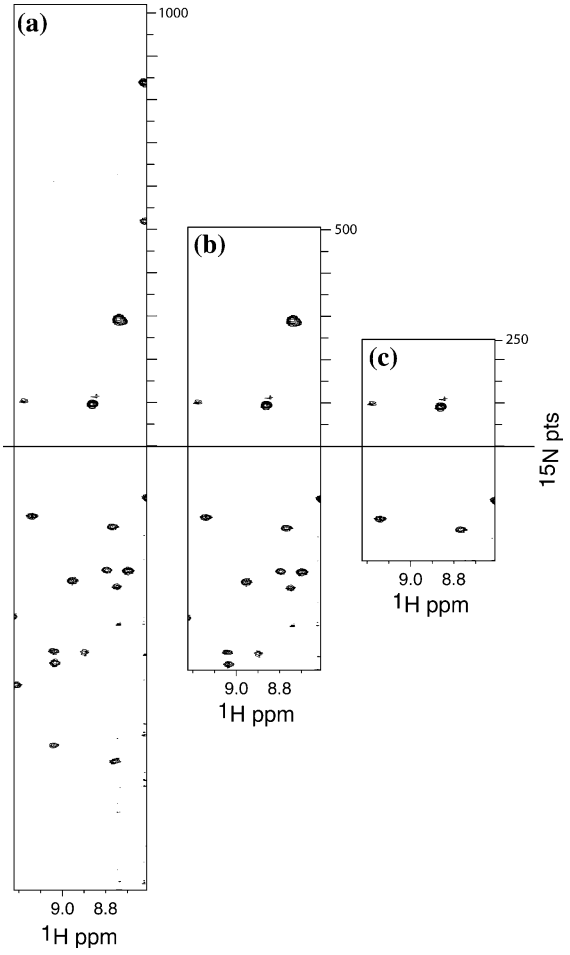

Figure 5. Comparison of different processing of the same data set recorded using non-linear acquisition. The Lagrange interpolation (see material and methods) allows the computation of an equispaced FID of M complex data points from a non-equispaced FID of N complex points. A ^{1}H–^{15}N HSQC spectrum was acquired with the non-linear scheme shown in Figure 1a ($N = 256 \times 2$ points) and interpolated to $M = 1024 \times 2$ points (inset a), 512×2 points (inset b) and 256×2 points (inset c). The 3 spectra are plotted with the same contour levels (positive + negative) and with the full ^{15}N spectral width. When compared to the normal ^{15}N spectral width required with a linear acquisition scheme ($SW_1 = 2100$ Hz), the apparent spectral widths are respectively 525 Hz (inset a) 1050 Hz (inset b) and 2100 Hz (inset c). Noise ridges parallel to F_1 (see text) are visible (on the lower right-hand corner of c) as a result of the approximation of sinusoids by polynomials.

Conclusions

In summary, the continuously varying pitch method, combined with the Lagrange interpolation [as implemented according to Dutt and Rokhlin (1995)] provides better resolved spectra than classical methods, for a given amount of experimental time. The first fast sampled data points allows to identify correctly the global position of the peaks while the points sampled for longer increments at a lower rate leads to precise frequency characterization. We have shown that this rather robust method does not require any parameter adjustment (in contrast to other iterative methods) except the choice of the spectral width of interest. It does not introduce any salient bias in the derived spectral parameters and the price to pay in terms of sensitivity is moderate. In addition to random noise, some systematic oscillation ridges come in sight shifted by half of the spectral width with respect to the original peak, because the approximation of damped sinusoids by polynomials is not perfect. Further analysis of these artefacts will be discussed elsewhere. Combined with cryogenic probes, this method could be implemented in any indirect dimension of (reduced dimensionality) 3D and 4D experiments to speed up acquisition on perdeuterated proteins. Other applications involve kinetics measurements, fast recording of 2D experiments of small organic compounds or even correction of corrupted data due to ADC overflow.

Acknowledgements

The SufA protein was provided by S. Ollagnier-de Choudens and N. Duraffourg (CEA–Grenoble). The author would like to thank Dr. Frank Delaglio (NIH) for providing assistance and technical details on *nmrPipe* software.

References

Barna, J.C.J., Laue, E.D., Mayger, M.R., Skilling, J. and Worrall, S.J.P. (1987) *J. Magn. Reson.*, **73**, 69–77.

Brutscher, B., Simorre, J.-P., Caffrey, M. and Marion, D. (1994) *J. Magn. Reson. B*, **105**, 77–82.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Dutt, A. and Rokhlin, V. (1995) *Appl. Comp. Harm. Anal.*, **2**, 85–100.

Frydman, L., Scherf, T. and Lupulescu, A. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 15858–15862.

Gesmar, H. and Led, J.J. (1989) *J. Magn. Reson.*, **83**, 53–64.

Hoch, J.C. and Stern, A.S. (2001) *Meth. Enzymol.*, **338**, 159–78.

Hu, H., De Angelis, A.A., Mandelshtam, V.A. and Shaka, A.J. (2000) *J. Magn. Reson.*, **144**, 357–366.

Kupce, E. and Freeman, R. (2004) *J. Am. Chem. Soc.*, **126**, 6429–6440.

Macura, S., Huang, Y., Suter, D. and Ernst, R.R. (1981) *J. Magn. Reson.*, **43**, 259–281.

Otting, G., Widmer, H., Wagner, G. and Wüthrich, K. (1986) *J. Magn. Reson.*, **66**, 187–193.

Ollagnier-de Choudens, S., Nachin, L., Sanakis, Y., Loiseau, L., Barras, F. and Fontecave, M. (2003) *J. Biol. Chem.*, **278**, 17993–18001.

Orekhov, V.Y., Ibraghimov, I.V. and Billeter, M. (2001) *J. Biomol. NMR*, **20**, 49–60.

Potts, D., Steidl, G. and Tasche, M. (1998) In *Modern Sampling Theory: Mathematics and Applications*. Benedetto, J.J. and Ferreira, P., (Eds.), Birkhäuser Publishing Ltd. Basel, Switzerland, 249–274..

Rovnyak, D., Hoch, J.C., Stern, A.S. and Wagner, G. (2004) *J. Biomol. NMR*, **30**, 1–10.

Rovnyak, D., Filip, C., Itin, B., Stern, A.S., Wagner, G., Griffin, R.G. and Hoch, J.C. (2003) *J. Magn. Reson.*, **161**, 43–55.

Simorre, J.-P., Brutscher, B., Caffrey, M. and Marion, D. (1994) *J. Biomol. NMR*, **4**, 325–333.

Stern, A.S., Li, K.-B. and Hoch, J.C. (2002) *J. Amer. Chem. Soc.*, **124**, 1982–1993.

Szyperski, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993a) *J. Biomol. NMR*, **3**, 127–132.

Szyperski, T., Wider, G., Bushweller, J.H., Wüthrich, K. (1993b) *J. Am. Chem. Soc.*, **115**, 9307–9308.

Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N. and Montelione, G.T. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 8009–8014.

Tang, J. and Morris, J.R. (1988) *J. Magn. Reson.*, **78**, 23–30.

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L.E. (2002) *J. Am. Chem. Soc.*, **124**, 10025–10035.