Journal of Magnetic Resonance 206 (2010) 81–87

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Combining methods for speeding up multi-dimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins ☆

Dominique Marion *

Laboratoire de RMN, Institut de Biologie Structurale Jean-Pierre Ebel, UMR5075 CNRS–CEA–UJF, 41, Rue Jules Horowitz, 38027 Grenoble Cedex, France

ARTICLE INFO

Article history:
Received 9 April 2010
Revised 31 May 2010
Available online 12 June 2010

Keywords:
Fast acquisition
BEST-sequence
Non-uniform sampling
Intrinsically disordered proteins
Maximum entropy reconstruction

ABSTRACT

Resonance assignment of intrinsically disordered proteins is made difficult by the extensive spectral overlaps. High-resolution 3D and 4D spectra are thus essential for this purpose. We have adapted the series of 3D BEST-experiments proposed by Lescop et al. [E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, J. Magn. Reson. 187 (2007) 163–169] to the case of unfolded proteins. Longer acquisitions in the indirect dimensions are obtained by implementing semi-constant time evolution and sparse sampling. Using maximum entropy reconstruction for the indirect dimensions, the artifact intensity due to sparse sampling can be reduced to a level similar to the other sources of noise. The reduction of the sampled increments and the shorter duration of individual transients makes it possible to record a 4D experiment with reasonable resolution in less than 60 h.

1. Introduction

Multidimensional (nD) NMR provides detailed information on biological macromolecules at the atomic resolution. However, it suffers from two drawbacks, low sensitivity and long measurement times. At first glance, these seem to be inextricably linked. Until recently, the length of a nD experiment was determined by its intrinsic (low) sensitivity and many 3D experiments were acquired overall several days only to achieve a reasonable signal-to-noise ratio. For this purpose, one would usually repeat a given pulse sequence without any change in the timing, phase setting and pulsed field gradient.

With the advent of high magnetic field and cryoprobe technology, the sensitivity has drastically improved over the recent years. Consequently, it is now possible to overcome the low sensitivity of a single scan. The design of nD experiments can be optimized with different goals in mind: the acquisition of a highly resolved spectrum with higher dimensionality within the minimal amount of time. Three different tracks have been followed by several research groups: reducing the length of an individual scan, collecting more information per scan (by never repeating exactly the same scan) and sampling the indirect dimension in a clever manner.

The first approach is based on the finding that the largest amount of time spent per individual scan is devoted to the relaxation delay where the spins are allowed to recover towards their thermodynamic equilibrium. It attempts to reduce this delay by selective excitation, which leaves other spins unaffected: this is the underlying principle of fast pulsing methods initially proposed by Pervushin et al. [1] and implemented in the SOFAST [2] and BEST-experiments (Band-selective excitation short transient) [3]. Single scan acquisition techniques rely on spatial encoding [4,5] as in MRI: within a single experiment, the spins located at various coordinates in the sample experience different evolution and the information usually gathered from different increments is here collected simultaneously. Finally, alternate sampling techniques aim at replacing the standard Cartesian grid used in indirect dimensions by sparse schemes [6–10]. In most cases, the inherent limitations of the processing algorithm dictate how the points are sampled [11,12]. Non-uniform sampling methods (NUS) can be combined with several processing techniques such as generalized FT [6,7], maximum entropy reconstruction [8,9] or three-way decomposition [10]. These sampling techniques inevitably lead to artifacts in the processed spectra that need to be minimized at both the acquisition and processing level. Rather than actually increasing the number of dimensions sampled independently, APSY [13] uses 2D projection of 6D experiments combined with an efficient algorithm (GAPRO) to identify all projected peaks.

Based on different principles, these three groups of techniques (imaging, fast pulsing and alternate sampling) can be combined at will. For example, single scan methods have been recently merged with fast pulsing techniques. In this communication, we explore how to combine fast pulsing techniques with non-linear sampling. While fast pulsing techniques were initially developed

☆ Part of this work was carried out during a stay at UMR CIRAD–INRA 1309 "Contrôle des maladies animales exotiques et émergentes" 97170 Petit Bourg, Guadeloupe (F.W.I.)

* Fax: +33 4 38 78 54 94.
E-mail address: Dominique.Marion@ibs.fr

1090-7807/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2010.06.007

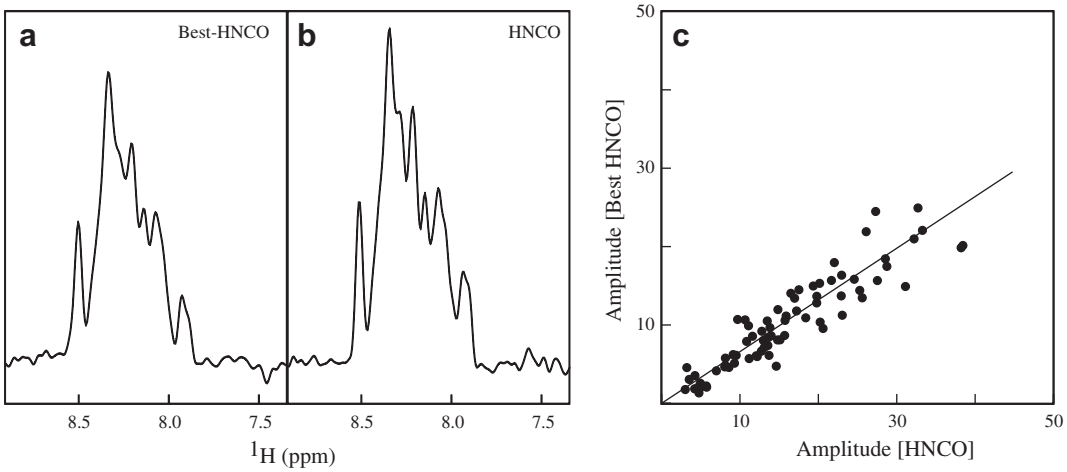

Fig. 1. Comparison of the absolute sensitivity of the BEST-HNCO experiment with a standard HNCO. A 2D experiment corresponding to a ^{1}H–^{13}CO plane was acquired (sw_1 = 1200 Hz, 128 × 2 increments and 16 scans). The interscan delay (d1) was set to 1.0 s for the HNCO sequence and 0.2 s for the BEST-HNCO. The overall experimental time for this 2D experiment decreases from 90 mn for the HNCO to 26 mn for the BEST-version. The spectra were Fourier transformed using the Rowland NMR toolkit [8]. Insets (a) and (b) compare the normalized amplitude of the first FID for the two schemes on protein Hbxip while inset (c) compares the cross-peak intensities in the two ^{1}H–^{13}CO planes. Starting from the peak picking on the HNCO experiment, the intensities of the cross-peaks in both spectra were optimized using nlinLS [30] assuming a lorentzian line-shape in both dimensions.

for globular folded proteins, we are extending their application to unfolded proteins which have recently attracted attention.

More than 30% of the proteins in the human proteome [14] are either fully unfolded or contain unfolded regions of significant length and do not fold into stable three-dimensional structures. Despite their lack of structure, these proteins play important roles in many physiological and pathological processes. In the case of folded proteins, the shortening of the interscan delay is made possible by the fast selective relaxation rate of the ^{1}H [1] when their neighbors remain close to the equilibrium value. This property is clearly associated with the local mobility of the polypeptidic chain which is likely to be quite different in unfolded proteins.

In this communication, we demonstrate that BEST-experiments can be applied to intrinsically disordered proteins. As these macromolecules lack persistent secondary and tertiary structures, their NMR spectra exhibit two features: in the absence of conformational shifts, the chemical shift dispersion of a given type of signals is moderate but the line-width of each of them is relatively narrow due to the fast conformational averaging. Therefore, their resonance assignment as well any further study requires 3D spectra with enhanced digital resolution. In contrast to single scan experiments [5], fast pulsing methods such as BEST-experiments are fully compatible with non-linear sampling methods. A first combination of these two concepts has been reported by Maciejewski et al. [15] while studying aliasing in non-uniform sampling. Here, we will show that well-resolved 3D spectra can be obtained overnight, and 4D experiments in two days by combining fast pulsing and non-uniform sampling. When processed with maximum entropy reconstruction, the sampling artifacts along the sparse dimensions can be minimized to a reasonable level. As a model protein, we have used a 99 residue (91 + a poly-His tag) protein called Hbxip [16], known to interact with protein Hbx, a non-structural viral proteins of Hepatitis b. Preliminary results (Pierre Gans, personal communication) show that this protein is mainly unstructured in solution.

2. Results and discussion

2.1. Experimental sensitivity

Resonance assignment for intrinsically disordered proteins presents a different challenge than for folded ones. While the constraint on sensitivity is generally relaxed, digital resolution becomes even more critical for disordered proteins to the discrimination of nearly degenerate signals. For a fixed amount of spectrometer time, higher digital resolution in indirect dimensions can be achieved only at the expense of reducing the duration of each individual scan. BEST-experiments [3] seem *a priori* well suited for shortening this duration because they allow interscan delay as short as 0.2 s. However, the laws of spin physics that make these pulse sequences effective may not be fully applicable to the case of intrinsically disordered proteins. BEST-experiments rely on the assumption that the rate of recovery of the H^N protons is much faster when other backbone and side-chain protons are not excited. This fast recovery can be explained by compensation of auto- and cross-relaxation mechanisms with other protons. NMR relaxation properties of intrinsically disordered proteins [17] are clearly different from that of globular proteins and it can be anticipated that this compensation might not be as efficient. On the other hand, amide protons in unfolded proteins are solvent exposed and thus interact more closely with water protons by chemical exchange. Note that these two effects are exploited in the Het-SOFAST experiment [18] to assess the structural compactness of polypeptide chains.

As a detailed investigation of these effects is beyond the scope of the present study, the sensitivity of the BEST-sequence has been *experimentally* compared to that of the standard triple resonance sequence on protein Hbxip [16]. Inspection of the ^{1}H–^{15}N HSQC spectrum of this protein (data not shown) reveals that 76 out of 89 assigned residues have their H^N between 8.0 and 8.5 ppm. This is a clear signature of the absence any persistent secondary structures under the NMR experimental conditions. In Fig. 1a and b, the amplitude of the first increment in a BEST-HNCO is compared with that from a regular HNCO on protein Hbxip. For the HNCO sequence, a delay of 1.0 s were chosen, a value routinely used in our laboratory for most proteins. For the BEST-version, a 0.2 s delay was selected for the most efficient use of spectrometer time. For a quantitative comparison, the cross-peak amplitudes in a ^{1}H–^{13}CO plane of a 2D BEST-HNCO were plotted against the amplitude in the HNCO version (Fig. 1c). The *absolute* intensities in the BEST-version are on average 64% of that in the regular experiment: if the ratio of the experimental time (nearly 3) is taken into account, this means that both sequences have a similar *relative* sensitivity. The scattering of the intensity correlation in Fig. 1c demonstrates that the relaxation processes and possibly the exchange with bulk

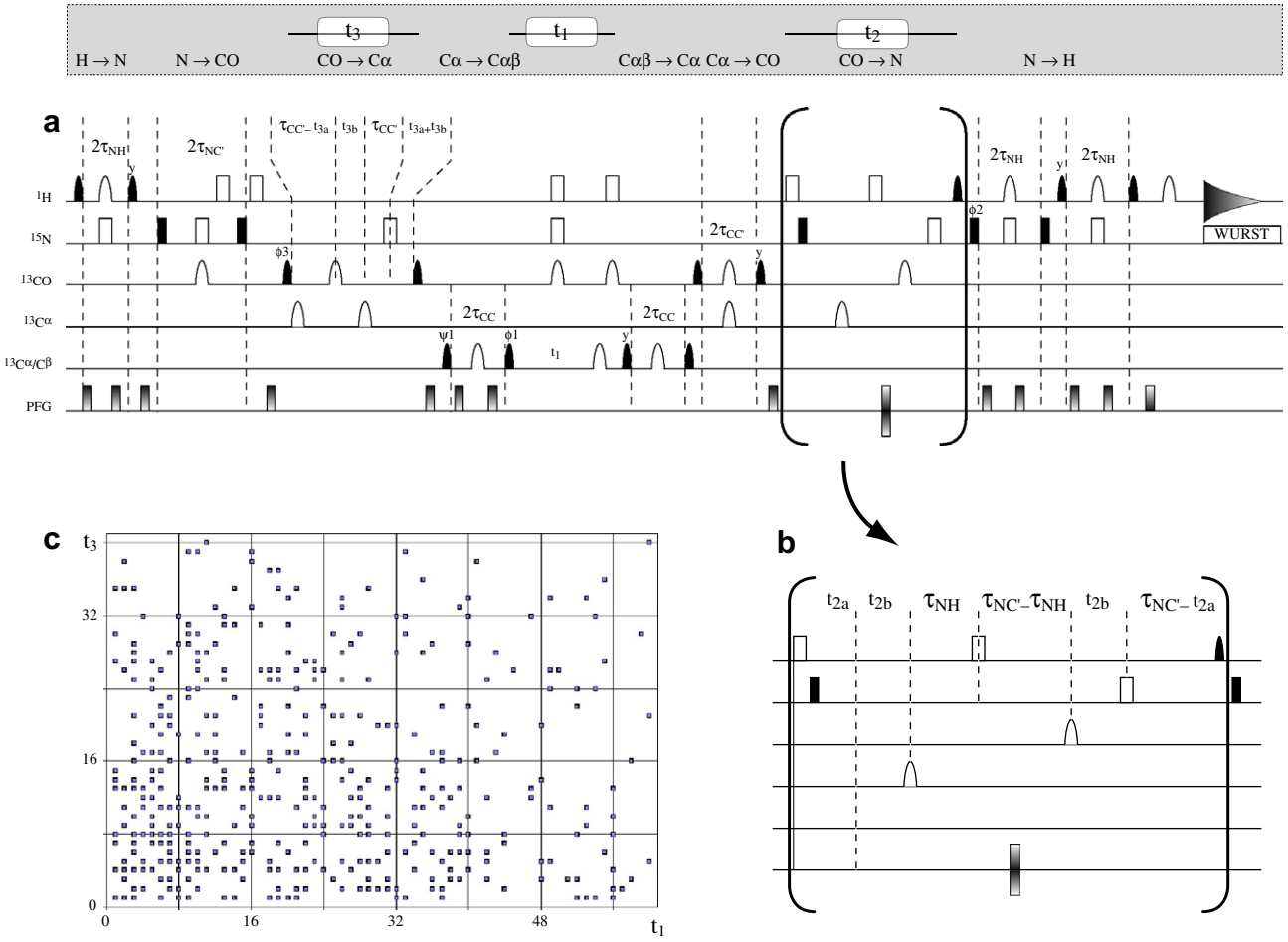

Fig. 2. Pulse sequence used to record a 4D BEST-HNCO[CACB] on intrinsically unfolded proteins (inset a). Based on the 3D BEST-HN(CO)[CACB] described earlier [3], it correlates the following nuclei: H^N(i), N(i), CO(i-1) and simultaneously C^α and C^β(i-1). The corresponding 3D version BEST-HN(CO)[CACB] is obtained by setting $t_{3a} = t_{3b} = 0$. Both ^{15}N and ^{13}CO frequencies are labeled during a transfer delay in a semi-constant manner (see below). Filled and open pulse symbols indicate 90° and 180° rf pulses. All selective ^{1}H pulses are centered at 8.2 ppm, covering a bandwidth of 4.0 ppm: PC9 or E-BURP2 shapes are used for ^{1}H 90° and REBURP for the 180°. CO pulses have sinx/x shape whereas C^α and C^α/C^β pulses are rectangular with zero excitation at the CO frequency. $\tau_{NH} = 2.4$ ms $\approx 1/4J_{NH}$, $\tau_{NC'} = 14.5$ ms $\approx 1/2J'_{NC}$, $\tau_{CC'} = 4.5$ ms $\approx 1/2J'_{CC}$, $\tau_{CC} = 3.5$ ms $\approx 1/4JC\alpha C\beta$ (sampling of both C^α and C^β resonances) or $\tau_{CC} = 7.1$ ms $\approx 1/2JC\alpha C\beta$ (sampling of C^β resonances). Note that these actual delays encompass the duration of the 180° refocusing/decoupling pulses placed at their center. Quadrature detection in t_1 and t_3 is obtained by time-proportional phase incrementation of ψ_1 (and φ_1) and φ_3, respectively according to TPPI-states. For quadrature detection in t_2, echo–antiecho data are recorded by inverting the sign of gradient in the center of the bracket and phase φ_2. Pulsed field gradients are applied along the z-axis (PFG) with durations of 200 μs–2 ms and field strengths ranging from 5–40 G/cm. See Lescop et al. [3] for further details. The Varian sequence and processing scripts are available from the author upon e-mail request. Inset b show the detailed timing of semi-constant evolution (t_2) for ^{15}N during the delay $2 \times \tau_{NC'}$. During this period the contributions of the J_{NH} and $J_{NC}\alpha$ are refocused while the $J_{NC}\alpha$ is active. Let us define t_2^{max} as the largest increment in the (linear or non-linear) sampling schedule and Δt_2 the dwell time associated with the ^{15}N spectral width. t_{2a} and t_{2b} are incremented in parallel according to: $t_{2a} = k.(t_2^{max}/\tau_{NC'})$. Δt_2 and $t_{3a} = k.[1.0 - (t_2^{max}/\tau_{NC'})]$. Δt_2 with $k \in \mathbb{N}$. If a NUS schedule is used for this dimension, the largest value t_2^{max} cannot be computed from the number of points but rather should be estimated by scanning the schedule. A similar computation is performed for the t_3 delay in the case of a 4D experiment. Inset c displays the 2D sampling pattern used for the 4D experiments. We have arbitrarily chosen to sample the two ^{13}C dimensions in a non-linear manner (sw_3 [^{13}C′] = 1200 Hz and sw_1 [$^{13}C^\alpha/C^\beta$] = 11,000 Hz). The schedule was generated using the sampling generator at the Univ. of Connecticut (http://sbtools.uchc.edu/nmr/sample_scheduler/): 400 points out of 2400 (40 × 60) possible points were selected. These pairs of integers are entered as arrays in the Varian vnmrj program and used inside the pulse sequence to evaluate the delays. The ^{15}N dimension was sampled linearly (36 × 2 data points) but the spectral width was reduced to 17 ppm leading to the aliasing of some resonances. The evaluation of the delays for the semi-constant time is achieved inside the *c*-pulse program according to the formula given above.

water interfere in a complex manner with the fast pulsing approach when applied to unfolded proteins. These features are currently investigated in detail in our laboratory (Bernhard Brutscher, personal communication).

2.2. Modification of the BEST pulse sequences

The reduction of the experimental time brought by the fast pulsing can be used to boost the resolution in unfolded proteins: because of their favorable relaxation properties, the indirect dimensions in a 3D experiment can be sampled for longer t_1 or t_2 values. In the original design of the BEST-sequence proposed by Lescop et al. [3], the ^{13}C dimension is sampled in a conventional manner but the ^{15}N dimension is sampled in a constant time way. Whereas the resolution in the ^{13}C dimension can be readily increased, the pulse sequence has to be modified to accommodate longer sampling in the ^{15}N dimension. With a ^{15}N spectral width of 1600 Hz, only 46 increments can be measured during the delay $2 \times \tau_{NC'}$ used to transfer the ^{13}C magnetization back to ^{15}N. To circumvent this limitation, one can either use a double constant time evolution [19] or a semi-constant time one [20,21]. The first method, designed specifically for "out-and-back' experiments, makes use of two transfer delays to sample the indirect frequency. For this to function properly, one should be able to restore the coherences at the beginning of the second period as they were generated at the end of the first one. Pulsed field gradient used to purge unwanted coherence might be detrimental. We have chosen to use a semi-constant-time approach [20] as depicted for a BEST-HNCO[CACB]

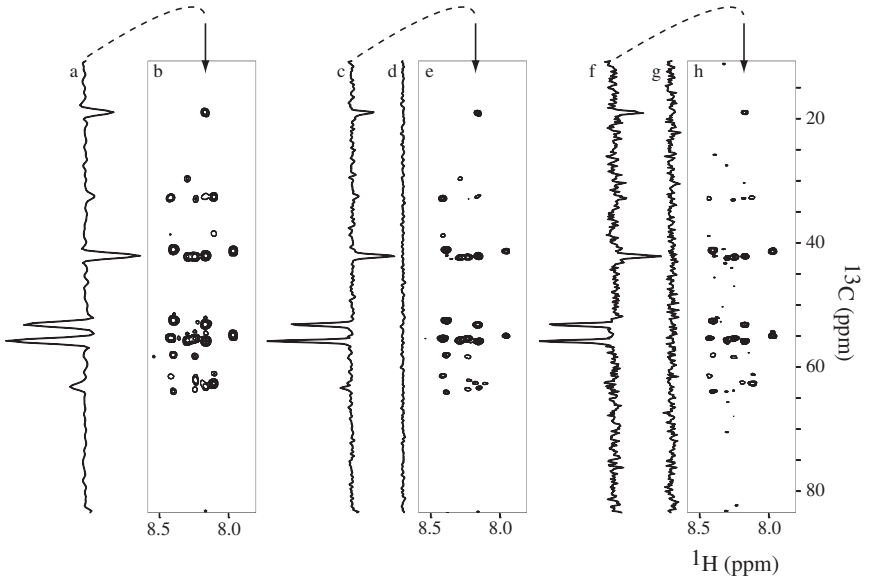

Fig. 3. Influence of the sampling schedule on the resolution and signal-to-noise. BEST-HNCACB spectra were acquired with various schedules for the ^{15}N–^{13}C dimensions with a constant number (4000) of (t_1, t_2) hypercomplex points. A ^{13}C–^{1}H plane at ^{15}N = 123.9 ppm is displayed as well as a 1D slice taken at the arrow position. In inset (b), a linear schedule of 50 × 80 hypercomplex points was used and the data processed with FT in both indirect dimensions. In inset (e), a sparse sampling schedule with ($t_1^{max} \times t_2^{max}$) = 66 × 137 and in inset (h) with ($t_1^{max} \times t_2^{max}$) = 106 × 229. Spectra (e) and (h) were processed automatically using the 2D MaxEnt reconstruction routine ("msa2d" [24]) with constant λ and no deconvolution. The script for processing was generated using the *RNMRTK script generator* at http://sbtools.uchc.edu/nmr/nmr_toolkit/. Optimization was carried out using similar parameters for both spectra – (Def, λ) = (3.95, 1.03) for (e) and (3.24, 1.44) for (h) and converged in less than 18 cycles. The 1D slices (a), (c) and (f) were taken along the ^{13}C dimension at the position of the arrow. Slices (d) and (g) taken at a position which contains no signal to provide an estimate of the noise (thermal + t_1 noise). The ^{13}C line-width of the two cross-peaks around 50 ppm is 150 Hz in (a), 80 Hz in (c) and 70 Hz in (f). Processing of the 3D data set using FT took 30 s and using MaxEnt processing took about 20 mn on a MacBook computer with a 2 GHz Intel Core 2 Duo.

scheme in Fig. 2B. The timing of this building block seems rather convoluted because the ^{15}N evolution is measured during a ^{13}C to ^{15}N delay during which a number of passive couplings (J_{NH}, $J_{NC\alpha}$) are refocused. In our case, t_{2a} and t_{2b} are incremented in parallel in contrast to choices made elsewhere [22], where a non-constant time evolution is included only for longer delays. These alternate approaches should affect the signal line-shapes marginally.

2.3. Digital resolution and artefacts in the indirect dimensions

When longer time increments for indirect dimensions are sampled, resolution is expected to improve, but at the expenses of signal-to-noise ratio. We have compared three BEST-HNCACB experiments recorded on Hbxip for the same amount of experimental time (16 h, 4000 hypercomplex (t_1, t_2) increments). As a reference experiment, a linear sampling scheme 50 × 80 was acquired and Fourier transformed (Fig. 3a). Two experiments with a sparse sampling schemes (66 × 137 and 106 × 229) were collected. Undersampling of the first scheme is moderate (44%) and the other more severe (16%).

These data sets were processed by maximum entropy (MaxEnt) reconstruction using the Rowland NMR toolkit[1] implementation. MaxEnt searches for a spectrum **f** that has the highest entropy among several spectra that are consistent with the measured FID. The optimization has to achieve a dual goal, minimizing the difference $C(\mathbf{f})$ between the calculated and measured data and maximizing an entropy term $S(\mathbf{f})$. A linear combination of the two functions $Q(\mathbf{f}) = S(\mathbf{f}) - \lambda C(\mathbf{f})$ is thus minimized. Compatible with phase sensitive NMR data, the signal entropy defined by Hoch et al. [23].

$$S(f) = \sum_{n=0}^{N-1} F_n \log\left(\frac{1}{2}\cdot\left(F_n + \sqrt{4+F_n^2}\right)\right) - \sqrt{4+F_n^2} \text{ with } F_n = \frac{|f_n|}{def} \quad (1)$$

contains a scale factor *def* related to the spectrometer sensitivity. A measure of the noise level gives an estimate of *def* and a suitable value of λ is evaluated on representative rows or planes of the spectrum. Once these parameters are determined (cf Mobli et al. [24] for details), MaxEnt reconstruction becomes a nearly automated procedure. An *a posteriori* validation of these parameters is given by the number of loops required for convergence (typically <50).

Let us compare the ^{13}C–^{1}H planes of the BEST-HNCACB experiment shown in Fig. 3b and e. Spectrum 3b was processed using apodization and FT along t_1 and t_2 and spectrum 3e using MaxEnt reconstruction. When FT is used in combination with zero-filling, a apodization window is applied that slightly broaden the signals. Note that marginal resolution enhancement could be obtained using the linearly acquired data (data used in Fig. 3b) processed either with MaxEnt reconstruction or by signal extrapolation using linear prediction prior to apodization and FT (spectra not shown). The narrower line-width in the ^{13}C dimension in Fig. 3e is primarily the result of the longer acquisition time (t_{max} = 137 × Δt(^{13}C) instead of 80 × Δt(^{13}C)). The weaker intensity of some peaks in spectrum 3e (as compared to 3b) can be rationalized by the higher resolution in the orthogonal ^{15}N dimension: a peak that is slightly off-resonance will not appear in the adjacent planes if the digital resolution is higher.

For longer acquisition in the indirect dimensions (Fig. 3h), no further improvement in resolution is obtained but the signal-to-noise ratio is lower. This deterioration can be ascribed to two effects: a less-favorable relaxation and artifacts due to incomplete sampling. In the $^{13}C^{\alpha}$/$^{13}C^{\beta}$ and ^{15}N dimensions, the semi-constant time acquisition leads to a longer pulse sequence during which relaxation will be more effective. Data points sampled at longer increments contain a higher proportion of noise but the peak integrals remains preserved. It should be remembered that the integral of a spectrum in the frequency domain is given by the amplitude of the first increment at t_1 = 0 in the time-domain. Rovnyak et al. [25] have shown that the number of sampling points (in indirect

[1] http://rnmrtk.uchc.edu/rnmrtk/RNMRTK.html.

dimensions) that lead to the maximum signal-to-noise ratio does not coincide with the number required for the optimal resolution: if R_2 is the transverse relaxation time of a nucleus detected indirectly, the most sensitive spectrum is obtained at $\sim 1.2 \times R_2^{-1}$ and the best resolution at $\sim 3 \times R_2^{-1}$. Although the transverse relaxation rates in protein Hbxip have not been quantified, the sampling schemes used for Fig. 3eand h are somewhere between these two optima.

When NUS is used, another type of artifacts adds to the real noise (either thermal or t_1 noise). Starting from the sampling pattern, one can evaluate the point spread function (PSF) by Fourier transformation (a sampled point is set to 1, a non-sampled to 0). If randomness is introduced into the sampling schedule, then the PSF will resemble noise [26,27] despite its deterministic nature. Each peak in the spectrum is convoluted by the PSF and thus the artifacts are proportional to the main peak. This point is illustrated in Fig. S1 (Supplementary information): using a synthetic data set where only 3% of the complete sampling schedule is retained, the artefact amplitude is more than 10% of the main peak. MaxEnt partially deconvolutes the PSF from the spectrum [27] and thus reduces these artifacts, as do several other methods (but not FT). Fig. S1 shows that the artifacts visible in the FT of the synthetic data can be considerably reduced by MaxEnt processing. By suitable choices of the threshold parameter *def*, the amount of reduction can be tuned to some extent. Other processing techniques such as the three-way-decomposition [10] may also be to also reduce artifacts.

To evaluate the relative contribution of noise and sampling artifacts in experimental data, one can compare the "noise" in two slices: one that contains a parent peak (Fig. 3c) and one where no signal is expected (Fig. 3d). In the case of BEST-HNCACB shown in Fig. 3, the thermal and/or t_1 noise present in any slice is dominant, showing that MaxEnt has successfully removed a large fraction of sampling artifacts. Note that this finding might not be true in all cases, for example in spectra containing intense diagonal signals.

2.4. Could the pulse sequences be simplified?

Our data are intended to be processed with MaxEnt reconstruction, a strategy that offers more flexibility at the acquisition level than standard FT in two ways. On one hand, sampling a first data point at time t_1 or $t_2 = 0$ is no longer required. It is a common practice to sample signal shifted by ½ dwell time when required despite the slight baseline distortion induced by the necessary phase correction to restore an absorptive profile. On the other hand, spectral deconvolution can be carried out during processing, i.e. some unwanted splittings can be removed at the processing level. These two options open new avenue for designing pulse sequences by removing a number of extraneous pulses. In general, each time a new pulse is introduced, a small amount of the coherence is lost because pulse imperfections.

In all BEST-sequences the following building block is encountered several times:

Spin $^{13}C_1$			180°			180°	
Spin $^{13}C_2$	90°	$t_1/2$		$t_1/2$	180°		90°

where $^{13}C_1$ and $^{13}C_2$ represent two kinds of ^{13}C, such as ^{13}CO, $^{13}C^{\alpha}$ or $^{13}C^{\beta}$ (cf. Fig. 2a). Note that all pulses are selective and thus are of finite duration. The frequency of the single quantum coherence associated with spin $^{13}C_2$ is sampled during this t_1 evolution. The first 180° pulse on C_1 aims at decoupling the J_{CC} coupling but introduces a Bloch–Siegert shift that is cancelled by the last 180° on C_1. The 180° on C_2 helps to remove any first-order phase correction owing to the refocusing properties of a 270° pulse (180° + 90°). We can remove these three pulses altogether, while being aware of the outcomes: the J_{CC} will remain active during the t_1 evolution and the finite duration of the 90° pulses will prevent the recording of a point at $t_1 = 0$.

Two 3D BEST-HNCO spectra [3] were acquired, one with the standard sequence and another where these three pulses were omitted. In the absence of the 180° pulse on C_2, it becomes impossible to record a spectrum which strictly corresponds to $t_1 = 0$. If the sampling grid is shifted by a non integral value of the dwell

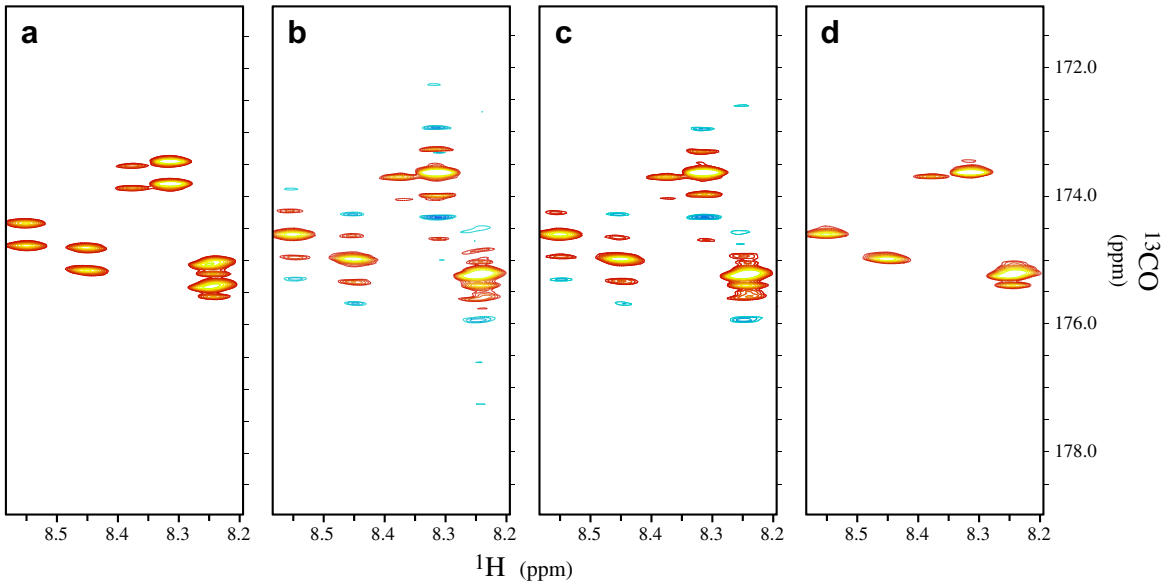

Fig. 4. Simplification of the BEST-HNCO pulse sequence by the removal of 3 ^{13}C selective 180° pulses. Spectra (a), (b) and (c) were acquired with the simplified sequence with the first point set at $t_1 = 1/\Delta W(^{13}C)$. Spectrum (d) was acquired with the complete sequence with the first point set at $t_1 = 0$. Spectra (a) and (d) were processed with MaxEnt reconstruction without deconvolution and spectra (b) and (c) with the deconvolution of an *in-phase J*-coupling of 51 and 53.5 Hz, respectively. The side-bands visible in the ^{13}C dimension in (b) and (c) are due to the incomplete deconvolution of the $J(CO–C^{\alpha})$ coupling. Despite these artifacts, a small increase in the sensitivity of the main peaks (10–20%) is detected in (b) and (c) as compared to the reference spectrum (d).

time, restoring an absorptive line-shape would require a large frequency-dependent phase correction, leading to baseline rolls. In contrast, if the grid is shifted by one increment, the result is a special case of sparse sampling that can be processed using MaxEnt without further phase correction. Fig. 4 shows representative slices of a BEST-HNCO experiment recorded on protein Hbxip using the original sequence (Fig. 4d) and without the 180° pulses (Fig. 4a–c). Comparison of Fig. 4a and d shows that the timing issue is properly handled by MaxEnt reconstruction; in the absence of any frequency phase correction, absorptive signals are obtained with a flat baseline. In contrast, the removal of the $^1J(CO–C^\alpha)$ by deconvolution is incomplete (Fig. 4b and c) for any splitting value used for this purpose (51 and 53.5 Hz, respectively). If vertical slices are taken along the ^{13}C dimension in Fig. 3a, the asymmetry of the doublet is clearly visible indicating that a model assuming purely in-phase splitting is not appropriate. Cross-correlation effects between relaxation mechanisms (^{13}CO CSA and $^{13}C–^{13}C$ dipole–dipole) [28] lead to some asymmetry in the doublet. As this feature may vary from one residue to another, it is not feasible to define a common shape that could be deconvoluted for all spin systems. Such cases would call for a modified MaxEnt algorithm where the function defining the pattern would be optimized during the reconstruction. Despite the occurrence of artifacts, a slight increase (between 10% and 20%) of the signal amplitude is detected between Fig. 4d and b (or 4c). The artifacts preclude the full benefit of this sensitivity increase as the three 180° pulse are intertwined and cannot be suppressed independently.

2.5. Extension to 4D experiments

The benefits of sparse sampling can be extended to higher dimensionality, such as 4D and 5D experiments. In the process of resonance assignment of unfolded proteins, ambiguities due to near degeneracy of chemical shifts can be resolved either by increasing the digital resolution or adding the frequency of an additional nucleus [13]. Inspection of several out-and-back pulse sequences shows that the coherence pathway from the H^N to the side-chains migrates through several intermediate spins. In this respect, one can consider using all intermediate steps to label frequencies. For instance the 3D BEST-HN(CO)CACB experiment involves transfer through CO, and can be converted at no cost into a 4D BEST-HNCOCACB. The pulse sequence details are shown in Fig. 2a and b and the sampling schedule in inset 2c. While all indirect dimensions could be in principle be sampled in a non-linear manner, practical considerations dictate several choices. At the time these data were processed, the Rowland NMR toolkit was only able to handle two-dimensional sampling schemes and carry out 2D MaxEnt reconstruction. Alternate methods such as spectral folding/aliasing could be combined with these techniques, but the spectral dispersion of unfolded proteins defines some boundaries. For instance, extensive aliasing as proposed by Lescop et al. [29] (ASCOM) shows rather limited efficiency for proteins that lack any secondary chemical shifts.

Within 60 h of instrumental time, a 4D BEST-HNCOCACB was acquired on our model protein and processed by means of FT and MaxEnt reconstruction (cf. caption to Fig. 5 for details). In order to yield the same digital resolution, a standard experiment would require roughly 18 times more spectrometer time. This time saving can be split into two components: a factor of three provided by fast pulsing (cf. Fig. 1 and associated discussion) and a factor of six provided by sparse sampling 6 (400 increments instead of 2400). Our shorter approach would have a lower sensitivity because the overall acquisition time determines the global S/N of the experiment. However, this is usually not an issue for intrinsically disordered proteins. An illustrative 2D slice of the 4D spectrum is displayed in Fig. 5 and 1D slices provide visual insight into the resolution

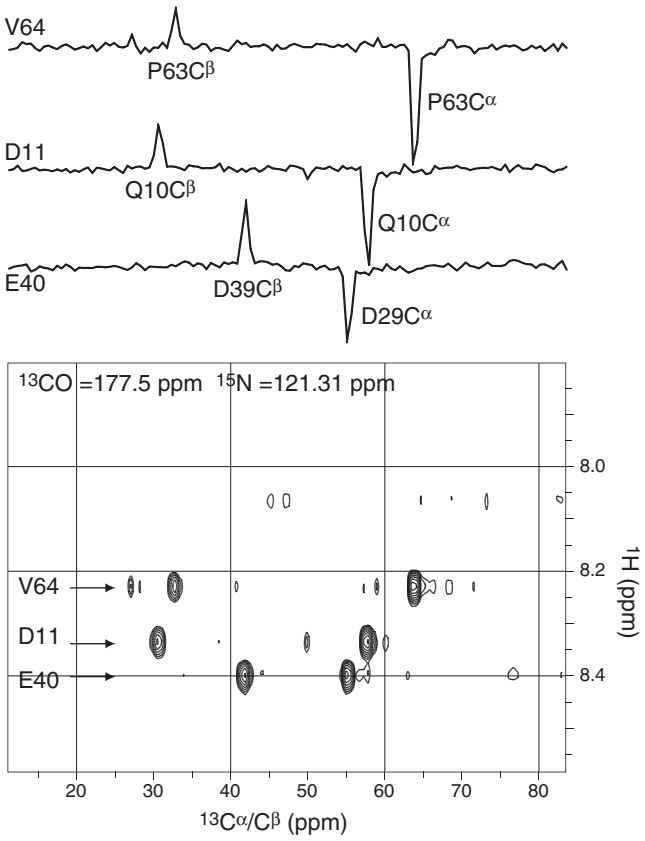

Fig. 5. A $^1H/^{13}C^{\alpha\beta}$ plane of a 4D BEST-HNCOCACB recorded with a Varian 600 MHz spectrometer on protein Hbxip using the pulse sequence depicted in Fig. 2a. The overall experimental time was 60 h. The ^{13}CO (t_1) and the $^{13}C^{\alpha\beta}$ (t_3) dimension were recorded using NUS while the ^{15}N (t_2) was partially folded. The spectral widths are respectively 1200, 1027 and 11,000 Hz for t_1, t_2 and t_3 and the maximum increments are 60, 36 and 40. The 2D sampling schedule for the ^{13}CO (t_1) and $^{13}C^{\alpha\beta}$ (t_3) is shown in fig. 2c. Starting from the time-domain data containing $36 \times 400 \times 602$ hypercomplex data points as $[t_2 \times (t_1, t_3) \times t_4]$, one obtains a 4D frequency-domain matrix of $64 \times 128 \times 128 \times 256$ data points [^{13}CO, ^{15}N, $^{13}C^{\alpha\beta}$, ^1H]. To minimize the shared memory requirements, the conventionally sampled dimensions (t_2, t_4) were first Fourier transformed and the ^1H spectral region devoid of peaks was eliminated. For each (F_2, F_4) point, a 2D time-domain (t_1, t_3) file was then created. The *def* value was evaluated on the basis of the noise and the suitable λ value by processing representative (t_1, t_3) planes. All (t_1, t_3) planes were then processed using "msa2d" with the same *def* and λ value. Subsequently, the (F_1, F_3) planes were then combined along the F_4 acquisition dimension, leading to a collection of [^{13}CO, $^{13}C^{\alpha\beta}$, ^1H] 3D matrices for various ^{15}N frequencies. Whether a given dimension has to be reversed for display (i.e. whether the quadrature phase shift is +90° or −90°) is more conveniently determined on simple 2D planes recorded on the spectrometer using standard linear sampling and FT.

and signal-to-noise achievable. The digital resolution of the processed spectrum is 18, 8 and 85 Hz, respectively along the ^{13}CO, ^{15}N, $^{13}C^{\alpha\beta}$ dimensions. Due to the length of the 4D acquisition, it is not feasible to optimize the resolution in the three indirect dimensions beforehand: The resonance scattering should be taken into account as well as the relaxation properties of the various spins. These properties might be field dependent in the case of CO chemical shift anisotropy. Using the 4D experiment described above and other 3D spectra recorded on Hbxip, nearly complete assignment of the backbone resonances could be obtained.

3. Conclusions

In this paper, we have demonstrated that fast pulsing methods (BEST-sequences) initially designed for folded proteins can be adapted to unstructured macromolecules, despite their markedly different relaxation properties. As compared to standard triple-res-

onance experiments, a relaxation delay between transients as short as 0.2 s can be employed. The BEST-approach is fully compatible with sparse sampling techniques in the indirect dimensions: improved digital resolution can be obtained by sampling longer time increments and processing the data with 2D maximum entropy reconstruction. The pulse sequences have to be slightly modified to semi-constant time evolution in some of the dimensions. The MaxEnt reconstruction reduces the sampling artifacts – in the case of triple-resonance experiments – to a level similar to or below the standard noise (thermal and t_1 noise). In theory, this approach could relax a number of restrictions on pulse sequence design (elimination of some refocusing pulses) although some cross-correlation effects hinder this strategy. Finally, extension to 4D can be easily implemented in several schemes for which high-resolution spectra could be collected within a few days. All these techniques could simplify the resonance assignment of unfolded proteins where severe spectral overlaps frequently lead to ambiguities. Their implementation, though slightly more complex than for regular experiments, is likely to be simplified in the near future by optimization of the interface between the various software program used.

Acknowledgments

The author would like to thank Dr. Pierre Gans for the biophysical studies on Hbxip, Dr Dimitrios Skoufias and Françoise Lacroix for providing the Hbxip sample, Drs Jérome Boisbouvier and Bernhard Brutscher for stimulating discussions of methodological aspects, Dr Laura Lerner for correcting this manuscript and Prof. Jeff Hoch (Univ. of Connecticut) for providing the MaxEnt software used in this studies.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2010.06.007.

References

[1] K. Pervushin, B. Vogeli, A. Eletsky, Longitudinal 1H relaxation optimization in TROSY NMR spectroscopy, J. Am. Chem. Soc. 124 (2002) 12898–12902.
[2] P. Schanda, B. Brutscher, Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds, J. Am. Chem. Soc. 127 (2005) 8014–8015.
[3] E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, J. Magn. Reson. 187 (2007) 163–169.
[4] L. Frydman, A. Lupulescu, T. Scherf, Principles and features of single-scan two-dimensional NMR spectroscopy, J. Am. Chem. Soc. 125 (2003) 9204–9217.
[5] Y. Shrot, L. Frydman, Single-scan NMR spectroscopy at arbitrary dimensions, J. Am. Chem. Soc. 125 (2003) 11385–11396.
[6] D. Marion, Fast acquisition of NMR spectra using Fourier transform of non-equispaced data, J. Biomol. NMR 32 (2005) 141–150.
[7] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, Random sampling of evolution time space and Fourier transform processing, J. Biomol. NMR 36 (2006) 157–168.
[8] 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.
[9] D.P. Frueh, Z.-Y.J. Sun, D.A. Vosburg, C.T. Walsh, J.C. Hoch, G. Wagner, Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins, J. Am. Chem. Soc. 128 (2006) 5757–5763.
[10] V.A. Jaravine, A.V. Zhuravleva, P. Permi, I. Ibraghimov, V. Yu Orekhov, Hyperdimensional NMR spectroscopy with nonlinear sampling, J. Am. Chem. Soc. 130 (2008) 3927–3936.
[11] E. Kupče, R. Freeman, Projection–reconstruction technique for speeding up multidimensional NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 6429–6440.
[12] E. Kupče, R. Freeman, Fast multidimensional NMR: radial sampling of evolution space, J. Magn. Reson. 173 (2005) 317–321.
[13] S. Hiller, R. Joss, G. Wider, Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY), J. Am. Chem. Soc. 130 (2008) 12073–12079.
[14] P. Tompa, The interplay between structure and function in intrinsically unstructured proteins, Febs Lett. 579 (2005) 3346–3354.
[15] M.W. Maciejewski, H.Z. Qui, I. Rujan, M. Mobli, J.C. Hoch, Nonuniform sampling and spectral aliasing, J. Magn. Reson. 199 (2009) 88–93.
[16] M. Melegari, P.P. Scaglioni, J.R. Wands, Cloning and characterization of a novel hepatitis B virus x binding protein that inhibits viral replication, J. Virol. 72 (1998) 1737–1743.
[17] K. Houben, L. Blanchard, M. Blackledge, D. Marion, Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P, Biophys. J. 93 (2007) 2830–2844.
[18] P. Schanda, V. Forge, B. Brutscher, HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains, Magn. Reson. Chem. 44 (2006) S177–S184.
[19] M.A. McCoy, Improved resolution from double constant–time evolution of 3D and 4D triple-resonance experiments, J. Magn. Reson. 130 (1998) 341–345.
[20] S. Grzesiek, A. Bax, Amino-acid type determination in the sequential assignment procedure of uniformly 13C–15N enriched proteins, J. Biomol. NMR 3 (1993) 185–204.
[21] T.M. Logan, E.T. Olejniczak, R.X. Xu, S.W. Fesik, A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments, Biomol. NMR 3 (1993) 225–231.
[22] V. Tugarinov, W.-Y. Choy, E. Kupče, L.E. Kay, Addressing the overlap problem in the quantitative analysis of two dimensional NMR spectra: application to 15N relaxation measurements, J. Biomol. NMR 30 (2004) 347–352.
[23] J.C. Hoch, A.S. Stern, D.L. Donoho, I.M. Johnstone, Maximum-entropy reconstruction of complex (phase-sensitive) spectra, J. Magn. Reson. 86 (1990) 236–246.
[24] M. Mobli, M.W. Maciejewski, M.R. Gryk, J.C. Hoch, Automatic maximum entropy spectral reconstruction in NMR, J. Biomol. NMR 39 (2007) 133–139.
[25] D. Rovnyak, J.C. Hoch, A.S. Stern, G. Wagner, Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy, J. Biomol. NMR 30 (2004) 1–10.
[26] N. Pannetier, K. Houben, L. Blanchard, D. Marion, Optimized 3D-NMR sampling for resonance assignment of partially unfolded proteins, J. Magn. Reson. 186 (2007) 142–149.
[27] J.C. Hoch, M.W. Maciejewski, B. Filipovic, Randomization improves sparse sampling in multidimensional NMR, J. Magn. Reson. 193 (2008) 317–320.
[28] T. Wang, D.S. Weaver, S. Cai, E.R.P. Zuiderweg, Quantifying Lipari–Szabo modelfree parameters from 13CO NMR relaxation experiments, J. Biomol. NMR 36 (2006) 79–102.
[29] E. Lescop, P. Schanda, R. Rasia, B. Brutscher, Automated spectral compression for fast multidimensional NMR and increased time resolution in real-time NMR spectroscopy, J. Am. Chem. Soc. 129 (2007) 2756–2757.
[30] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, Bax: NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR. 6 (1995) 277–293.