Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 186 (2007) 142–149
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Communication

Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins

Nicolas Pannetier, Klaartje Houben, Laurence Blanchard, Dominique Marion *

Institut de Biologie Structurale Jean-Pierre Ebel CEA - CNRS – UJF, 41, Rue Jules Horowitz, 38027 Grenoble Cedex, France

Received 12 December 2006; revised 10 January 2007
Available online 23 January 2007

Abstract

Resonance assignment of NMR spectra of unstructured proteins is made difficult by severe overlap due to the lack of secondary structure. Fortunately, this drawback is partially counterbalanced by the narrow line-widths due to the internal flexibility. Alternate sampling schemes can be used to achieve better resolution in less experimental time. Deterministic schemes (such as radial sampling) suffer however from the presence of systematic artifacts. Random acquisition patterns can alleviate this problem by randomizing the artifacts. We show in this communication that quantitative well-resolved spectra can be obtained, provided that the data points are properly weighted before FT. These weights can be evaluated using the concept of Voronoi cells associated with the data points. The introduced artifacts do not affect the direct surrounding of the peaks and thus do not alter the amplitude and frequency of the signals. This procedure is illustrated on 60-residue viral protein, which lacks any persistent secondary structure and thus exhibits major signal overlap.

Keywords: Non-linear sampling; Random acquisition; Discrete Fourier transform; Unstructured proteins; Voronoi cells

1. Introduction

There is an increasing body of evidence that a large fraction of eukaryotic proteins (>30%) are partially or completely disordered, while retaining some biological activity [1,2]. Such proteins are known as *natively unfolded* or *intrinsically unstructured* (IUP) and can either be fully disordered throughout their length or have large disordered regions embedded in a compact fold. Until recently, the tenet of structural biology was that proteins require a well-defined conformation to be biologically active. Proteins or protein regions with no predefined fold do however play important roles in numerous cellular processes and signaling events and are involved in protein deposition diseases (Alzheimer's, Parkinson's [3] and Huntington's diseases).

In contrast to X-ray crystallography, NMR is well adapted to the study of unstructured proteins in solution. Unfolded proteins are prone to both proteolytic cleavage and aggregation, but suitable experimental conditions (ionic strength, pH,...) can generally be found to prevent these problems: once a stable sample is prepared, one might ordinarily expect NMR spectra with reasonably narrow resonances as discussed below. This permits thus to investigate a number of spectral parameters (chemical shifts [4], scalar couplings, spin relaxation [5] and exchange rates) that report on the structural and dynamical behavior of the protein. However, it should be kept in mind that any interpretation at the atomic level requires specific resonance assignment. In this respect, unstructured proteins differ from globular ones: whereas the resonance line-width is generally more favorable, the spectral dispersion is unfortunately very limited. A disordered protein lacks specific secondary and tertiary structure and is composed of an ensemble of interconverting conformations with distinct φ and ψ angles. Two main contributions to chemical shift are

* Corresponding author. Fax: +33 476 88 54 94.
E-mail address: Dominique.Marion@ibs.fr (D. Marion).

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.01.013

identified: the first one originates from the covalent structure of the amino-acids and the second from the secondary and tertiary structure [6]. As a result of the averaging process taking place in unfolded proteins, only the first one remains and thus the chemical shift dispersion is rather limited for all nuclei (^{1}H, ^{13}C and ^{15}N). On the other hand, if the averaging process occurs on a fast time scale (typically less than a few ns), it decreases the efficiency of the dipolar relaxation in the transverse plane, which is dominated by $J(0)$ terms in well-folded proteins. Sequential resonance assignment of a doubly labeled (^{13}C and ^{15}N) protein is best achieved using a set of triple resonance 3D experiments (HNCO, HNCA, CBCANH, CBCA(CO)NH...) [7]. In this procedure, the chemical shifts in a given spin-system as measured in one experiment are matched with resonances found in another experiment, to in the end sequentially assign all resonance frequencies. Ambiguities arise typically when several combinations match, i.e. when several candidates can be found within a given frequency range. Therefore, assigning an unfolded protein is, for a given sequence length, a more challenging task than for a globular one. In this communication, we propose a modification of the acquisition and processing schemes to improve the resolution of 3D NMR spectra with the goal of easing the resonance assignment of unfolded proteins.

2. Theory

The conventional procedure for a multidimensional experiment consists of sampling the time domain on a Cartesian grid. This requires significant measuring time for three- and more dimensional experiments, as each new data point in an indirect dimension requires a new experiment. Alternatives to the conventional Cartesian sampling of the time domain, such as reduced dimensionality techniques [8,9] or radial sampling, were proposed more than 10 years ago to achieve better resolution in less time. For several years these data sets were FT transformed only along the radial dimension and interpreted as a sum and difference of two (or more) frequencies. More recently, projection reconstruction [10] was employed to restore spectra that can be visualized in a conventional manner, i.e. a single spin frequency per dimension. Realizing that reduced dimensionality experiments boils down to sampling data in radial coordinates, we and others [11–13] have proposed to Fourier transform them completely and rewrite the discrete FT algorithm in the appropriate coordinate system. Note that there is no rationale for replacing the regular acquisition scheme and the fast FT [14] in the acquisition dimension.

In a Cartesian coordinate system, the 2D Fourier transform is a double integral given by:

$$F(\nu_1,\nu_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t_1,t_2)\exp(-2\pi i\nu_1 t_1)\times\exp(-2\pi i\nu_2 t_2)\,dt_1\,dt_2 \quad (1)$$

To evaluate it numerically, this double integral can be replaced by the following double sum:

$$F(\nu_1^m,\nu_2^n)=\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} f(t_1^j,t_2^k)B_{jk}^{mn}\cdot\Delta t_1^j\cdot\Delta t_2^k \quad (2)$$

where $B_{jk}^{mn}=\exp(-2\pi i\nu_1^m t_1^j)\cdot\exp(-2\pi i\nu_2^n t_2^k)$

It has been shown that this expression can be recasted for other sampling schemes, by appropriately changing the coordinate system. Whereas the integration area remains constant for all points in the case of Cartesian sampling ($\Delta t_1^j\times\Delta t_2^k$), this is no longer true for an arbitrary acquisition scheme and views differ on whether or not the surface associated with each data point can be discarded. We have shown that for radially sampled data, neglecting this factor leads to a broadening of the signals due to incorrect weighting [13]. A radial sampling scheme is necessarily associated with some local undersampling in the azimuthal direction, a feature that leads to baseline-distortion away from the peak of interest [12]. These artifacts share the geometrical pattern as the acquisition scheme and could potentially be misinterpreted by a peak-picking program, at the locations where two ridges intersect.

Because these artifacts are intrinsically linked to the sampling scheme and not induced by the FT algorithm, alternated sampling patterns were proposed by Kazimierczuk et al.: spiral acquisition [11] and random acquisition[15]. In this communication, we further explore the random acquisition scheme with the aim of reducing the overall experimental time while preserving spectral resolution and avoiding spectral folding. As shown for radial sampling, the proper weighting of the time domain data has a definite impact on the quality of the frequency domain spectra. However, since the points are sampled in a random manner, the surface associated with one sample cannot be derived analytically but will be computed as a *Voronoi cell* [16]. Although Voronoi cells can be defined for Euclidian space of any dimension, we will focus here on a plane.

Let us consider a discrete set of points S ($s_a, s_b, \ldots$) in plane. The *Voronoi cell* (C_i) associated with the point s_i is made of all points x of the plane that are closer to s_i than any other points of the set S.

$$C_i=\{x:|s_i-x|\leqslant|s_j-x|,\forall j\neq i\} \quad (3)$$

Voronoi cells are also called Dirichlet domains and are described as a convex polygon in a 2D. Among several softwares able to evaluate the area of Voronoi cells for arbitrary sets of points, we have chosen "Qhull", a robust practical convex hull algorithm [17] developed at "The Geometry Center" of the University of Minnesota, USA. Note that this program can also handle higher dimensionality cases and could thus be used for randomly sampled data along three indirect dimensions.

To illustrate this method, 3D triple resonance experiments were recorded on an intrinsically unstructured protein, a 60-residue fragment (N_{TAIL}[443–501]) of the

nucleoprotein N from the paramyxovirus Sendai (NCBI #Q07097). In this virus, the nucleoprotein N packages the genomic RNA and N_{TAIL} interacts with the RNA polymerase during replication and encapsidation. In addition of being not folded, the primary sequence of N_{TAIL}[443–501] contains amino-acid repetitions (Glu-Glu-Glu and Arg-Arg ×2) that lead to severe spectral overlap.

3. Materials and methods

NMR spectra were acquired on a Varian Inova Spectrometer ($^1H = 600$ MHz) equipped with a Penta™ probe. Pulse sequences provided in the Biopack package were used ('gcbca_co_nhA.c') and modified to accommodate random acquisition by means of defining explicit d2 (t_1) and d3 (t_2) arrays. Prior to acquisition, the incremented delays are computed using a C-written program, that creates a macrofile to be read on the spectrometer. This program based on the GSL library (http://www.gnu.org) uses the function 'gsl_ran_Gaussian( )' that returns a Gaussian random variate for a given standard deviation. For practical reasons, this infinite distribution is truncated at a given threshold (at 50% height in the present case). The delays are defined with a resolution of 0.1 μs, a value much larger than the time base (12.5 ns) of the data acquisition controller board on the Varian Inova spectrometer. Fig. 1a shows the incremented delays used for both experiments: the largest increments are (0.0048 and 0.051 s) for Cartesian (^{13}C and ^{15}N) sampling and (0.0117 and 0.1386 s) for random sampling.

Using this random distribution of delays (t_1 and t_2), the associated area are evaluated using programs 'qvoronoi' and 'qconvex' which are parts of the Qhull package (http://www.qhull.org/). Caution should be taken to avoid any non-finite surface for the data points located close to the edges of the rectangular area depicted in Fig. 1a (smallest and largest t_1 or t_2 values). The following procedure (see Fig. 2) has been used for this purpose. The experimental

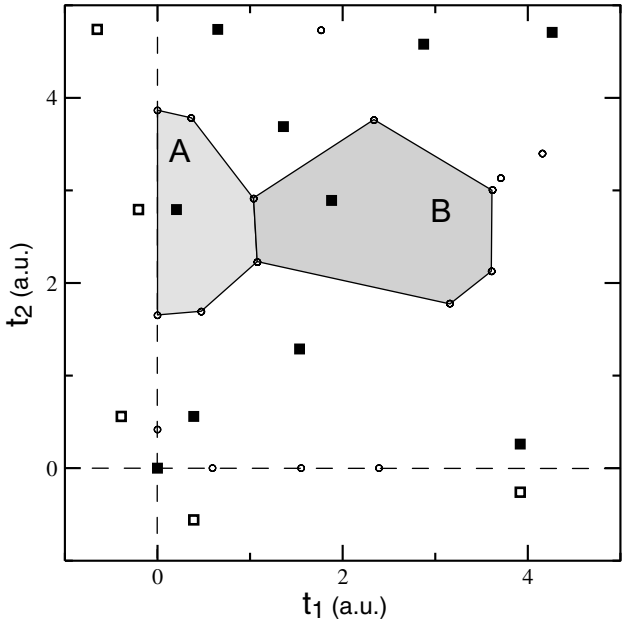

Fig. 2. Procedure to compute the area for data points located at the edge of the grid area. An expansion of the sampling pattern in the vicinity of the time origin ($t_1 \to 0$ and $t_2 \to 0$) is shown. The measured data points are indicated by filled squares (■). Dummy experimental points (open squares □) are generated by mirroring the experimental ones with respect to the two axes at $t_1 = 0$ or $t_2 = 0$. Using Qhull, the Voronoi cells and areas are evaluated and the vertices of the polygons (open circles ○) identified. Only two of them, labeled A and B, have been drawn for sake of clarity. Note that as a result of the mirroring procedure, one edge of polygon A overlaps with the origin axis. Due to the causality principle, the NMR signal is absent prior to the excitation pulse (i.e. for negative time) and this scaling of the data points thus fulfills this criterion.

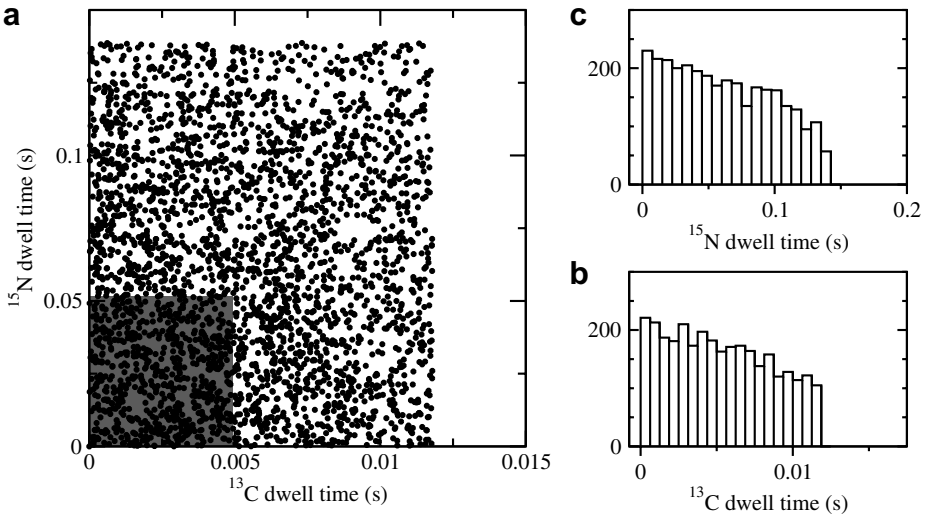

Fig. 1. Sampling scheme used for the CBCA(CO)NH experiment. (a) 48 × 64 (^{13}C and ^{15}N) data set was acquired for Cartesian sampling (gray area) and 3120 points for random sampling. For Cartesian sampling, the ^{13}C dwell time is 100 μs (SW = 10,000 Hz) and the ^{15}N dwell time is 800 μs (SW = 1125 Hz). Insets (b) and (c) show the dwell time distribution for both increments. As the number of increments is roughly identical in both experiments, the sampling density in the random experiment is roughly 15% of the Cartesian one.

data points are mirrowed with respect to the edges prior to the evaluation of the area of the Voronoi cells. As a result, the polyhedra associated with the points at the borders have two vertices on the edge. This method provides a correct scaling of all data points, including those close to the edge. Although the area evaluation is computationally demanding, it is performed only once for all (t_1 and t_2) planes of the 3D data set.

In contrast to previous sampling schemes, the two indirect time variables are no longer separable. Thus, instead of having a 2D series $(t_1(j), t_2(k))$ of FID, one has to deal with a 1D vector $(t_1(m'), t_2(m'))$ that will arbitrarily be stored into a 2D matrix of suitable size, to conveniently preserve the file architecture of nmrPipe. The (t_1, t_2, t_3) 3D data is first Fourier transformed (FFT) using nmrPipe along t_3, leading to a series of (t_1, t_2) "planes". These planes are apodized and processed using a modified version of the C-program described previously for radial sampling [13]. Whereas the area associated with each point was internally computed in the previous implementation, the Fourier transform program uses the Voronoi surfaces computed by Qhull. Therefore, the processing time for random acquisition is similar to that needed for a radially sampled set of the same size. These processing routines, to be used together with nmrPipe, are available from the authors for various operating systems (Linux, MacOS...). Resonance assignment was carried out using the graphical interface Smartnotebook [18] combined with nmrView on an iMacG5.

4. Results

Resonance assignment of N_{TAIL}[443–501] was carried out using mainly two data sets, a CBCA(CO)NH [19] experiment combined with a CBCANH experiment [20]. Connection between residues involves the comparison of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ frequencies in these two spectra, a process severely hindered by the poor spectral and digital resolution along the ^{13}C dimension. To solve this problem, it is necessary to sample more data points in the ^{13}C and ^{15}N indirect dimension. Fig. 1 compares the random sampling pattern used for the CBCA(CO)NH experiment (circles) with a standard Cartesian sampling (gray area). A Gaussian distribution (truncated at 50%) was used as shown on the histograms in Figs. 1b and c. Note that the overall acquisition length of both experiment is nearly identical.

Severe undersampling occurs globally for the random acquisition scheme, as the density of sampled data points is much lower (Fig. 1). Rather than quantifying it independently along each dimension (the dimensions are no longer separable for random sampling), the area associated with each sampled point provides a global estimate. For Cartesian sampling ($\Delta t_1 \times \Delta t_2$), this area is 0.08×10^{-6} s^2 while the *average* value for random sampling is 0.52×10^{-6} s^2 (i.e. 6.5 times larger). We will see later that, despite of this large value, no folding (as experienced with standard acquisition) occurs in the transformed spectra. Let us finally note that a fair number of data points for longer t_1 and t_2 values are associated with a quite large area (10% has an area larger than 1.0×10^{-6} s^2). It should be emphasized that various authors assign a different meaning to random sampling. Rovnyak et al. [21] consider a random set as a pattern of dots that fall on a periodic grid ($t_1 = k \times \Delta t_1$, with $k \in \mathbb{N}$), whereas we consider a set of delays that have no common denominator ($t_1 \in \mathbb{R}$). In practice, a common denominator does exist in the latter case, which corresponds to the resolution defined above (0.1 μs). Note that this value is roughly three orders of magnitude smaller than the dwell time: because *folding* at Nyquist frequency originates from the periodicity in the time-domain, we can safely forget this aspect in our case. As discussed below, this does not imply the lack of any aliasing artifacts.

In the experimental section, the procedure for evaluating the area for $t_1 \to 0$ and $t_2 \to 0$ data points has been described (Fig. 2). Due to the causality principle, the NMR signal is absent prior to the excitation pulse (i.e. for negative time) and this scaling meets this requirement. The FT theorem states that the first data point in the time domain is related to the spectrum integral: thus, an incorrect scaling of the early points will induce a DC offset in the spectrum. We will show later on that no offset is present in the spectra obtained using random sampling due to proper scaling.

Longer acquisition along both ^{13}C and ^{15}N dimensions should improve resolution. However, increasing the length of the indirect dimensions involves a slight modification of this pulse sequence, which employs a ^{15}N semi-constant time [22] and a ^{13}C constant time acquisition. For ^{15}N, this only involves recomputing the various delays of this building block on the basis of the largest increment used. For ^{13}C, the C^{α} and C^{β} coherences evolve during a delay $2T_{ab}$ (as defined in [20]) under the influence of the $^1J_{CC}$ couplings with carbons other than with the CO. The transfer function, which depends on the number of adjacent carbons, is reported in Fig. 3. Grzesiek and Bax [20] suggested to choose $T_{ab} = 3.3$ ms, a value that does not permit the sampling of data points for long ^{13}C evolution. Inspection of Fig. 2 shows that $T_{ab} = 9.9$ ms is an alternate choice that leads to efficient transfer to the C^{a} (then to the CO) for all residue types. The relative sign of correlation peaks depend on the number of neighbors and residues with a *single* C^{γ} (Glu, Gln, Lys,...) can be easily discriminated. A signal loss due to relaxation in the transverse plane is expected for a longer constant-time delay, but the impact remains limited for unstructured proteins.

Two CBCA(CO)NH data sets were recorded on protein domain N_{TAIL}[443–501]: one with the random sampling scheme proposed here and one with the cartesian sampling. Fig. 3 compares the two (^{15}N–^{13}C) planes for a crowded spectral area. In a well-resolved ^{1}H–^{15}N HSQC spectrum, 3 overlapping correlation peaks are visible at ($^{1}H = 8.41$ ppm, $^{15}N = 121.7$ ppm). Discriminating them in a 3D spectrum is more difficult along these dimensions because of the lower resolution. The lack of secondary structure—and thus of secondary

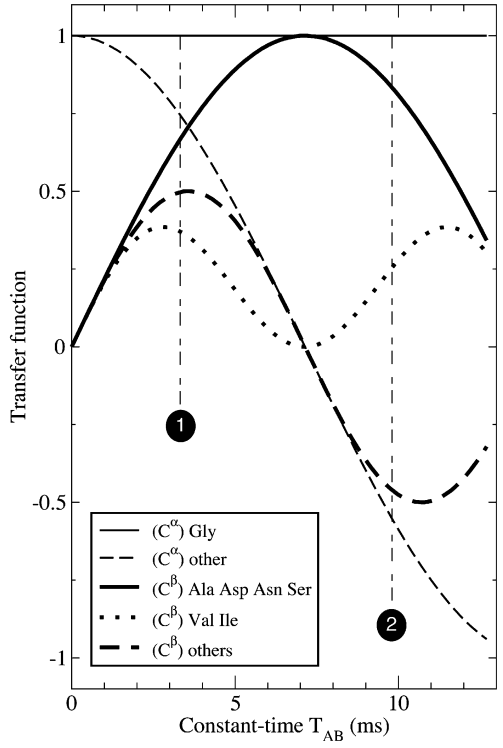

Fig. 3. Plots of the coherence transfer function (to the C^{α} starting either from the C^{β} or the C^{α}) versus the delay T_{AB}. These plots use Eq. (1) of Grzesiek and Bax [20] and ignores relaxation. The transfer function depends upon the number of adjacent carbons via the $^{1}J_{CC}$.

chemical shifts—leads also to severe overlap in the ^{13}C dimension. This feature is illustrated in Fig. 4 by the two sets of cross-peaks corresponding to Glu residues ($^{13}C \approx 57$ and 31 ppm). Insets b, c and e, f illustrate the resolution brought by the random acquisition scheme, where the frequencies can be more precisely determined in both ^{13}C and ^{15}N dimension. Due to the small chemical shift dispersion, a large number of possible candidates for pairing intra- and inter-residue correlation are found for any $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ frequencies. For instance, the C^{β} of Glu 469 and 474 (see Fig. 4), which are separated by less than 20 Hz, are difficult to discriminate along the ^{13}C dimension where a 10 kHz spectral width is necessary to cover the shifts of all types of residues.

The opposite side of the coin is the decrease of the signal-to-noise ratio for random acquisition (see the 1D slices in Fig. 4). Actually, the "noise" in data acquired randomly contains three components. The first one, or thermal noise, is present in any NMR experiment and is induced by all electronic components involved in the acquisition process (amplifiers, ADC...). The second, called t_1 noise in 2D NMR, is visible only along the indirect dimensions—which are sampled point-by-point—and originates from long-term fluctuations on the spectrometer. Strictly speaking, the last source should not be called noise but aliasing artifacts, because it can be predicted, knowing the acquisition scheme.

Simulated data without thermal and t_1 noise have been generated to assess the nature of the artifact. Let us consider a 2D oscillating damped signal:

$$f(t_1,t_2) = \exp(2\pi\nu_1 t_1)\exp(-R_2^1 t_1)\cdot\exp(2\pi\nu_2 t_2) \times \exp(-R_2^2 t_2) \quad (4)$$

Using a Cartesian sampling and FFT, the obtained spectrum will show no noise or baseline fluctuation far way from the signal of interest. If one turns to radial sampling—or any deterministic scheme—, some oscillations become visible in the spectrum. We and others [13,12] have shown that they are not located in the direct vicinity of the peak and their geometry can be rationalized on the basis of the acquisition pattern. These artifacts can be analysed using the point spread function (PSF or point response function), which is the Fourier transform of a Dirac signal.

Fig. 5 displays the resulting spectrum using the random acquisition scheme depicted in Fig. 1. Even if the time-domain data does not contain noise, the baseline of the spectrum is not flat. These artifacts that look at first glance "random" are related to the acquisition scheme by the PSF. A symmetrical pattern (about a vertical and a horizontal spectrum) is clearly visible. Spectrum in Fig. 5a is obtained using the Voronoi cell weighting described earlier and in Fig. 5b using the same weight for all data points as proposed by Kazimierczuk et al. [11]. As noted earlier for deterministic sampling patterns [12,13], the artifacts remain clustered far away from the peak, when the integration area is taken into account.

The effect of data weighting on artifacts cannot be investigated on a single spectrum due to the random nature of our scheme. One specific sampling scheme may be superior to another with respect to artifacts but not to resolution. Let us consider a single oscillating signal in the time domain and different random acquisition schemes with the same Gaussian distribution. These FIDs are processed with and without Voronoi cell weighting; then, for each point of the spectrum, we evaluate the mean and the standard deviation of the amplitude distribution, displayed in Figs. 5c–h. Note that for the two weighting schemes, the mean of the artifacts is smaller than their standard deviation. Two different trends are observed on the mean value: when no weighting is used, the signal is broader (inset d) and a larger DC offset is observed over the entire spectrum range (inset f). The first effect originates from the smaller contribution of the last data points (a truncated Gaussian distribution was used and the absence of weighting does not compensate for it) and the latter from the incorrect scaling of the first time-domain data point. The flat profile of the standard deviation in inset h confirms that the artifacts are spread uniformly over the spectrum when no weighting is used in contrast with the profile in inset g. Let us compare the standard deviation of the artifacts (insets g and h) with the amplitude of the parent signal (50 and 30, insets c and d): far away from the peak, their *relative* amplitudes (6.5:50 versus 3.3:30) are roughly similar

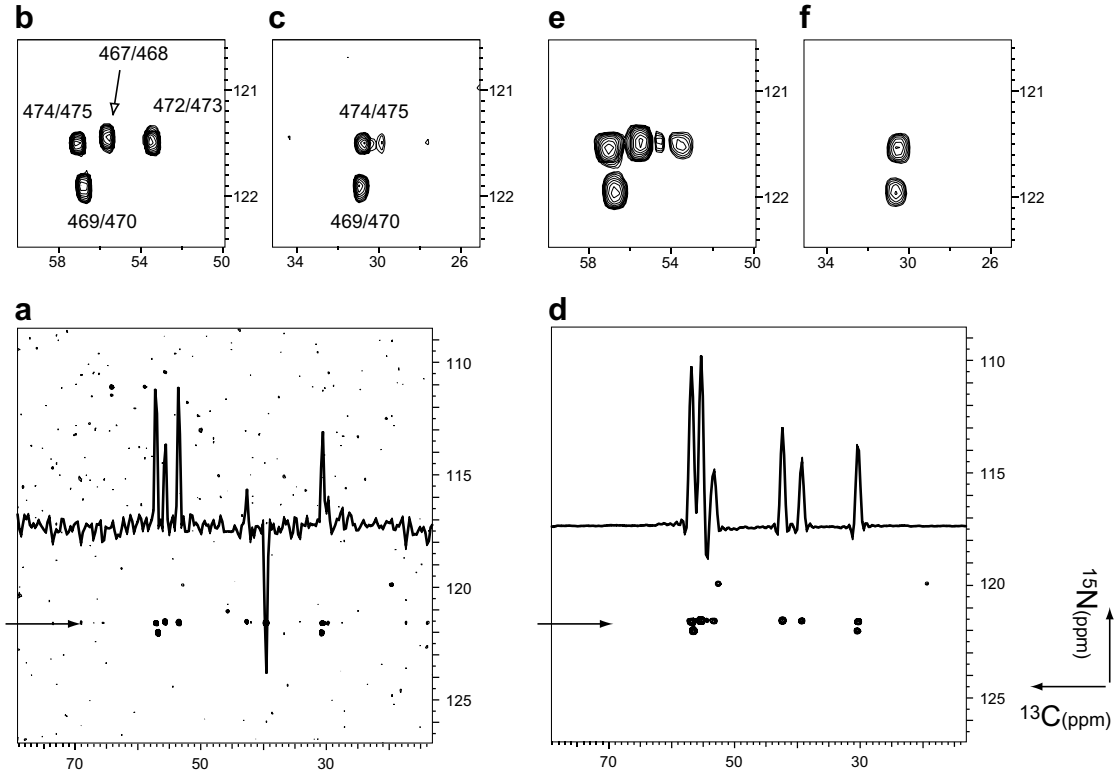

Fig. 4. CBCA(CO)NH experiments recorded on N_{TAIL}[443–501]. Contour plots of two ^{13}C–^{15}N planes were taken at $^{1}H = 8.43$ ppm. Inset (a) has been acquired with random sampling and (d) with Cartesian sampling. Insets (b), (c), (e) and (f) show an expansion of the C^{α}-N correlations with the corresponding assignment (C^{α} (i) with NH($i+1$)). The 1D slice is taken at the location of the arrow. A negative C^{β}–N correlation is observed for residues without C^{γ} (Asn 472) and a positive for residues with one C^{γ} (Glu 474 and Leu 467). N_{TAIL}[443–501] sample at 0.88 mM concentration was prepared in phosphate buffer (50 mM) (pH 6.0) with 500 mM NaCl, 0.02% NaN_3, protease inhibitor cocktail and 10% D_2O.

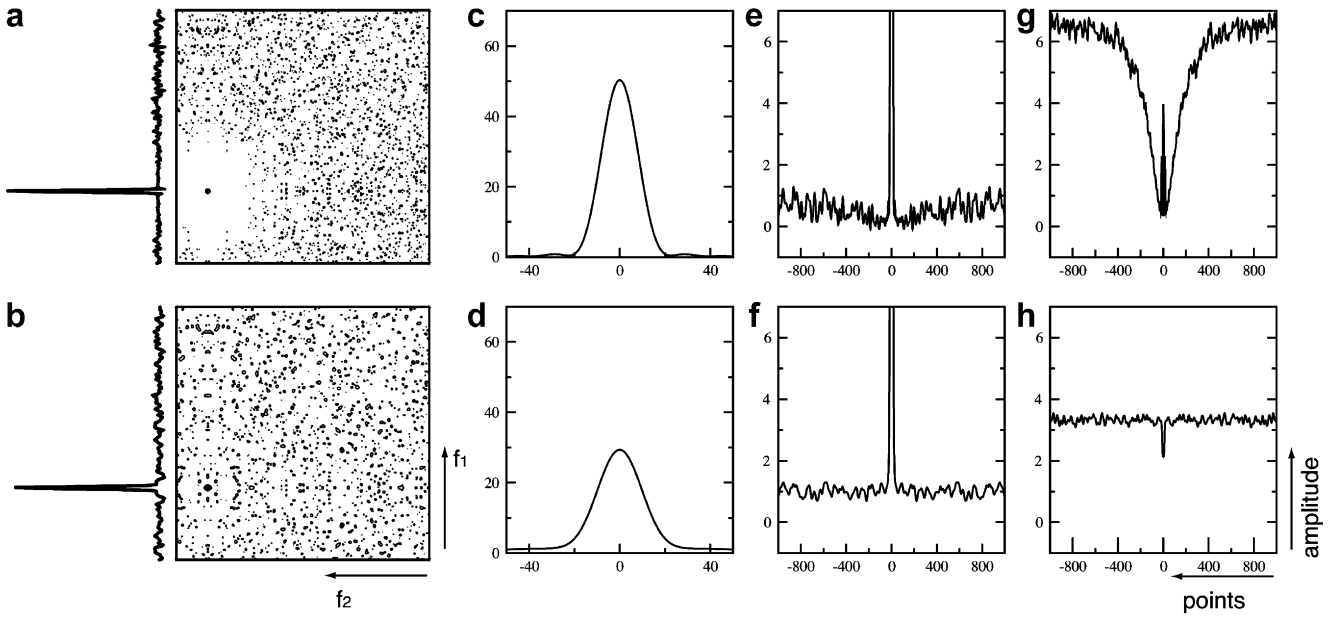

Fig. 5. Effect of data weighting on Fourier transformed spectra. Simulated data *without experimental noise* were used to analyze the artifacts associated with random acquisition. Insets a and b illustrate the artifacts distribution in a 2D spectrum with a single signal. Random samples were selected according to the scheme depicted in Fig. 1 and Fourier transformed. Spectrum A uses the Voronoi cell weighting and spectrum b an uniform weighting (cf. Fig. 4 of Kazimierczuk et al. [15]). Normalized 1D slices are shown on the left. To evaluate the differences between the two processing schemes, we have generated 1000 1D spectra: the signal remains unchanged but the sampling scheme is varied in a random manner. A Gaussian distribution (truncated at 50%) of 48 points was used where 128 points would be necessary to span the FID at the Nyquist frequency. This corresponds to an average undersampling of 40%. The statistical analysis of these spectra is reported in insets c/d and e/f for the mean (note the different scaling of the panels) and in insets g/h for the standard deviation. Whereas the artifacts are uniformly spread when uniform weighting is used, they are shifted far away the main peak with Voronoi cell weighting.

for both methods, but the Voronoi cell weight reduces drastically the artifacts in the direct vicinity of the signal. The price to pay for shifting the artifacts far away from the peak is thus limited. The central narrow peak visible in Fig. 5g is related to slight fluctuations of the main peak line-width and amplitude, depending which data points were sampled in the time-domain signal. The lack of artifacts in the direct vicinity of the signal when Voronoi cell weighting is used leads to the conclusion that random acquisition combined with weighted FT can thus be used for quantitative NMR. In spectra containing several peaks (usually of similar amplitude in heteronuclear NMR), the artifacts associated with one signal will obviously interfere with other peaks, but the advantage of weighting remains anyway.

Using only two high-resolution CBCA(CO)NH and CBCANH experiments, the resonances of the HN, N, C^{α} and C^{β} nuclei of N_{TAIL}[443–501] were completely assigned. For folded proteins, ambiguities during the assignment process are resolved by adding the information gathered from a pair of HNCO and HN(CA)CO experiments, but it is unlikely that this would help for unstructured molecules, as the chemical shift dispersion for ^{13}CO is smaller than for other aliphatic ^{13}C. Based on these assignments, structural features of the protein domain can thus be analyzed using the chemical shift index (CSI) [23], which reports on protein secondary structures. The C^{α} carbons experience an upfield shift in helices and a downfield shift in β-strand while the opposite trend is observed for C^{β}. The secondary chemical shift for N_{TAIL}[443–501] are reported in Fig. 6 for HN, N, C^{α} and C^{β} nucleus. These data confirm the lack of secondary structure elements in this protein although a closer inspection reveals a group of upfield shifted C^{α} between residues 477 and 491. Weak helix propensity was predicted based on sequence analysis for residues 477–495 and our NMR data support this trend. For comparison purposes, helical structures in N_{TAIL} interaction partner (protein PX [24]) are associated with stretches of C^{α} shifts deviating from random coil value by more than 2 ppm. Preliminary NMR results (Houben et al. to be published) shows that this region with weak helical propensity represent the PX interaction site.

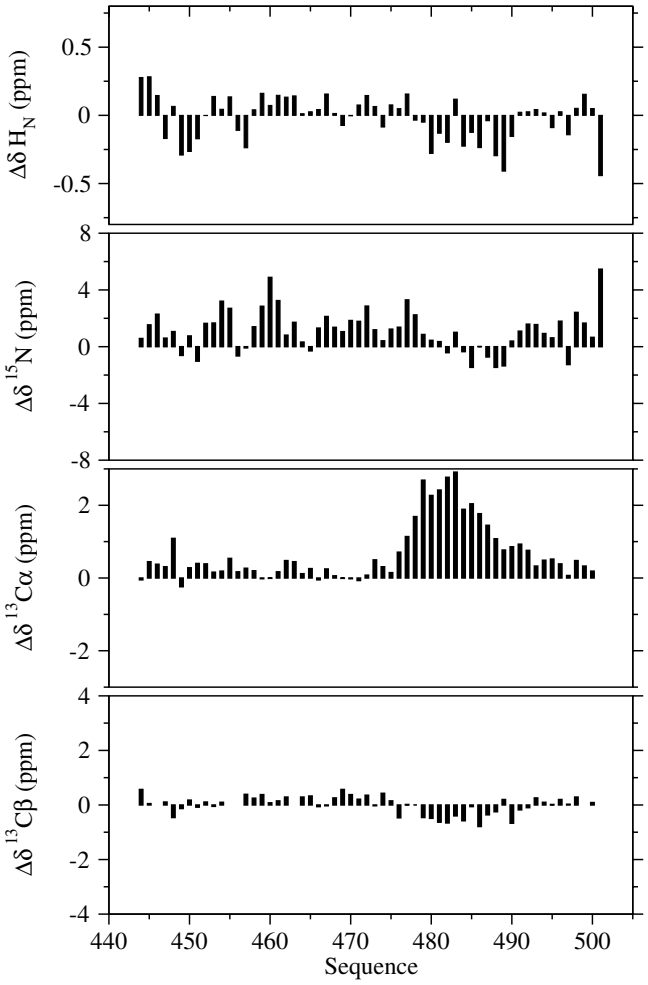

Fig. 6. Secondary chemical shifts observed for protein domain N_{TAIL}[443–501]. These shifts were obtained by substracting random chemical shifts published by Wishart and Sykes [23] from the observed shifts. No secondary structure elements are detected, except a weak trend visible only on the C^{α} between residues 477 and 491.

5. Conclusions

Assigning resonances of unfolded proteins is a prerequisite for their study by NMR. The lack of secondary structure elements leads to numerous overlap in the NMR spectra and consequently resonance assignment becomes a tedious task. However, the flexibility of these proteins usually leads to narrow signals, a feature that can be exploited to resolve the overlap ambiguities. In this communication, we have shown that non-linear sampling combined with Fourier transform can be easily implemented for this purpose. The ''random'' appearance of the artifacts does not introduce any bias in the frequency and amplitude of the signals and makes any misinterpretation unlikely, this in contrast to artifacts with well-defined geometry. With the advance of cryogenic probes, one can compromise the apparent decrease of the signal-to-noise ratio for a better digital resolution in the indirect dimension.

Whereas FT is the most commonly used processing method in the field of NMR, several more sophisticated methods have been proposed such as maximum entropy reconstruction (MEM [25]), filter diagonalisation method (FDM [26]) or multi-way decomposition ([27]...). These methods share a number of common features that distinguish them from FT: they are more computationally demanding and aim at the final result through an iterative process, that has to be conducted with great care by the spectroscopist. Some of them (FDM [26] and multi-way decomposition [27]) deal with all frequency dimensions simultaneously, in contrast with FT where the individual spin frequencies are treated independently. In the case of maximum entropy, the iterative computation searches for an optimal balance between the data and the signal entro-

py, a process that is highly non-linear. In the hands of experienced users, these algorithms perform very well, but artifacts due to the processing (and not the sampling as discussed earlier) may occur. The main advantage of the method proposed here (random sampling combined with Voronoi cell weighting and FT) relies in its linearity and the lack of any adjustable parameter. On the other hand, the point spread function (PSF) can be used as a tool to optimize the sampling scheme for specific purposes. Finally, it can be implemented with minimal changes in standard processing schemes and the longer processing time required for random sampling (as compared to standard FFT) will clearly not be an issue in the near future with the increasing computing power.

References

[1] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci. 11 (2002) 739–756.

[2] A.L. Fink, Natively unfolded proteins, Curr. Opin. Struct. Biol. 15 (2005) 35–41.

[3] R. Bussell Jr., D. Eliezer, Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein, J. Biol. Chem. 276 (2001) 45603–45996.

[4] P. Garcia, L. Serrano, D. Durand, M. Rico, M. Bruix, NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation, Protein Sci. 10 (2001) 1100–1112.

[5] B. Brutscher, R. Brüschweiler, R.R. Ernst, Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ^{1}H, ^{13}C, and ^{15}N nuclear magnetic resonance spectroscopy, Biochemistry 36 (1997) 13043–13053.

[6] D.S. Wishart, B.D. Sykes, The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data, J. Biomol. NMR 4 (1994) 171–180.

[7] L.E. Kay, K.H. Gardner, Solution NMR spectroscopy beyond 25 kDa, Curr. Opin. Struct. Biol. 7 (1997) 722–731.

[8] T. Szyperski, G. Wider, J.H. Bushweller, K. Wüthrich, 3D ^{13}C–^{15}N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C–^{15}N-double-labeled proteins, J. Biomol. NMR 3 (1993) 127–132.

[9] J.P. Simorre, B. Brutscher, M.S. Caffrey, D. Marion, Assignment of NMR spectra of proteins using triple-resonance two-dimensional experiments, J. Biomol. NMR 4 (1994) 325–334.

[10] E. Kupče, R. Freeman, Projection-reconstruction of three-dimensional NMR spectra, J. Am. Chem. Soc. 125 (2003) 13958–13959.

[11] K. Kazimierczuk, W. Kozminski, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, J. Magn. Reson. 179 (2006) 323–328.

[12] B.E. Coggins, P. Zhou, Polar Fourier transforms of radially sampled NMR data, J. Magn. Reson. 182 (2006) 84–95.

[13] D. Marion, Processing of ND NMR spectra sampled in polar coordinates: a simple Fourier transform instead of a reconstruction, J. Biomol. NMR 36 (2006) 45–54.

[14] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19 (1965) 297–301.

[15] K. Kazimierczuk, A. Zawadzka, W. Kozminski, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, J. Biomol. NMR 36 (2006) 157–168.

[16] G. Voronoi, Nouvelles applications des paramètres continus à la théorie des formes quadratiques, Journal fur die Reine und Angewandte Mathematik 133 (1908) 97–178.

[17] C.B. Barber, D.P. Dobkin, H.T. Huhdanpaa, The Quickhull algorithm for convex hulls, ACM Trans. Math. Softw. 22 (1996) 469–483.

[18] C.M. Slupsky, R.F. Boyko, V.K. Booth, B.D. Sykes, Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments, J. Biomol. NMR 27 (2003) 313–321.

[19] S. Grzesiek, A. Bax, Correlating backbone amide and sidechain resonances in larger proteins by multiple relayed triple resonance NMR, J. Am. Chem. Soc. 114 (1992) 6291–6293.

[20] S. Grzesiek, A. Bax, An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, J. Magn. Reson. 99 (1992) 201–207.

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

[22] S. Grzesiek, J. Anglister, A. Bax, Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$ enriched proteins by isotropic mixing of ^{13}C magnetization, J. Magn. Reson. B 101 (1993) 114–119.

[23] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J. Biomol. NMR 5 (1995) 67–81.

[24] L. Blanchard, N. Tarbouriech, M. Blackledge, P. Timmins, W.P. Burmeister, R.W. Ruigrok, D. Marion, Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution, Virology 319 (2004) 201–211.

[25] M. Mobli, A.S. Stern, J.C. Hoch, Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy, J. Magn. Reson. 182 (2006) 96–105.

[26] J. Chen, D. Nietlispach, A.J. Shaka, V.A. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, J. Magn. Reson. 169 (2004) 215–224.

[27] V.Y. Orekhov, I. Ibraghimov, M. Billeter, Optimizing resolution in multidimensional NMR by three-way decomposition, J. Biomol. NMR 27 (2003) 165–173;
D. Malmodin, M. Billeter, Multiway decomposition of NMR spectra with coupled evolution periods, J. Am. Chem. Soc. 127 (2005) 13486–13487.