Nonuniform sampling and maximum entropy reconstruction applied to the accurate measurement of residual dipolar couplings

Jayne A. Kubat [a], James J. Chou [b], David Rovnyak [a,*]

[a] *Bucknell University, Department of Chemistry, Lewisburg, PA 17837, USA*
[b] *Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, Boston, MA 02115, USA*

Abstract

Residual dipolar couplings (RDC) provide important global restraints for accurate structure determination by NMR. We show that nonuniform sampling in combination with maximum entropy reconstruction (MaxEnt) is a promising strategy for accelerating and potentially enhancing the acquisition of RDC spectra. Using MaxEnt-processed spectra of nonuniformly sampled data sets that are reduced up to one fifth relative to uniform sampling, accurate $^{13}C'$–$^{13}C_\alpha$ RDCs can be obtained that agree with an RMS of 0.67 Hz with those derived from uniformly sampled, Fourier transformed spectra. While confirming that frequency errors in MaxEnt spectra are very slight, an unexpected class of systematic errors was found to occur in the 6th significant figure of $^{13}C'$ chemical shifts of doublets obtained by MaxEnt reconstruction. We show that this error stems from slight line shape perturbations and predict it should be encountered in other nonlinear spectral estimation algorithms. In the case of MaxEnt reconstruction, the error can easily be rendered systematic by straightforward optimization of MaxEnt reconstruction parameters and self-cancels in obtaining RDCs from nonuniformly sampled, MaxEnt reconstructed spectra.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Residual dipolar coupling; Maximum entropy reconstruction; Nonuniform sampling; Calmodulin

1. Introduction

Dipolar couplings measured in marginally aligned biological molecules are global orientation restraints important for molecular structure determination and validation by NMR [1–3]. Due to the orientation degeneracy of the dipolar coupling function, rigorous structure refinement against RDCs usually requires measurement of RDCs between different pairs of nuclei sampling independent orientations. In the case of the protein backbone, these measurements include the large ^{15}N–^1H and $^{13}C_\alpha$–$^1H_\alpha$ couplings, as well as the smaller $^{13}C'$–$^{13}C_\alpha$, ^{15}N–$^{13}C'$, and $^{13}C_\alpha$–$^{13}C_\beta$ couplings. The ^{15}N–^1H, $^{13}C'$–$^{13}C_\alpha$, and ^{15}N–$^{13}C'$ bond vectors lie in the peptide plane whereas $^{13}C_\alpha$–$^1H_\alpha$ and $^{13}C_\alpha$–$^{13}C_\beta$ bonds sample the out-of-plane

orientation. For larger proteins requiring deuteration, $^{13}C_\alpha$–$^1H_\alpha$ couplings cannot be measured and are therefore replaced by the small $^{13}C_\alpha$–$^{13}C_\beta$ couplings. Measurement of small $J/J+D$ splitting has traditionally required very long J-modulated evolution in the indirect dimensions to obtain sufficient resolution and accurate peak positions. More recently, quantitative-J methods have been shown to be highly effective for measuring the small $J/J+D$ couplings for heteronuclear spin pairs such as ^{15}N–$^{13}C'$ and $^{13}C'$–$^{13}C_\alpha$ [4,5]. However, the quantitative-J approach cannot be applied to the homonuclear $^{13}C_\alpha$–$^{13}C_\beta$ coupling and measurement of this coupling usually requires very long C_β-coupled C_α evolution, and therefore long experimental time. Broadly, heteronuclear RDCs have the unique potential to report on proton-poor macromolecular regions, such as metal-binding sites, which are difficult to constrain with other NMR observables. New methods should be developed to accelerate and enhance the

---

* Corresponding author. Fax: +1 570 577 1739.
 E-mail address: drovnyak@bucknell.edu (D. Rovnyak).

measurement of heteronuclear, weak RDCs. In this study, we show that the measurement of RDCs can be accelerated up to fivefold by the combined use of nonuniform sampling, maximum entropy reconstruction, and rigorous peak fitting.

The current environment of many diverse options for alternative data acquisition and processing [6–32] highlights the importance of rationally pairing these methods with experimental problems that best suit them and which most need their potential benefits. RDCs are conventionally obtained by measuring chemical shifts with high precision and accuracy by acquiring data to very long evolution times in the indirect recoupled dimension. Extensive sampling improves digital resolution in the resulting spectra so that peak positions may be measured with high precision; additionally, sampling at long evolution times is strictly necessary for obtaining the maximal available resolution in NMR spectra [33]. Therefore the need to sample indirect evolution periods to very long times represents a sampling requirement intrinsic to the challenge of accelerating RDC spectra. Also, to our knowledge, RDC spectra have not been handled by any type of non-Fourier transform to date, or by any data extrapolation methods, which can introduce frequency errors. For example, RDC data must not be extended in the time domain by linear prediction (LP) to artificially enhance the digital resolution since LP can produce significant frequency errors, especially when the signal to noise ratio is low [6]. So there is also a requirement for very high fidelity processing when handling RDC data sets.

The accordion principle [34] has been applied to the efficient measurement of $^{13}C'$–$^{13}C_\alpha$ RDCs from a series of 2D spectra when the $^1H\{^{15}N\}$ HSQC is well resolved [9]. The strengths of this approach are the ability to scale the couplings and to achieve high digital resolution in the ^{15}N dimension, the ease of data interpretation, and the utilization of the FFT for all processing stages. For larger proteins in which resolution must be improved, the method needs to be scaled to higher dimensions, limiting the time savings, and relaxation signal losses may be serious; however the combination of accordion and FFT strategies addresses the sampling and processing criteria discussed above and should be of utility. We develop here an alternative that is based on the combination of nonuniform sampling and maximum entropy (MaxEnt) reconstruction. In this paper, nonuniform sampling is the practice of acquiring data corresponding to a reduced set of samples on the Nyquist grid in the indirect dimensions of multidimensional NMR spectroscopy using an identical number of transients for each sample. Nonuniform sampling retains samples at very long evolution times in the recoupled dimension of RDC spectra to obtain high resolution while reducing the net experimental time by not acquiring data for all intervening samples (e.g., Fig. 6b and c). Stern et al. demonstrated that maximum entropy reconstruction is highly resistant to frequency errors, and produces NMR spectra with significantly fewer frequency errors

compared to spectra processed with the aid of linear prediction [6]. Further, multi-dimensional NMR spectra obtained by MaxEnt reconstructions of nonuniformly sampled data have been shown to have sufficient fidelity to be suitable for many biomolecular NMR applications in liquids [35,36] and solids, [37] for analysis with automated assignment routines, [38] and for enabling novel pulse sequences for the study of large proteins [39]. Given that MaxEnt reconstruction is a nonlinear spectral estimation algorithm and is not free from frequency errors, the broader question that motivated this work is if the combined used of nonuniform sampling and MaxEnt reconstruction could provide accurate RDCs while yielding meaningful time savings, which we find is possible. Addressing this question led to an investigation into the occurrence of slight frequency errors in MaxEnt spectral estimations that have not been characterized before, and which are a result of slight peak distortions due to nonlinear amplitude bias in MaxEnt reconstructions.

2. Materials and methods

Uniformly ^{15}N- and ^{13}C-labeled recombinant Xenopus calmodulin (CaM) was overexpressed in *Escherichia coli* (strain AR58) and purified to homogeneity as previously described. Three NMR samples were used for the present study, each prepared in 250 μl of 95% H_2O/5% D_2O, pH 7.0, using 280 μl Shigemi microcells. The isotropic sample contains 1 mM CaM, 16 mM $CaCl_2$, and 100 mM KCl. The aligned sample used for structure determination contains 15 mg/ml of the filamentous phage Pf1 (Asla Labs, http://130.237.129.141//asla/asla-phage.htm), 1 mM CaM, 16 mM $CaCl_2$, and 10 mM KCl. A second aligned sample was also made, containing 18 mg/ml Pf1, 0.5 mM CaM, 6 mM $CaCl_2$, and 100 mM KCl. NMR experiments were conducted at 30 °C on a 500 MHz Bruker spectrometer equipped with cryogenic probe. The $^{13}C'$–$^{13}C_\alpha$ couplings were obtained from the standard 3D HNCO recorded with 120 ms of $^{13}C^\alpha$—coupled $^{13}C'$ evolution.

Processing of the full length data sets with the fast Fourier transform (FFT) was performed with NMRPipe [40] and with the Rowland NMR Toolkit [7]. MaxEnt processing of the protein RDC data sets was carried out with the program 'msa2d' implemented in the Rowland NMR Toolkit which performs holistic two-dimensional reconstructions of nonuniformly sampled planes while maintaining constant λ across all planes [41]. Acquisition of 3D-HNCO data consisted of ($180 \times 40 \times 512$) complex points in $^{13}C'$, ^{15}N, ^1H dimensions respectively for all data sets. FFT and MaxEnt data were processed to identical final digital resolutions of ($512 \times 128 \times 1024$) complex points in $^{13}C'$, ^{15}N, ^1H dimensions respectively. For the MaxEnt reconstructions, we utilized single- and dual-processor linux and Apple OSX workstations, with typically 1 GB RAM and 2 GHz or greater processor speed. Whereas MaxEnt reconstructions of typical 3D assignment spectra may require 15 min or less of computation time on these types

of computers, the coupled 3D-HNCO spectra require substantially higher digital resolution and more stringent MaxEnt parameters *def* and *λ*, and therefore require much greater computation times. We processed the spectra in approximately half-spectral-width batches, requiring a total time of about 5–6 h on a dual 2 GHz G5 MacOSX workstation. With the recent advent of dual/quad core processing, faster CPUs, faster memory access, and higher RAM capacities, MaxEnt computation times would decrease substantially on such equipment.

Simulated two-dimensional doublet data was obtained with an in-house C program and processed by either the FFT, by two-dimensional MaxEnt reconstruction ('msa2d'), or by constant-*λ* row-wise reconstruction. One-dimensional cross-sections (a.k.a. slices) of the data were fit to a double Lorentzian function, requiring six independent variables, by nonlinear least squares fitting in either SigmaPlot 9.0 (Systat Software Inc.,CA) or Grace (http://plasma-gate.weizmann.ac.il/Grace/),

$$
S(\omega) = \frac{A}{\pi} \left[\frac{\frac{1}{2}\delta_A}{(\omega - \omega_A)^2 + \left(\frac{1}{2}\delta_A\right)^2} \right]
$$
$$
+ \frac{B}{\pi} \left[\frac{\frac{1}{2}\delta_B}{(\omega - \omega_B)^2 + \left(\frac{1}{2}\delta_B\right)^2} \right] \tag{1}
$$

where A and B are the amplitudes of each component, δ_A and δ_B are the full widths at half maximum of each component, and ω_A and ω_B are the frequencies of each component. Such fits were used to identify peak positions. Results obtained with SigmaPlot and Grace were validated against each other and found to be indistinguishable. Full three-dimensional data sets were also analyzed with the automated peak picking routine implemented in NMR-ViewJ (One Moon Scientific, NJ).

The maximum entropy reconstruction strategy described in full by Hoch and Stern [7] is briefly outlined here. MaxEnt reconstruction is implemented as an inverse algorithm that obtains a frequency-domain spectrum **f** that is consistent with the experimental data **d**. Two constraints are used to obtain **f**. First, a chi-squared statistic comparing the agreement between the experimental data **d** and the inverse discrete Fourier transform of **f**, denoted **d′** (i.e., **d′** = IDFT(**f**)) is composed,

$$
C(\mathbf{f}, \mathbf{d}) = \sum_{i=0}^{M-1} \left| d'_i - d_i \right|^2 \tag{2}
$$

where the sum takes place only over the M nonuniformly acquired sample points. Additional regularization is needed beyond $C(\mathbf{f}, \mathbf{d})$ in order to obtain unique, smooth spectral estimates for **f**. A modified Shannon equation [42,43] is summed over the N points in the frequency domain to yield

$$
S(\mathbf{f}) = \sum_{n=0}^{N-1} \frac{|f_n|}{def} \log \left(\frac{|f_n|/def + \sqrt{4 + |f_n|^2/def^2}}{2} \right)
$$
$$
- \sqrt{4 + |f_n|^2/def^2} \tag{3}
$$

an entropy assigned to **f**, where maximal entropy is taken as the regularization condition. Both constraints can be achieved together by the use of a LaGrange multiplier *λ* such that the problem of finding **f** is to maximize

$$
Q(\mathbf{f}, \mathbf{d}) = S(\mathbf{f}) - \lambda C(\mathbf{f}, \mathbf{d}) \tag{4}
$$

Qualitatively the procedure may be viewed as finding the least signal in the frequency domain that is consistent with the experimental data. In practice, the method is nontrivial [7].

3. Results

The maximum entropy algorithm implemented in the Rowland NMR Toolkit uses a modification to the Shannon entropy functional [42,43] to regularize spectral estimates and has been described previously [7]. We have shown that MaxEnt reconstructions of typical 3D protein backbone assignment spectra [35] are very insensitive to the exact choices for the two parameters *def* and *λ* that are required in the algorithm. The parameter *def* is a scale factor in the entropy functional [see Eq. (3)], while *λ* is a weighting parameter [LaGrange multiplier, see Eq. (4)] that determines the degree of nonlinearity in the resulting spectral estimate. Generally, decreasing *def* and increasing *λ* lead to more stringent conditions for spectral reconstructions. In a preliminary study, we initially carried out MaxEnt processing of the isotropic and aligned coupled 3D-HNCO spectra for the measurement of weak $^{13}C'–^{13}C_\alpha$ RDCs for Xenopus calmodulin (CaM) using 2/3 data reduction by nonuniform sampling and values of *def* and *λ* that we have found to be effective and time-efficient for many types of data sets (values were ca. 1 for each parameter). However, RDCs determined from these data agreed poorly with RDC's obtained from Fourier processed full length data (not shown) when utilizing either automated two-dimensional peak picking (NMRViewJ) or least squares peak fitting of one-dimensional cross-sections of the data. An RDC measurement is traditionally taken as the difference between the doublet separations in isotropic (J coupled only) and partially aligned (J-coupled with a residual dipolar splitting, J+D) protein samples. In a successive analysis of four peak positions even small intrinsic errors can propagate into large uncertainties in the final RDC determination, so analysis of weak $^{13}C'–^{13}C_\alpha$ RDCs (ca. 1–5 Hz) represents one of the most strenuous tests of non-Fourier processing strategies to date. To explore whether optimization of *def* and *λ* would lead to improved RDC measurements, we performed MaxEnt reconstructions over a broad parameter space for *def* and *λ* (Fig. 1, details in caption).
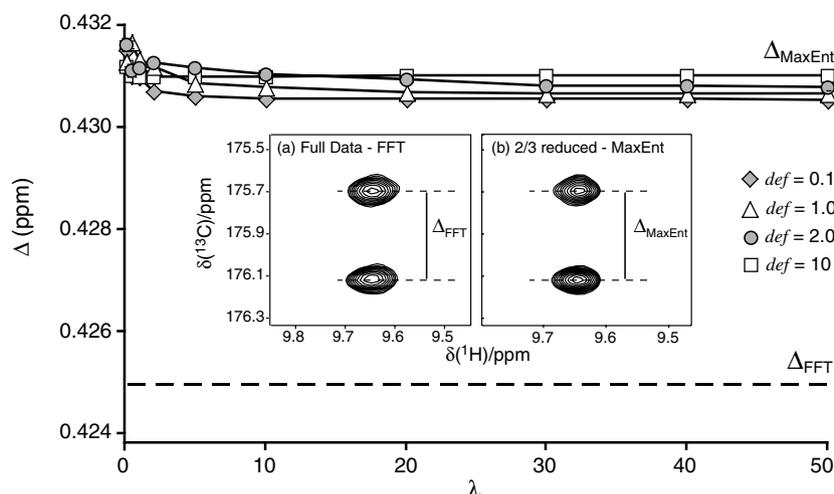
Fig. 1. Variations in MaxEnt parameters *def* and λ are compared to the peak separation measured by nonlinear least squares fitting of the resulting spectra for one experimental doublet in the isotropic spectrum. Forty spectra corresponding to *def* values of 0.1, 1.0, 2.0, and 10 and λ values of 0.1, 0.5, 1.0, 2.0, 5.0, 10.0, 20.0, 30.0, 40.0, and 50.0 were obtained by MaxEnt reconstruction. Higher values of *def* and lower values of λ produced considerably more uncertainty in the observed splittings in the simulated doublets, while low *def* and high λ values produced consistent, though not accurate, splitting in the doublets. The splitting in the FFT processed spectrum was 0.425 ppm and the MaxEnt splitting for $\lambda > 10$ was between 0.4306 and 0.4310 ppm.

We carried out MaxEnt processing of the isotropic data set in which 2/3 of the data was eliminated according to a sampling function that was exponentially weighted in both the nitrogen and carbon indirect evolution periods (Fig. 6b, 'sampsched2d' program in the Rowland NMR Toolkit). The insets of Fig. 1 compare a doublet obtained via either FFT or MaxEnt to indicate the agreement, and raise the point that $^{13}C'$–$^{13}C_\alpha$ RDC's will be on the order of 1–5 Hz, or just a few thousandths of a ppm at high fields (500 MHz or higher). We analyzed the cross-section (a.k.a. slice) of the data taken through the ^1H peak maximum by nonlinear least squares fitting of the spectrum to a double Lorentzian function. No combination of *def* or λ yielded a peak separation that was in agreement with that determined from the FFT processed doublet. The results in Fig. 1 also indicated that the MaxEnt processing yielded more uncertainty in the peak separation for small values of λ (e.g., $\lambda \sim 1$), which were used in our preliminary studies. The net deviation of the peak separation measured in the MaxEnt spectrum from the accepted value determined from the FFT processed spectrum is about 0.006 ppm when λ is large. In either regime of λ, the discrepancies would not influence the interpretation of protein 3D assignment spectra (e.g., HNCA, HNCO, HNCACO, etc. Also, taking the measured separation to be the accurate *J* coupling would usually be appropriate in this case since the error (0.006 ppm < 1 Hz) is much less than the typical $^{13}C'$–$^{13}C_\alpha$ scalar coupling of about 55 Hz. However a 1 Hz RDC corresponds to a difference in peak separation between the isotropic and aligned spectra of 0.008 ppm (assuming 125 MHz ^{13}C Larmor frequency as in this study). Therefore the discrepancy in peak separation between the FFT doublet (Δ_{FFT}) and the MaxEnt doublet (Δ_{MaxEnt}) in the isotropic spectrum is on the order of the value of the RDC we need to measure, especially when λ

is small. Yet Fig. 1 also suggests that the error between Δ_{FFT} and Δ_{MaxEnt} is regulated at larger values of λ (>10 for these data). If the error was systematic (i.e., constant) for both the isotropic and recoupled splittings, then it would self-cancel when subtracting doublet separations, and accurate RDCs should be measurable. The test in Fig. 1 suggested that there was a more fundamental discrepancy in doublet separations measured from FFT and MaxEnt processed data and led us to conduct more detailed analyses.

A series of simulated two-dimensional doublet spectra was created, including noise, and consisting of a single line in the first frequency dimension and split into a doublet in the second frequency dimension. The doublet separation was varied, as parameterized by the line width. Doublet separation was measured by one-dimensional nonlinear least squares fitting of the cross-section taken through the value of the peak maximum in the first dimension to a double Lorentzian function. In Fig. 2a, we show the percent difference in doublet separation between spectra obtained by MaxEnt reconstruction utilizing $\lambda = 10$ and by standard FFT processing. When noise is included and 2/3 of the data is eliminated (open diamonds), the percent difference is found to be proportional to the known signal separation specified in the simulations. When the data is noise free (shaded diamonds) the discrepancy still exists and is still proportional to the frequency separation between the signals. And for MaxEnt processing of the full uniformly sampled data set (filled diamonds), the difference again is observed. The results of Fig. 2a address several questions. First, it is apparent that the absolute values of the errors are very slight and would not normally be noted in typical protein NMR applications; only by normalizing the error against the line width can the error be characterized, and then the error amounts to only very small fractions of
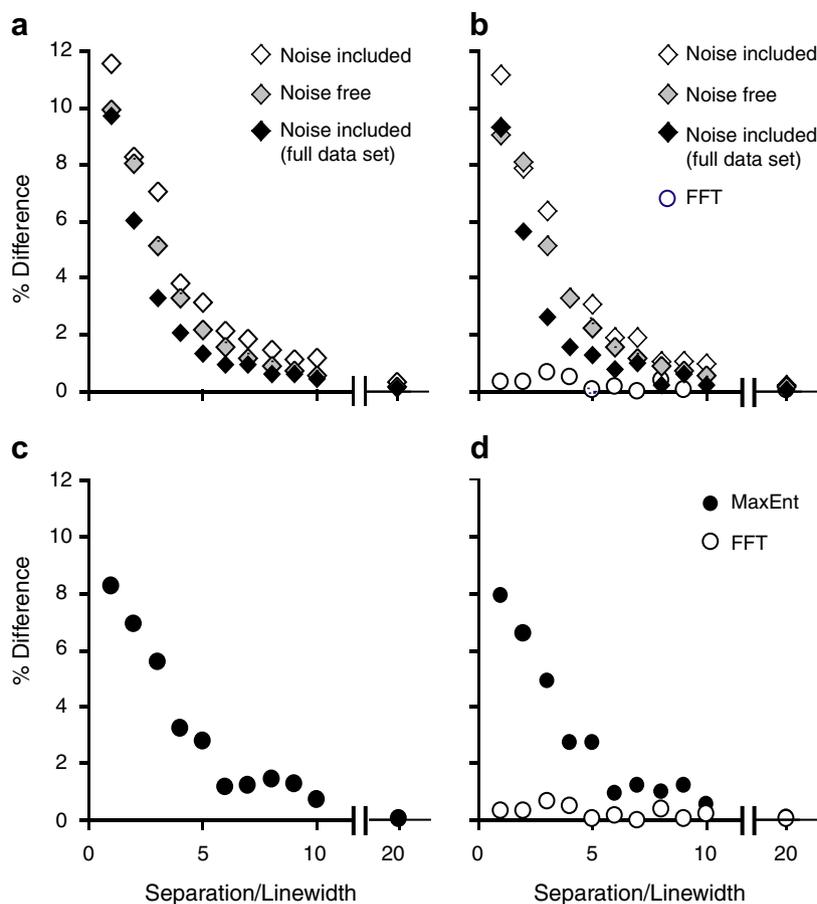
Fig. 2. Analysis of peak separations determined from FFT and MaxEnt processing of simulated doublets that possessed similar digital resolution and signal to noise as the experimental CaM spectra. The separation is parameterized as the ratio of the separation to the line width and was determined by nonlinear least squares fitting of a double Lorentzian function to the cross-section of the data taken at the peak maximum. (a and c) Examine the difference between doublets obtained by analysis of MaxEnt and FFT processed spectra. (b and d) Examine the difference between the known input values of the doublet separations and the values obtained by analysis of MaxEnt and FFT processed spectra. (a and b) Depict additional tests that show the error in the measured doublet separations is not attributable to the presence of noise or to the use of a reduced number of samples.

the line width. Next, Fig. 2a indicates that the error is not principally a result of reduced constraints on the signal due to the use of 2/3 fewer samples, or of uncertainty in the frequency position due to noise. It is observed in Fig. 2a that the error is only negligibly attenuated in the noise free data but somewhat more attenuated in the full, uniformly sampled data, indicating that these issues do contribute to uncertainty in the measured frequency positions, but are not the fundamental cause of the observed error. Since the signals were defined in the time domain simulations as Lorentzian (i.e., monoexponential decay), the error is observed when fitting the data with the corresponding Lorentzian line shape function, which may not be strictly correct (*vide infra*). Finally, Fig. 2a demonstrates that there is a non-random, systematic character to the error: overestimation of the splittings as a function of the separation to linewidth ratio.

Where Fig. 2a tested agreement between MaxEnt and FFT processed spectra, Fig. 2b presents the accuracy of each of these methods since the correct separations are defined in the simulations. As expected, the peak separa-

tions in FFT processed data (open circles) agree within <1% of the defined separations. This <1% variation against the known values is due to uncertainty in the peak fitting due to noise. This analysis of the FFT processed peaks validates our acceptance of peak separations determined from FFT processed data as the correct, control values. In Fig. 2b the FFT processed doublet separations closely reflected the input values, with the largest percent difference of only 0.65%. The MaxEnt processed doublet splittings significantly deviated from the input values as the separation to line width ratio became smaller, just as in Fig. 2a. Percent differences for the 2/3 reduced, noise free and 2/3 reduced, 4/5 reduced, and full data sets at a separation to line width ratio of 20 were 0.24%, 0.15%, 0.12%, and 0.06%. At a separation to line width ratio of 5, the percent differences were 3.1%, 2.2%, 2.7%, and 1.3%, while at a ratio of 1, the percent differences were 11.2%, 9.0%, 7.9%, and 9.4%, respectively.

Importantly Fig. 2a and b support that the error will be a constant in $^{13}C'$–$^{13}C_\alpha$ doublet separations that are measured independently for isotropic and aligned samples,

since the error has a negligible dependence on noise (which could vary between data sets) but is a strong function of the separation to line width ratio. We measured typical carbon line widths of 8 Hz in the calmodulin spectra, which places the doublets in a regime where the separation to line width ratio is about 7 (based on a typical $^{13}C'–^{13}C_\alpha$ ^1J of ca. 55 Hz). This ratio corresponds to relatively shallow regions of Fig. 2a and b, and will not be significantly changed by variations in the doublet separation due to the additional weak RDC of 1–5 Hz. Therefore the error will be approximately constant in either aligned or isotropic RDC doublets.

Fig. 2a and b also indicate that the separation to line width ratio is an objective, frequency independent means to determine the presence of the error. When the separation to line width ratio is greater than 10 the error is vanishingly small and is indistinguishable from the noise.

It is also apparent in Fig. 2a and b that MaxEnt processing of the full data set does attenuate the error relative to 2/3 reduced data. We further investigated the dependence of the error on data reduction by performing MaxEnt processing of a 4/5 reduced data set (Fig. 2c and d). The nonuniform sampling for 2/3 data reduction and 4/5 data reduction followed identical probability densities and yet the error observed in spectra obtained from the 4/5 reduced data set is slightly less than that obtained from the 2/3 reduced data set, and is comparable to the error obtained from the full data set. We will comment on this further in the discussion.

The evidence in Fig. 2 of a systematic character to the error in doublet splittings of MaxEnt processed data indicates that it should be possible to determine a rational explanation for its cause. Individual MaxEnt processed doublets (from simulated data) were overlaid against FFT processed doublets as a function of the peak separation to line width ratio (Fig. 3). At a separation to line width ratio of 3 (Fig. 3a), the tails of the resonances overlap significantly at the mean position of the two signals. The signal intensity at the mean of the doublet position in the MaxEnt processed spectrum lies significantly below that of the FFT spectrum, distorting the shape of the MaxEnt processed doublet. This observation is consistent with the well known nonlinear behavior of MaxEnt reconstructions to bias strong signals in the presence of weak signals. Specifically, the peak maxima are biased relative to the weak signal at the region of overlap of the Lorentzian tails, slightly sharpening the inside edges of the peaks of the doublets, and therefore effectively increasing the separation between them. The effect of this mechanism is particularly clear in Fig. 3a where the separation to line width ratio is 3 and the MaxEnt spectrum has peak maxima that are visibly shifted relative to the peaks in the FFT spectrum. Even when the separation to line width ratio is 10 (Fig. 3c), the MaxEnt spectrum does not fully reproduce the intensity of the FFT spectrum in the frequency region between the peaks, although at this point the error becomes comparable to uncertainty resulting from noise. We propose that the source of the error is a peak shape distortion that is due
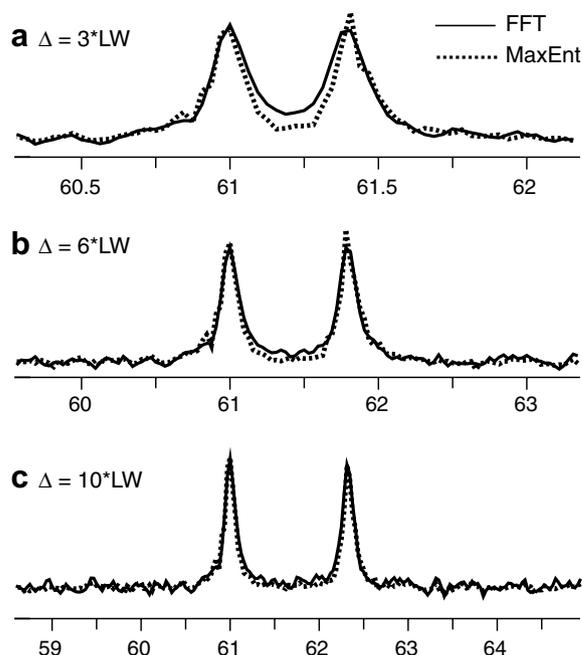

Fig. 3. One dimensional cross-sections taken from FFT processing of full length simulated data and from MaxEnt processing of the same data but reduced by 2/3 are superimposed. Separations of 3, 6, and 10 times the line width are considered in (a), (b) and (c) respectively. The horizontal axis is an arbitrary axis assuming carbon evolution at 125 MHz. The MaxEnt spectra (dashed lines) do not fully reproduce the intensity at the mean position of the two peaks, even when the separation is 10 times the linewidth.

to nonlinear amplitude bias of MaxEnt reconstructions and that artificially spreads the distance between peaks.

To further test this proposed mechanism, we performed MaxEnt reconstructions of the simulated data sets using deconvolution [7,44,45] with an exponential function in order to improve resolution and therefore increase the ratio of the doublet separation to line width. If the error is primarily a function of this ratio, then the error should decrease with such deconvolution. In Fig. 4a, MaxEnt processing of a simulated doublet consisting of 20 Hz line widths and using deconvolution with 5 Hz and 10 Hz exponential decays results in only very minor reductions of the error. However deconvolution of an exponential decay from the signal will bias noise substantially and thereby introduce greater frequency and line width uncertainty into the resulting spectra. The simulated data used in Fig. 4a already have a relatively low intrinsic signal to noise ratio (SNR) of about 30:1 (measured by analysis of the FFT processed spectrum) in order to be comparable to the experimental spectra. If the SNR is increased four-fold to about 120:1 then deconvolution does indeed result in significant decreases in the measured error, further supporting the proposal that nonlinear amplitude bias induces peak distortion and concomitant frequency errors in MaxEnt processing of doublet spectra.

The isotropic and aligned CaM 3D-HNCO spectra were reprocessed with MaxEnt reconstruction using $\lambda = 30$, which Fig. 1 suggests is a stable regime for the systematic
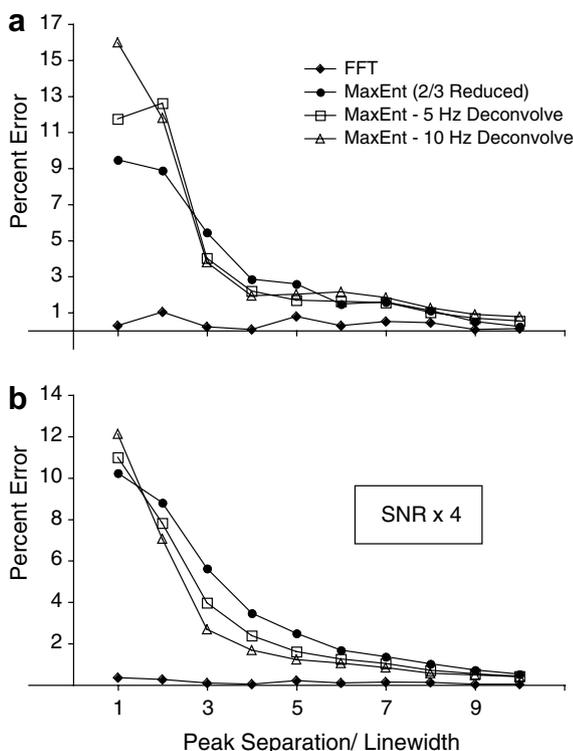
Fig. 4. Analysis of peak separations obtained from simulated spectra using FFT processing of uniform data, MaxEnt processing of data reduced by 2/3, and MaxEnt processing of 2/3 reduced data also employing deconvolution with 5 and 10 Hz exponential decays. Each peak in the doublet has a 20 Hz native line width. The percent error is for the peak separation determined by fitting a spectrum to a double Lorentzian function relative to the known input value in the simulation. Percent errors are shown for FFT processing (filled diamonds), MaxEnt reconstruction with no deconvolution (filled circles), MaxEnt with 5 Hz deconvolution (open squares) representing 25% line width reduction, and MaxEnt with 10 Hz deconvolution representing 50% line width reduction. The signal to noise ratio (SNR) was about 30:1 in (a) and about 120:1 in (b) for the FFT processed spectra.

frequency error in these data sets, and without the use of deconvolution. We selected at random 52 doublets to analyze for RDC values, which were obtained by measuring peak separations by nonlinear least squares fitting of one-dimensional cross-sections for both FFT and MaxEnt processing of the data. The comparison between RDCs extracted from FFT processing of the full data sets and MaxEnt processing of 2/3 and 4/5 reduced data sets is shown in Fig. 5. Good agreement is obtained in both cases, with RMSD values that are significantly below 1 Hz. If we assume an error of 0.25 Hz (0.002 ppm for ^{13}C at 125 MHz) in the measurement of a single peak position, the RDCs possess an inherent measurement uncertainty of 0.5 Hz and the agreement between FFT and MaxEnt derived RDCs is at or near the intrinsic uncertainty of the measurement.

4. Discussion

We observe good agreement between $^{13}C'$–$^{13}C_\alpha$ RDCs extracted from MaxEnt processing of either 2/3 reduced

or 4/5 reduced data and from FFT processing of full, uniformly sampled data. Nonuniform sampling preserves resolution by ensuring that experimental data at long evolution times is acquired, while MaxEnt processing is shown to have very high fidelity for measuring extremely small perturbations in chemical shifts. A class of slight systematic frequency errors that arise from MaxEnt processing of doublet NMR spectra and Lorentzian signal fitting are found to result from nonlinear amplitude bias in MaxEnt spectral estimations that leads to very slight peak shape distortions. The errors are shown to be systematic for large λ, and so are removed by subtractive analysis of J and J+D values.

A general approach to ensuring that def and λ are well chosen for the purpose of rendering the error systematic is to perform an abbreviated version of the test in Fig. 1. One first extracts from the 3D data a 2D plane that contains a representative doublet and processes that plane for a modest array of values of def and λ very quickly. In particular, one should select λ such that further increasing λ does not yield any discernible change in measured peak splitting. This procedure should be followed if peak separations are to be rigorously analyzed, however we wish to repeat an early point that much less stringent (and therefore computationally faster) conditions for def and λ will give spectra of very high quality for typical backbone and sidechain assignment and analysis. Also, in assuming that the error will cancel in a subtractive analysis of independently acquired spectra, it is necessary that the spectra utilize identical sampling schedules and that the separation to linewidth ratios be similar for the doublets being compared.

Strictly, peak shape distortion gives rise to an apparent frequency error when analyzing data with Lorentzian line shapes: correct frequencies would be measurable if an analytic form for the distorted doublet could be found. Whether or not a closed solution is even possible, it is evident by inspection that an analytic approximation to the distorted doublet could be quite complex. Since the peak distortions are extremely slight, the double Lorentzian is a very close approximation to the true line shape. Although other types of analysis are possible, such as the measurement of a center of gravity of each peak shape, Fig. 3a shows that the signal maxima are visibly shifted relative to their correct positions so that other analysis procedures will measure frequency errors also. And Lorentzian fitting is shown here to yield repeatable, dependable results: it produces systematic errors when analyzing doublet spectra. Also, nonlinear least squares fitting of a spectrum to a double Lorentzian function is a straightforward operation and does not require special expertise.

The nonlinear nature of MaxEnt reconstructions has been well documented and characterized for the case of amplitude bias in the resulting spectral estimates, and several methods have been developed to manage this nonlinearity [41]. A constant-λ algorithm provides uniform nonlinearity throughout an entire two-dimensional data

a Uniform vs. 2/3 Reduced   **b** Uniform vs. 4/5 reduced

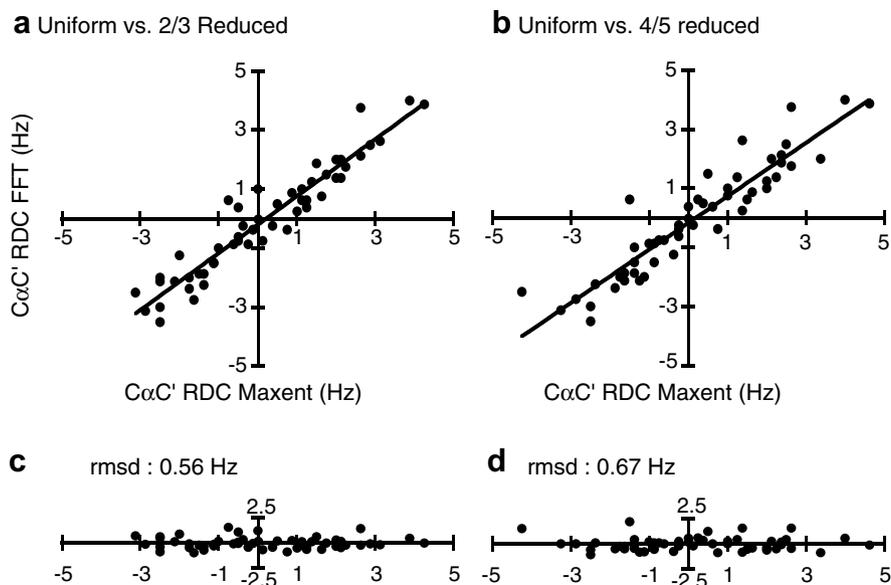**c**   rmsd : 0.56 Hz   **d**   rmsd : 0.67 Hz

Fig. 5. Comparison of RDCs measured for 52 randomly selected doublets between FFT processing of a uniformly sampled data set and MaxEnt processing of (a) the data reduced by 2/3 and (b) the data reduced by 4/5. Linear regressions are superimposed on each set of data. For (a) the linear regression was $y = 0.96x - 0.2$ with $R^2 = 0.91$. For (b) the linear regression was $y = 0.90x - 0.17$ with $R^2 = 0.87$. The residuals in (c) and (d) show no structure and have a random character, and the RMSD's for (c) and (d) are both substantially below 1 Hz.

set (or throughout an entire three dimensional data set as in this study) [41]. This measure ensures that nonlinear amplitude bias is systematic across the entire spectrum. Schmieder et al. also showed that the nonlinear amplitude response in a spectral reconstruction could be calibrated using injected signals and that the calibration could be used to accurately measure signal volumes [41]. In this study, we have shown that frequency errors are measurable in spectra of doublets when the separation to line width ratio is less than 10 and that these errors are attributable to nonlinear amplitude bias in MaxEnt spectral estimates that leads to very minor line shape distortion away from Lorentzian. In analogy to Schmieder et al., we also find that it is essential to have uniform nonlinearity in the reconstructed spectra in order to ensure that the frequency errors arising from nonlinear amplitude bias are systematic. Also, Fig. 2 represents a calibration of peak separation error using simulated data, which parallels the ability to calibrate peak volumes with simulated data.

As noted earlier in the text, the separation to line width ratio is an objective metric to determine if a given peak separation will be susceptible to the error characterized here. If this ratio is greater than 10, then the error is negligible and the separation measured from MaxEnt processed spectra may be safely inferred to be accurate and correct to a high degree of precision. For example, the separation to line width ratio will be greater than 10 in the case of ^{15}N–^{1}H ^{1}J couplings (ca. 92 Hz) in proteins, given that line widths in an indirectly detected ^{15}N dimension are commonly much less than 9 Hz. In this study we were operating in an intermediate regime for the separation to linewidth ratio of 7–8. The errors encountered in this regime were on the order of 1 Hz. If it is necessary to know the correct abso-

lute values of splittings when the separation to linewidth ratio is less than 10, then an *in situ* error analysis can be carried out, in analogy to the calibration of signal intensities introduced by Schmieder et al. [41]. Doublet signals of known separation and linewidths can be injected into vacant regions of the data and calibrations like those in Fig. 2 generated.

This work also suggests a broader finding about the origin of slight frequency errors in MaxEnt reconstructed spectra. The high accuracy of resonance positions found for MaxEnt processed spectra when the doublet separation is twenty times the line width confirms the fidelity of MaxEnt reconstructions of isolated signals. Since the sole condition for the occurrence of the frequency errors we have characterized is a proximity of signals relative to their line width, then this is a very general phenomenon and can help to explain the occurrence of slight frequency errors that have been previously documented to occur in MaxEnt reconstructions [6].

Spectral noise will introduce random fluctuations in peak positions regardless of the method of processing, and will not be considered further. But more work is needed to discern the role of the nonuniform sampling schedule in contributing to slight frequency errors of MaxEnt spectral estimates. We observed in simulated spectra in Fig. 2 that MaxEnt processing of the 4/5 reduced sampling schedule yielded slightly smaller errors in the analysis of peak separations than MaxEnt processing of 2/3 sampling. It is axiomatic that the 2/3 reduced schedule contains greater constraints on the spectrum and intrinsically should be more amenable to producing accurate spectral estimates. However, we also qualitatively observed a similar effect in the multidimensional RDC data shown in Fig. 5.

We noticed that many RDCs from the 4/5 reduced data agreed more closely with their corresponding FFT derived values than when derived from the 2/3 reduced data; however there is also lower signal to noise in the 4/5 reduced data which led to overall increased measurement error. Typical applications of nonuniform sampling of exponentially decaying signals may reduce the intrinsic signal to noise by 5–20% [33]. In Fig. 6 we illustrate the two sampling schedules and the probability density of each schedule; both schedules were created according to identical exponential weightings, although they still vary randomly within the constraint of the sampling density. Additional work is needed to identify more metrics beyond the sample density for gauging the quality of sampling schedules with respect to MaxEnt reconstruction. For example, the sparse sampling that occurs at long times (Fig. 6) could allow for an unfortuitous correspondence between samples and the signal nodes.

The time savings afforded by nonuniform sampling would have numerous benefits such as in enabling the study of unstable proteins. The significant advances of the last several years to accelerate NMR data acquisition has made it possible to efficiently and cost-effectively perform structural studies of proteins that are stable for 12–24 h. However RDC measurements, which are among the most time consuming experiments in the biomolecular NMR toolbox,

have not received the same degree of attention and so unstable proteins remain much less amenable to RDC analysis. This work shows a viable and robust alternative to making unstable proteins accessible to RDC analysis. The ability to achieve a fivefold savings in time also presents promising opportunities to reinvest the saved time and improve upon the intrinsic uncertainties in the measurements of RDCs. For example, although the intrinsic signal to noise ratio loss can be on the order of 5–20%, a portion of the saved time can be consumed with additional transients to offset the loss; or the entire time savings can be consumed with additional transients, which would yield useful net gains in SNR. Also, time can be invested to acquire more data at significantly longer indirect evolution times in order to improve the intrinsic spectral resolution, which is often severely under-realized at ultra-high fields [33]. This could enable the measurement of weak RDCs at much higher accuracy and precision than previously possible, particularly when combined with cryogenic probe technology to gain further signal-to-noise enhancement.

5. Conclusion

Elimination of up to 4/5 of experimental data is shown to lead to accurate RDCs by combined nonuniform sampling,
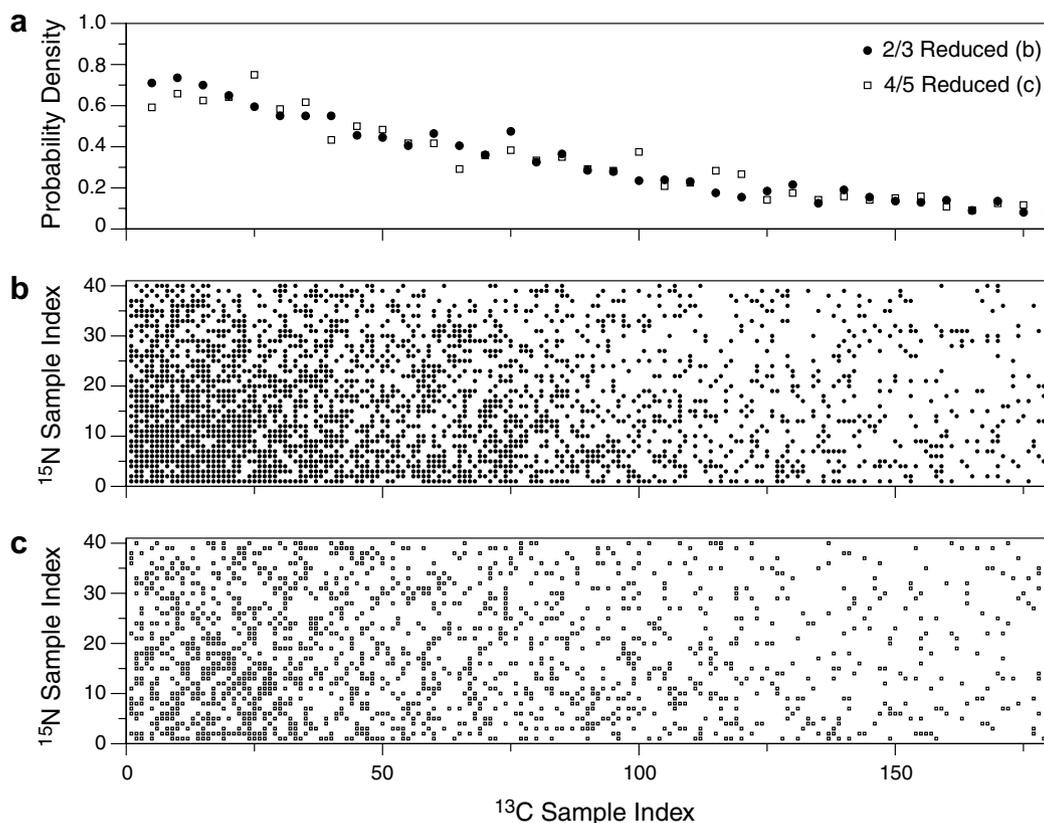

Fig. 6. Depiction of the sampling schedules used for the ^{15}N–^{13}C planes of the HNCO spectra. (a) Shows the sampling probability density for the 2/3 and 4/5 reduced schedules and shows they are qualitatively identical although random variations do exist. A value of 1 would indicate the use of all 40 uniform nitrogen samples at a given carbon evolution time. (b and c) Show the detailed samples in the nitrogen and carbon dimensions of the 2/3 and 4/5 reduced data respectively.

MaxEnt reconstruction, and nonlinear least squares fitting of the spectral data. We have also characterized a class of very slight frequency errors that arise in MaxEnt processing of spectra composed of uniform or nonuniformly distributed samples, and show evidence that the mechanism for this error stems from nonlinear amplitude bias inherent in MaxEnt reconstructions, and should in principle be observed with any nonlinear estimation algorithm. In the case of MaxEnt, the errors are extremely slight, generally confined to the 6th significant figure in the case of carbonyl chemical shifts in coupled 3D-HNCO spectra; the errors are systematic under stringent reconstruction conditions and do not prohibit the determination of RDC values by traditional subtractive J/J+D analysis. This methodology presents opportunities to perform RDC analyses of unstable proteins, or to utilize the time savings to significantly enhance the measurement of weak RDCs among heteronuclei, which could be of high value in constraining proton-poor regions of macromolecules.

Acknowledgments

References

[1] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution, Proc. Natl. Acad. Sci. USA 92 (1995) 9279–9283.

[2] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a diluted liquid crystalline medium, Science 278 (1997) 1111–1114.

[3] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, Use of dipolar H-1-N-15 and H-1-C-13 couplings in the structure determination of magnetically oriented macromolecules in solution, Nature Struct. Biol. 4 (1997) 732–738.

[4] J.J. Chou, F. Delaglio, A. Bax, Measurement of one-bond 15N–13C′ dipolar couplings in medium sized proteins, J. Biomol. NMR 18 (2000) 101–105.

[5] C.P. Jaroniec, T.S. Ulmer, A. Bax, Quantitative J correlation methods for the accurate measurement of 13C′–13Cα dipolar couplings in proteins, J. Biomol. NMR 30 (2004) 181–194.

[6] A.S. Stern, K.-B. Li, J.C. Hoch, Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of linear-prediction extrapolation and maximum-entropy reconstruction, J. Am. Chem. Soc. 124 (2002) 1982–1993.

[7] J.C. Hoch, A.S. Stern, NMR Data Processing, Wiley-Liss, New York, 1996.

[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] K. Ding, A.M. Gronenborn, Sensitivity-enhanced IPAP experiments for measuring one-bond 13C′–Ca and 13Ca–1Ha residual dipolar couplings in proteins, J. Magn. Reson. 167 (2004) 253–258.

[10] G. Armstrong, V.A. Mandelshtam, The extended Fourier transform for 2D spectral estimation, J. Magn. Reson. 153 (2001) 22–31.

[11] 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.

[12] J. Chen, A.J. Shaka, V.A. Mandelshtam, RRT: the regularized resolvent transform for high-resolution spectral estimation, J. Magn. Reson. 147 (2000) 129–137.

[13] V.A. Mandelshtam, The multidimensional filter diagonalization method I. theory and numerical implementation, J. Magn. Reson. 144 (2000) 343–356.

[14] H. Hu, A.A. DeAngelis, V.A. Mandelshtam, A.J. Shaka, The multidimensional filter diagonalization method II. Application to 2D projections of 2D, 3D and 4D NMR experiments, J. Magn. Reson. 144 (2000) 357–366.

[15] 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.

[16] P. Schanda, E. Kupce, B. Brutscher, SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, J. Biomol. NMR 33 (2005) 199–211.

[17] D. Frueh, H. Arthanari, G. Wagner, Unambiguous assignment of NMR protein backbone signals with a time-shared triple-resonance experiment, J. Biomol. NMR 33 (2005) 187–196.

[18] D. Frueh, D.A. Vosburg, C.T. Walsh, G. Wagner, Determination of all nOes in 1H-13C-Me-ILV-U-2H-15N proteins with two time-shared experiments, J. Biomol. NMR 34 (2006) 31–40.

[19] S. Kim, T. Szyperski, GFT NMR, A new approach to rapidly obtain precise high-dimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.

[20] B.E. Coggins, R.A. Venters, P. Zhou, Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein GB1 domain, J. Am. Chem. Soc. 126 (2004) 1000–1001.

[21] B. Shapira, L. Frydman, Arrayed acquisition of 2D exchange NMR spectra within a single scan experiment, J. Magn. Reson. 165 (2003) 320–324.

[22] L. Frydman, A. Lupulescu, T. Scherf, Principles and features of single-scan two-dimensional NMR spectroscopy, J. Am. Chem. Soc. 125 (2003) 9204–9217.

[23] E. Kupce, R. Freeman, Fast multi-dimensional NMR of proteins, J. Biomol. NMR 25 (2003) 349–354.

[24] E. Kupce, R. Freeman, The radon transform: a new scheme for fast multidimensional NMR, Concepts Magn. Reson. Part A 22A (2003) 4–11.

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

[26] E. Kupce, T. Nishida, R. Freeman, Hadamard NMR spectroscopy, Prog. NMR Spectr. 42 (2003) 95–122.

[27] R. Freeman, E. Kupce, Distant echoes of the accordion: reduced dimensionality, GFT-NMR, and projection–reconstruction of multi-dimensionalsSpectra, Concepts in Magnetic Resonance Part A 23A (2004) 63–75.

[28] V.Y. Orekhov, I.V. Ibraghimov, M. Billeter, MUNIN: a new approach to multi-dimensional NMR spectra interpretation, J. Biomol. NMR 20 (2001) 49–60.

[29] A. Gutmanas, P. Jarvoll, V.Y. Orekhov, M. Billeter, Three-way decomposition of a complete 3D (15)N-NOESY-HSQC, J. Biomol. NMR 24 (2002) 191–201.

[30] D. Malmodin, M. Billeter, Multiway decomposition of NMR spectra with coupled evolution periods, J. Am. Chem. Soc. 127 (2005) 13486–13487.

[31] R. Bruschweiler, F. Zhang, Covariance nuclear magnetic resonance, J. Chem. Phys. 120 (2004) 5253–5260.

[32] R. Bruschweiler, Theory of covariance nuclear magnetic resonance spectroscopy, J. Chem. Phys. 121 (2004) 409–414.

[33] 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.

[34] G. Bodenhausen, R.R. Ernst, The accordion experiment, a simple approach to three-dimensional NMR spectroscopy, J. Magn. Reson. 46 (1981) 367–373.

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

[36] Z.-Y.J. Sun, D. Frueh, P. Selenko, J.C. Hoch, G. Wagner, Fast assignment of 15N-HSQC peaks using high-resolution 3D HNco-caNH experiments with nonuniform sampling, J. Biomol. NMR 33 (2005) 43–50.

[37] D. Rovnyak, C. Filip, B. Itin, A.S. Stern, G. Wagner, R.G. Griffin, J.C. Hoch, Multiple-quantum magic-angle spinning spectroscopy using nonlinear sampling, J. Magn. Reson. 161 (2003) 43–55.

[38] Z.-Y.J. Sun, S.G. Hyberts, D. Rovnyak, S. Park, A.S. Stern, J.C. Hoch, G. Wagner, High-resolution aliphatic side-chain assignments in 3D HCcoNH experiments with joint H–C evolution and non-uniform sampling, J. Biomol. NMR 32 (2005) 55–60.

[39] D. Frueh, Z.-Y.J. Sun, D.A. Vosburg, C.T. Walsh, J.C. Hoch, G. Wagner, Non-uniformly sampled double-TROSY hNcaNH experi-ments for NMR sequential assignments of large proteins, J. Am. Chem. Soc. 128 (2006) 5757–5763.

[40] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J. Biomol. NMR 6 (1995) 277–293.

[41] P. Schmieder, A.S. Stern, G. Wagner, J.C. Hoch, Quantification of maximum-entropy spectrum reconstructions, J. Magn. Reson. 125 (1997) 332–339.

[42] 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.

[43] G.J. Daniell, P.J. Hore, Maximum entropy NMR—a new approach, J. Magn. Reson. 84 (1989) 515–536.

[44] J.C. Hoch, A.S. Stern, Maximum entropy reconstruction, spectrum analysis and deconvolution in multidimensional nuclear magnetic resonance, Methods Enzymol. 338 (2001) 159–178.

[45] N. Shimba, A.S. Stern, C.S. Craik, J.C. Hoch, V. Dotsch, Elimination of 13Cα splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction, J. Am. Chem. Soc. 125 (2003) 2382–2383.