


A non-uniform sampling approach enables studies of dilute and unstable proteins

Tomas Miljenović¹ · Xinying Jia¹ · Peter Lavrencic^{1,2} · Bostjan Kobe^{2,3,4} · Mehdi Mobli¹ 

Received: 1 December 2016 / Accepted: 23 January 2017 / Published online: 10 February 2017
© Springer Science+Business Media Dordrecht 2017

Abstract NMR spectroscopy is a powerful method in structural and functional analysis of macromolecules and has become particularly prevalent in studies of protein structure, function and dynamics. Unique to NMR spectroscopy is the relatively low constraints on sample preparation and the high level of control of sample conditions. Proteins can be studied in a wide range of buffer conditions, e.g. different pHs and variable temperatures, allowing studies of proteins under conditions that are closer to their native environment compared to other structural methods such as X-ray crystallography and electron microscopy. The key disadvantage of NMR is the relatively low sensitivity of the method, requiring either concentrated samples or very lengthy data-acquisition times. Thus, proteins that are unstable or can only be studied in dilute solutions are often considered practically unfeasible for NMR studies. Here, we describe a general method, where non-uniform sampling (NUS) allows for signal averaging to be monitored in an iterative manner, enabling efficient use of spectrometer time, ultimately leading to savings in costs associated with instrument and isotope-labelled protein use. The method requires preparation of multiple aliquots of the protein sample that are flash-frozen and thawed just before acquisition

of a short NMR experiments carried out while the protein is stable (12 h in the presented case). Non-uniform sampling enables sufficient resolution to be acquired for each short experiment. Identical NMR datasets are acquired and sensitivity is monitored after each co-added spectrum is reconstructed. The procedure is repeated until sufficient signal-to-noise is obtained. We discuss how maximum entropy reconstruction is used to process the data, and propose a variation on the previously described method of automated parameter selection. We conclude that combining NUS with iterative co-addition is a general approach, and particularly powerful when applied to unstable proteins.

Keywords Non-uniform sampling · Targeted acquisition · Jittered sampling · Maximum entropy reconstruction · Proteins

Introduction

Since the first protein structure in solution was solved using nuclear magnetic resonance (NMR) spectroscopy in 1984 by Wüthrich and co-workers (Havel and Wüthrich 1984), 11,628 NMR structures have been deposited in the Worldwide Protein Data Bank. While the biggest protein analysed by solution NMR spectroscopy has a molecular weight of 900 kDa (Fiaux et al. 2002), in practice NMR is mainly used for smaller proteins or protein domains due to the signal overlap and broadening associated with larger proteins. In routine applications of biomolecular NMR, structural characterisation of proteins is carried out by application of multidimensional triple resonance experiments that commonly require weeks of NMR data-acquisition time (Ikura et al. 1990). This approach becomes impractical for

✉ Mehdi Mobli
m.mobli@uq.edu.au

¹ Centre for Advanced Imaging, The University of Queensland, St Lucia, QLD, Australia

² School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, Australia

³ Australian Infectious Diseases Research Centre, The University of Queensland, St Lucia, QLD, Australia

⁴ Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, Australia

metastable proteins that are only fleetingly monodisperse and/or folded under the conditions required for structural analysis.

Non-uniform sampling (NUS) can dramatically reduce NMR data-acquisition time of triple resonance spectra, by sub-sampling the full time-domain grid (Rovnyak et al. 2004). NUS data, however, cannot be processed by the conventional discrete Fourier transform (DFT) approach and requires non-DFT methods (Mobli and Hoch 2014). Several spectral reconstruction methods have been presented that are capable of processing NUS data; a family of related iterative reconstruction methods—including iterative soft thresholding, compressed sensing and maximum entropy reconstruction—have proven to yield stable and reliable frequency-domain spectral representations and are currently employed routinely in many laboratories (Holland et al. 2011; Kazimierczuk and Orekhov 2011; Mobli and Hoch 2014; Stern et al. 2007).

The earliest application of NUS in NMR demonstrated that the distribution of sampling points strongly affects the outcome of the spectral reconstruction, and it was proposed that random samples drawn from an exponentially decaying probability density function (PDF) that mimics the decay of the signal would, much like window functions, yield the best outcome in terms of sensitivity and resolution (Barna et al. 1987). Thus it was clear that the distribution of the sampling points on the time-domain grid can affect both the sensitivity and resolution of the data, and formal relationships between the distribution and sensitivity later showed that NUS was, per unit time, capable of improving the signal-to-noise (S/N) of a given NMR experiments (Rovnyak et al. 2011). In cases where the signal is heavily under-sampled, it has been speculated that additional gains can be achieved by further augmenting the PDF through optimisation of some metric in a transform domain of the sampling function (Lustig et al. 2007). Later work, however, has shown that this is not possible for NMR data where the signal frequencies and phases are not known *a priori* (Schuyler et al. 2011). Instead, promising results have been found in efforts directed towards providing pseudo-random sampling distributions that retain the incoherence of random sampling, while reducing the overall dependence of the features of the reconstructed spectrum on the seed-dependent realisation of the sampling distribution (Hyberts et al. 2010; Kazimierczuk et al. 2008; Mobli 2015). To achieve this, jittered sampling has emerged as a general method that can be adapted to any arbitrary PDF, yielding highly seed-independent results (Mobli 2015; Worley 2016).

Therefore, at present the most robust and time-efficient method of acquiring multidimensional NMR data appears to be by random, weighted NUS, drawn from a PDF divided in equiprobable jittered regions, and processed by an iterative reconstruction method. Indeed, in the past

decade our group has solved the structure of ~30 peptides and proteins, where we have exclusively used NUS and maximum entropy reconstruction, in place of traditional sampling and DFT, for triple-resonance experiments used in resonance assignments. These have, however, in general involved highly stable proteins at moderately high concentrations (>300 μM).

Here, we present an approach involving the application of NUS and maximum entropy to a more challenging system not ideally suited for NMR analysis. The method is demonstrated on a metastable domain of the MyD88 adaptor-like (MAL) protein, which is a key component in the Toll-like receptor (TLR) signalling cascade of the human innate immune system (Thomas et al. 2012). Although MAL is a cytosolic protein, the four previously published crystal structures of the MAL TIR domain (MAL^{TIR}) (residues 79–221; ~16 kDa) show a mixture of reduced and oxidised cysteine residues (Lin et al. 2012; Snyder et al. 2014; Valkov et al. 2011; Woo et al. 2012). The role of the oxidised cysteines and their effect on the protein structure and function remains poorly understood. Solution-state studies by NMR promise to provide critical insights into the redox state of the protein and how this may affect the protein structure. Previous crystallography studies have optimised high-yield expression of the protein in *Escherichia coli* and the protein could be readily prepared and flash-frozen for subsequent analysis (Valkov et al. 2011). In our early attempts to stabilise the protein for NMR studies, it was apparent that the protein was only stable as a monomer at (i) low concentration (<300 μM), (ii) low temperatures (<19 °C), (iii) high salt concentrations (>200 mM NaCl) and (iv) high pH (>8.4). Even under such unfavourable NMR conditions, the protein is only stable for a brief (12 h) period of time, making it particularly ill-suited for NMR studies. Indeed, a survey of the BMRB's ~12,000 submissions shows that only 144 (1.2%) of these are at a pH greater than 8.4 and only 32 (0.3%) of these are at a temperature less than 19 °C.

To obtain NMR data of suitable quality with minimum time requirements, we first defined an NUS sampling schedule that would provide sufficient resolution in the indirect dimensions, and used jittered sampling with exponential weighting to ensure reproducibility and sensitivity enhancement. Within the time constraints of the experiment, we acquired eight scans (transients) per dataset, which in itself yielded a spectrum with insufficient sensitivity for analysis. To overcome this, we repeated the experiment multiple times, co-adding the resultant datasets, and evaluated the reconstructed spectrum until sufficient S/N had been obtained for subsequent analysis. This ensured that a minimum number of experiments were acquired for each multidimensional experiment. We demonstrate the application of this approach using the 3D HBHA(CO)NH

experiment. Our approach is related to the targeted acquisition approach previously described for NUS (Jaravine and Orekhov 2006). However, while acquisition of additional sample points poses a relatively small time burden, increasing the sensitivity of the data by increasing the number transients is a more time-consuming task, potentially providing a more significant time saving and is likely to be practically easier to implement. We discuss the advantages of the approach and how it may be implemented as a general tool in multidimensional NMR spectroscopy. We also discuss how parameter selection in maximum entropy reconstruction affects the outcome and propose a modification to automated parameter selection that aids spectral analysis in such challenging systems.

Methods

Production of MAL and NMR sample preparation

The plasmid coding for the C116A mutant of MAL^{TIR} (MAL^{TIRC116A}) was transformed into *E. coli* BL-21 (DE3) cells by heat-shock, and grown overnight in a starter culture of Luria's broth (LB) in the presence of 100 mg/L ampicillin while shaking at 37 °C. The protein was then expressed in the same media while shaking at 37 °C until the cells reached an optical density (OD) of 0.7 at the wavelength of 600 nm. The sample was centrifuged at 800 x g, washed in M9 salts and resuspended in minimal M9 media containing ¹³C glucose and ¹⁵N ammonium chloride, until the cell density reached an OD of 0.8 at the wavelength of 600 nm. The temperature was then reduced to 20 °C and the cells were induced with 1 mM isopropyl-1-thiogalactopyranoside (IPTG) for overnight expression. The protein was purified as previously described (Valkov et al. 2011). Aliquots of the protein were flash-frozen in N₂(l) and stored at -80 °C. NMR data were acquired at 291 K using a 900 MHz AVANCE III spectrometer (Bruker BioSpin, Germany) equipped with a cryogenically cooled probe.

300 μL of a solution containing 300 μM ¹³C/¹⁵N-labelled MAL^{TIRC116A}, in 20 mM TRIS buffer at pH of 8.6, 200 mM NaCl, and 5% D₂O was added to a susceptibility-matched 5 mm outer-diameter microtube (Shigemi Inc., Japan).

NUS data acquisition

The 3D HBHA(CO)NH experiment from the Bruker pulse sequence library was used and NUS data acquisition enabled through Topspin 3.0. Sampling schedules were prepared by first establishing the full time-domain grid, containing 64 increments (128 points - T_{2max} = 7.9 ms) in the indirect ¹H dimensions and 50 increments (100 - T_{2max} = 17 ms) in the indirect ¹⁵N dimension, resulting in

a total of 3200 coordinates containing 12,800 data records. 600 data coordinates (~19%) were sampled using NUS. The grid points were weighted by a PDF defined based on the decay of the signal according to Eq. 1.

$$p(i) = \prod_{k=1}^L e^{\frac{-ixlw_k}{sw_k}} \quad (1)$$

where lw_k and sw_k are the linewidth and spectral window in the k^{th} dimension, respectively. The ¹⁵N dimension is acquired using constant time and a nominal 1 Hz weighting was applied. The indirect ¹H dimension is acquired using a semi-constant time approach and an estimated weighting corresponding to a line-width of 15 Hz was applied. Based on the above PDF, the predefined time-domain grid was divided into 600 equiprobable regions as previously described (Mobli 2015). A sample point was drawn from each region based on the underlying PDF, using the one pass approach (Eq. 2) previously implemented for multidimensional NMR (Mobli et al. 2010), where each point is given a rank according to

$$r(i) = x(i)^{1/\rho(i)} \quad (2)$$

where where $x(i)$ is a random number between 0 and 1. The highest rank $r(i)$ is then retained from each region. To acquire a single repetition of the experiment required ~43 min. Each experiment was set up to acquire 8 transients/scans per increment resulting in a total of ~6 h. Two such datasets could be acquired within the stability window (12 h) of the sample. A total of 9 datasets—all using the same sampling schedule—were acquired (72 transients or ~51.6 h total). All datasets were acquired consecutively over a period of 5 days. Care was taken to optimise spectrometer settings for each sample and to use the same receiver gain values to allow for linear co-addition. All parameters were stable during this period include the phase of the individual datasets and required no further adjustments.

Spectral reconstruction

Non-Fourier analysis of our spectrometer data was performed using the Rowland NMR Toolkit (Stern et al. 2002). An automated processing script generator was used to convert the data into the Toolkit format. The direct dimension was processed using the traditional DFT approach (using a Gaussian window function with a ~20 Hz linewidth, with water suppression achieved through time-domain deconvolution). The indirect dimensions were processed using maximum entropy reconstruction. Although the theory has been provided in detail previously, we will here summarise some key aspects relevant to the current study (Daniell and Hore 1989). Here, the entropy term is defined as

$$S(f) = - \sum_{n=0}^{N-1} \left\{ x_n \log \left(\frac{x_n + \sqrt{4 + x_n^2}}{2} \right) - \sqrt{4 + x_n^2} \right\} \quad (3)$$

where $x_n = \frac{|f_n|}{def}$ and def is treated as an adjustable parameter. As noted previously (Daniell and Hore 1989), when the value of def becomes very small, the $\left(\frac{x_n + \sqrt{4 + x_n^2}}{2} \right)$ term goes towards x_n and the entropy resembles a scaled version of the Shannon entropy $\left(S(f) = - \sum_{n=0}^{N-1} x_n \log x_n \right)$. Hence, low values of def , drive the spin $\frac{1}{2}$ entropy (Eq. 3) towards the Shannon entropy. The maximum entropy spectrum is that which maximises the entropy (Eq. 3), while retaining consistency with the experimental data. Consistency is defined by first calculating the unweighted χ -squared statistic,

$$C(f, d) = \sum_{i=0}^{M-1} |m_i - d_i|^2 \quad (4)$$

where m_i is the inverse DFT of the candidate spectrum and d_i is the measured time response. Then a threshold, C_0 (referred to as *AIM* in the Rowland NMR Toolkit) is defined and consistency is achieved by maintaining $C(\mathbf{f}, \mathbf{d}) \leq C_0$. The two constraints (3 and 4) are optimised by maximising the object function:

$$Q(\mathbf{f}, \mathbf{d}) = S(\mathbf{f}) - \lambda C(\mathbf{f}, \mathbf{d}) \quad (5)$$

where λ is a Lagrange multiplier. Setting the value of C_0 to a very small value, or similarly by setting the value of λ to a large value, ensures that the mock data does not deviate from the experimental data, in general leading to more linear reconstructions (Paramasivam et al. 2012). In practice, either the value of C_0 or the value of λ is treated as an adjustable parameter.

Maximum entropy parameter selection

Automated maximum entropy reconstruction—AUTO (Mobli et al. 2007) The value of C_0 (*AIM*) is determined by evaluating the noise content of the last acquired FID. The noise value for the 9 experiments was determined and found to be consistent in all datasets. To ensure consistent setting of *AIM* and to aid robustness in comparisons, the value of *AIM* was determined for the co-addition of all 9 datasets, assuming this would produce the most accurate noise value. This value was then scaled by $\sqrt{n_{scans}}$ to find the value for different co-additions. The scaling factor,

def , was set based on the initial *AIM* value and the number of points in the dataset ($def = AIM \times \sqrt{M}/\sqrt{N}$), as described previously. The AUTO approach uses the constant λ method, where the strongest signals are used to perform trial reconstructions using the determined values of *AIM* and def , while the value of λ for each reconstruction is recorded and averaged. This averaged value is then used together with the predefined value of def to reconstruct the full spectrum.

Maximum entropy interpolation—MINT (Paramasivam et al. 2012). In this strategy the value of C_0 is set significantly below the noise level, in principle leading to overfitting of the data. Again to implement this approach in practice, the constant λ approach was used, and rather than setting a small value of C_0 , a large value of λ is chosen. The value of λ is iteratively raised until a limit is reached where the reconstruction converges in a reasonable timeframe. As described, the method does not discuss the setting of def and we have here used the same setting of def as defined by the AUTO approach.

Shannon-weighted augmentation of reconstruction parameters—SHARP. This approach is a variation of the AUTO approach where the value of def is scaled down, resulting in an entropy function, which is weighted towards the form of the Shannon entropy [see above and (Daniell and Hore 1989)]. In practice, scaling down def results in longer convergence times, similar to what is observed when increasing the value of λ . Although the reconstruction times increase dramatically as def is lowered, the appearance of the spectrum is most strongly affected by the initial reductions in def (by 10^{-1} – 10^{-2}) and stabilises at values scaled by 10^{-3} or less. In SHARP, the value of def as defined by AutoMaxEnt and then by default scaled by a factor of 10^{-3} ($def_{SHARP} = def_{AUTO}/10^3$). All other parameters, including setting of constant λ , is performed as in the AUTO approach.

Results and discussion

Nine sparsely sampled 3D HBHA(CO)NH experiments were acquired using the same sampling schedule. The data were acquired under identical conditions, including the receiver gain setting, to facilitate summation. All of the samples used were from the same original stock, so there was little/no variation in sample condition for each experiment. An important consideration in the presented approach is that the protein of interest must be produced in sufficient quantity that multiple NMR samples can be prepared and stored for subsequent analysis. In our experience, once overexpression is achieved in *E. coli*, the cost of preparing large amounts of protein is not prohibitive. For a 16 kDa protein, a 2 L bacterial culture producing a

yield of 10 mg/L of protein allows for preparation of ~4 mL of 300 μ M solution, which can be used to prepare >10 samples (in 300 μ L NMR cells). Further, we find that the majority of proteins can be flash-frozen in liquid nitrogen and thawed gently on ice without significant loss of protein. This is indeed the practice in many crystallography laboratories, where the procedure ensures strict maintenance of sample conditions during optimisation of crystal conditions.

The datasets were converted to the Rowland NMR Toolkit format and summed using the previously described *planemath* tool within the software suit (Mobli et al. 2006). This generated 9 datasets, the original experiment and eight additions to this by subsequent datasets (Fig. 1). This represents realisations of the data where the number of scans/transients is increased from 8 to 72 by increments of 8. The final dataset, being the sum of 9 experiments, will have 3 times the S/N of the original experiment (as signal strength increases linearly with each additional scan, while random noise contributions only increase by a factor of $\sqrt{n_{scans}}$). The co-additions of the experiments were reconstructed independently using maximum entropy reconstruction with automated parameter selection (AUTO method described above).

Figure 2 shows the plots where sensitivity is plotted as a function of number of increments acquired for a decaying signal (Rovnyak et al. 2011). As noted previously, the sensitivity increases rapidly as data are acquired initially, and reaches a maximum at $\sim 1.25 \times T_2$ and steadily decreases after this point. There is therefore a point for any given experiments where the maximum sensitivity is reached by sampling alone. At this point, the only way to improve S/N is to acquire additional iterations of the experiment and, as the signals add linearly and the noise does not, an improvement of $\sqrt{n_{scans}}$ can be achieved. In Fig. 2, we have plotted separate curves for nine co-additions of a given experiment. To aid in interpretation, the x-axis is plotted as the number of increments (rather than time), assuming a SW of 2500 Hz, and the curve is constructed assuming

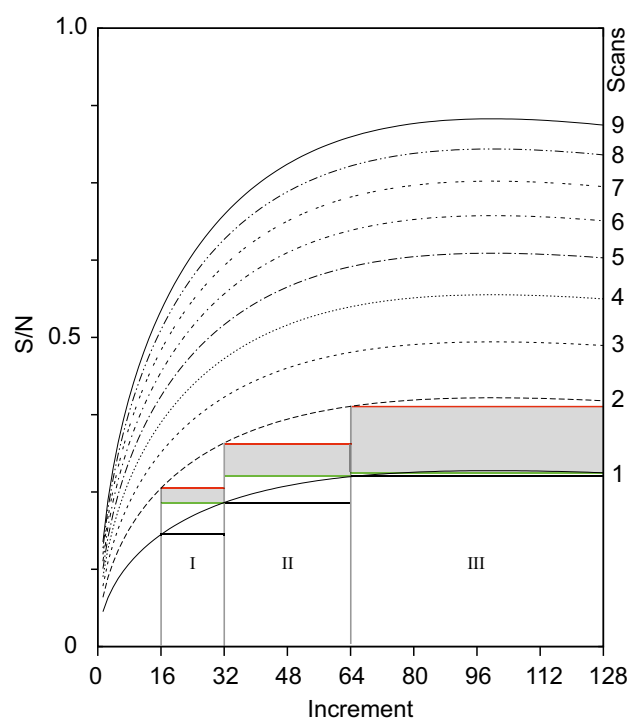
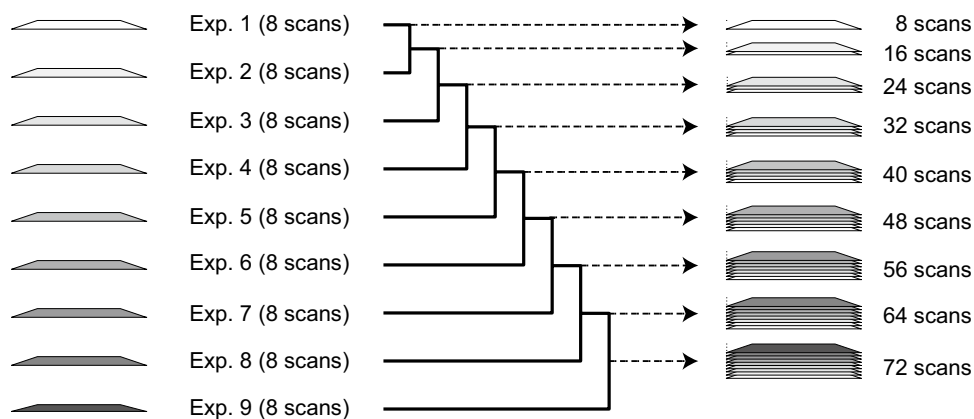


Fig. 2 Sensitivity improvements as a function of sampling (increments) and signal addition (scans). The plot is generated assuming a T_2 of 30 ms (10 Hz line-width) and a sweep width of 2500 Hz. Each line represents accumulation of an additional dataset, in effect doubling the acquisition time. The *nine lines* represent the sensitivity gain achieved by addition of each of the nine datasets used here. The *regions I, II and III* illustrate the gain in sensitivity by sampling additional increments (distance between the *horizontal black and green lines*), compared to increase in sensitivity if the same amount of time is spent on signal averaging (distance between the *black and red line*). The *grey boxes* illustrate the additional gain in sensitivity achieved by signal averaging compared to acquisition of additional time increments. It can be seen that signal averaging for later data increments provides a significantly higher S/N increase than at lower evolution times, compared to sampling additional time increments

Fig. 1 The co-addition scheme used to sum the nine HBHA(CO)NH datasets. Each individual experiment was acquired with 600 NUS samples taken from a 64×50 ($^1\text{H}/^{15}\text{N}$ respectively) grid using jittered random sampling, with eight scans per increment. The data were summed in the time domain and stored for subsequent reconstruction



a line-width of 10 Hz. The plots illustrate that as additional scans are added, the improvements in S/N reduce per unit time, a consequence of the nonlinear change in S/N with signal averaging ($S/N \propto \sqrt{n_{scans}}$). We note that the improvements in S/N are with regards to the underlying data and are therefore independent of the reconstruction method used.

In Fig. 2, we have further defined three regions (I, II and III), where each region signifies doubling of the acquisition time. The three regions are taken at the beginning, middle and late in the signal evolution time. The distance between the black and the green horizontal lines in each region signifies the improvement in S/N achieved by acquisition of additional time points. The distance between the black and the red horizontal lines shows the improvements in S/N that can be achieved by signal averaging instead of the acquisition of additional points. The shaded area shows the benefit of signal averaging, compared to acquisition of additional data points (non shaded area between black and green bars). The figure illustrates that while sampling the early time points $\sim 0.2 \times T_2$ (or in this case ~ 16 increments), there is barely any benefit in increasing the number of scans and time is better spent acquiring additional points, which in linear sampling additionally improves resolution, while in NUS reduces the level of noise due to sampling artifacts. However, as the evolution time is increased to $\sim 0.4 \times T_2$ (or ~ 32 increments) the benefits of signal averaging become more pronounced and beyond $0.8 \times T_2$ (~ 64 increments), S/N improvements are only possible through signal averaging. An intriguing consequence of this analysis (Fig. 2) is that signal averaging is more beneficial at later evolution times than at earlier evolution times. Non-uniform signal averaging is an area related to NUS, which has not yet been investigated in depth and our results are in agreement with a previous analysis showing a benefit where later time points are signal averaged more than early time points (Hodgkinson et al. 1996).

The above observations (Fig. 2) are indeed borne out in our experiments. We have monitored the spectral reconstructions by assessing the signal intensity in the ^1H - ^{15}N projection of the 3D experiment, and also by monitoring the signal intensity of known weak signals along the indirect ^1H dimension (Fig. 3). In our experiments, we can clearly see a sharp initial increase in S/N through co-addition and also the plateauing of improvements beyond the co-addition of the first 7 datasets, reaching a point of diminishing returns. In practical terms, to increase the S/N by a factor of 2 for the initial experiment requires co-addition of 4 experiments (from 8 to 32 scans); this results in ~ 18 h of additional experiment time. However, to again double the sensitivity requires an additional ~ 86 h of acquisition time (or co-addition of another 12 datasets). Therefore, for the majority of NMR experiments, extending the

experiment time beyond ~ 60 h by signal averaging would result in prohibitively long acquisition times (>7 days per 3D experiment).

Based on these findings, a general approach would be to: (i) define an NUS scheme that would ensure sufficient resolution in all dimensions, while optimising sensitivity per unit time (Rovnyak et al. 2011); (ii) acquire a minimum number of scans per increment (in some cases limited by phase cycling); (iii) reconstruct the spectrum using a suitable algorithm; (iv) assess the spectrum with respect to some spectral qualities, ideally in an automated manner as previously described (Jaravine and Orekhov 2006); (v.a.) if the data is of sufficient quality terminate the procedure (v.b.), if data quality is insufficient go to step (ii) and co-add to the existing dataset until convergence or (v.c.) terminate data acquisition if ~ 60 h of acquisition time is reached—beyond this time the experiment can be considered impractical in most cases. This approach would ensure that a minimal amount of time is spent on each experiment—while terminating impractical experiments. It should be noted that the reconstruction method used in step (iii) may influence the sampling schedule used in step (i) and what spectral quality is assessed in step (iv), this however does not influence the general procedure.

Finally, we sought to assess the effect of parameter selection on the reconstructions. The reconstructions so far have all been performed according to the AUTO scheme; however, a method of parameter selection was recently introduced, called MINT, to obtain more linear MaxEnt reconstructions. In contrast to these methods, it is also possible to weight the entropy function from that derived for a spin $1/2$ system, towards a form that closer resembles the Shannon entropy, by reducing the value of the parameter def ; we refer to this as SHARP, where the AUTO method is used to define C_0 close to the noise level, but the value of def is significantly scaled down (see also the Methods section). The value of λ is found in the same way as for AUTO, but instead using the scaled def value. The scaling of def has a significant effect on both the appearance of the spectrum and the convergence of the reconstruction.

In Fig. 4, we sought to demonstrate how scaling of def affects the number of iterations (loops) required for convergence and also show the effect of the scaling on the spectrum. The same trace as in Fig. 3f is used. The figure shows that def can be scaled by 2 or 3 orders of magnitude without significant computational burden, but that beyond this, the effect on the reconstruction is small while the effect on convergence becomes significant. These results are compared to the AUTO and MINT approaches previously described (Fig. 4). In the example where def is scaled by 10^{-4} , Fig. 4f, the plane used for this comparison converged after ~ 30 iterations. The spectrum as a whole, however, included planes that required many thousand iterations to

Fig. 3 Sensitivity improvements by co-addition. **a** The ^{15}N - ^1H HSQC spectrum of MAL^{TIRC116A}. Representative strong and weak resonances are highlighted by a *green* and *orange box*, respectively. **b–f** The reconstruction of 1, 3, 5, 7 and 9 co-additions of the HBHA(CO)NH experiment. The *traces* correspond to the indirect ^1H dimension (HBHA dimension) for the weak peak (*orange box* in a). **g–k** Same as **b–f**, but for traces corresponding to the strong peak (*green box* in a). All spectra reconstructed using maximum entropy reconstruction with automated (AUTO) parameter selection

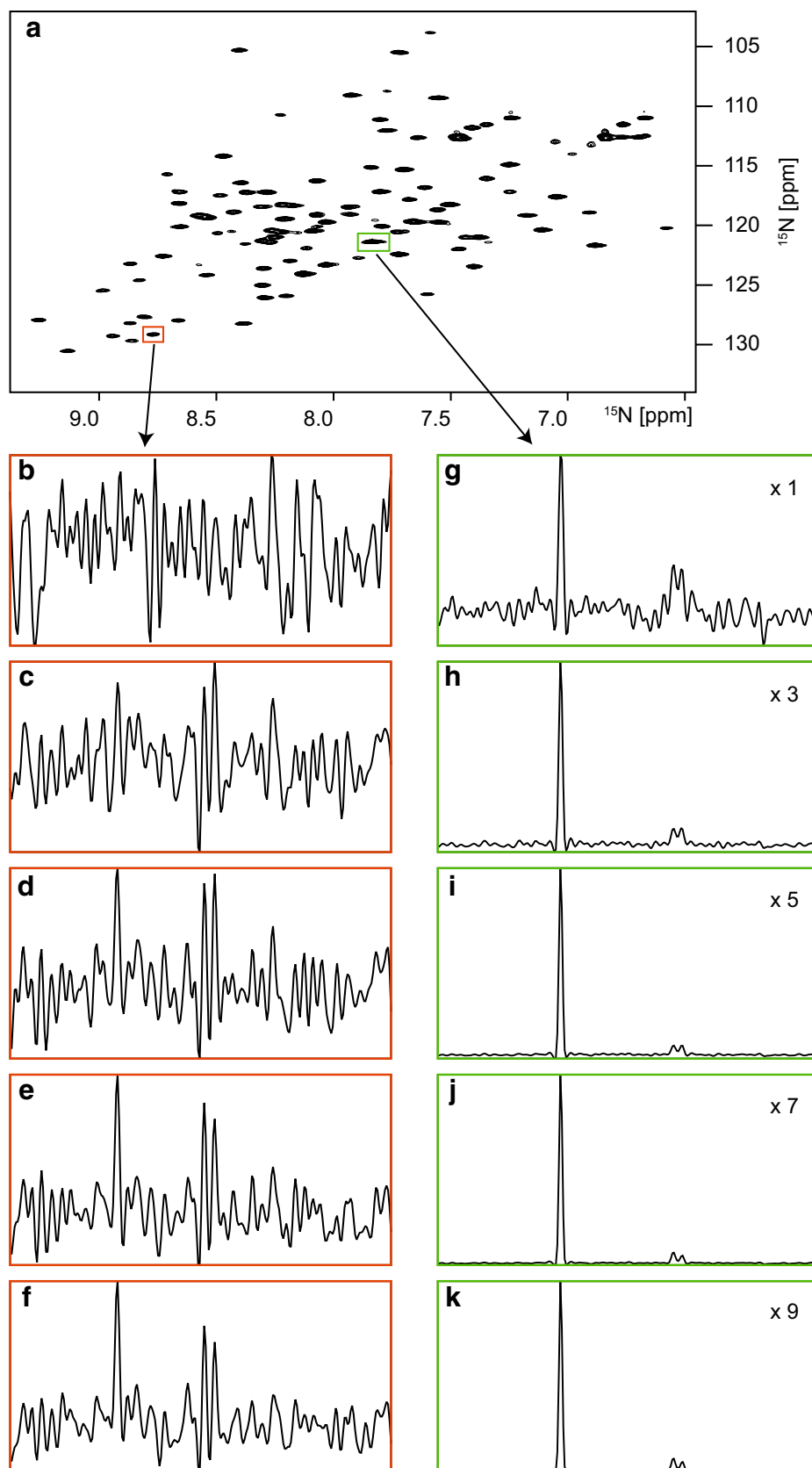
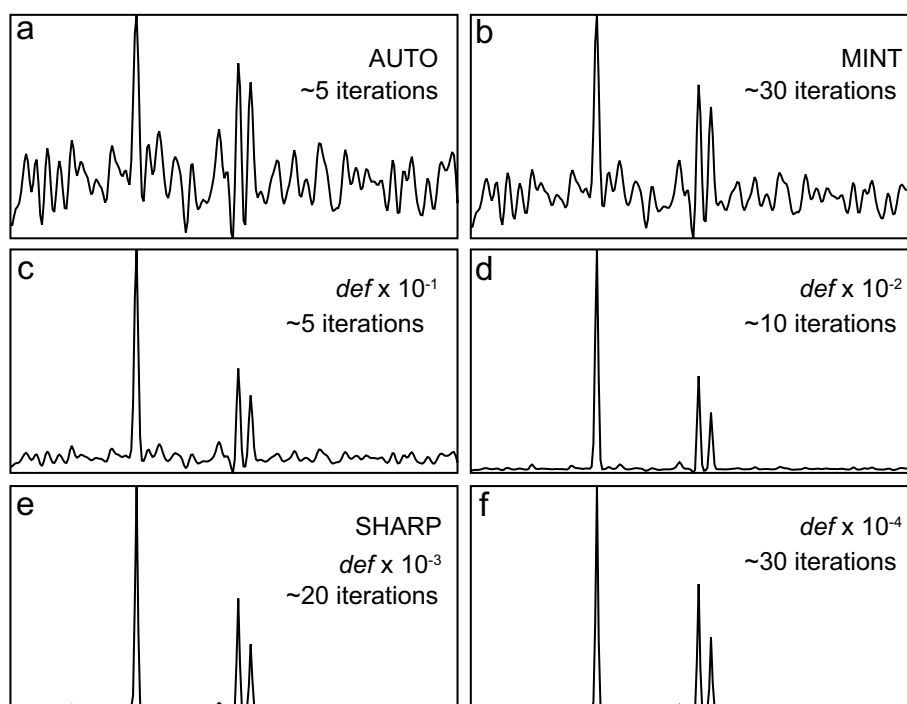


Fig. 4 The effect of parameter selection on the reconstruction. The *traces* are taken at the same coordinate as indicated in Fig. 3 **b–f**. **a** and **b** the previously reported AUTO and MINT approaches to parameter selection, both performing similarly. **c–f** The effect of scaling the value of *def* by 1–4 orders of magnitude. Scaling beyond three orders of magnitude results in no observable spectral improvements but is associated with an increase in computational time, as indicated by the number of iterations required for convergence. Scaling of *def* by 10^{-3} is defined as Shannon-weighted augmentation of reconstruction parameters (SHARP)



converge requiring very long reconstruction times overall (several hours on 24 CPUs). Where *def* is scaled by 10^{-3} the convergence of all the planes is much faster (full reconstruction in less than 1 h on 24 CPUs). It may be that in practice the value of *def* must be scaled incrementally similar to what is done with the value of λ in MINT parameter selection, however, in our experience most spectra converge rapidly with the 10^{-3} scaling of *def*, and we define this as the default scaling in SHARP.

The improvements observed, by scaling *def*, have indeed previously been noted, and argued to be a consequence of a narrowing of the dispersive imaginary components, leading to an increase in the absorptive real part of the spectrum (Daniell and Hore 1989). We note that there is some non-linear scaling in SHARP, where the smallest peak in Fig. 4e is ~30% of the height of the tallest peak in that trace, while the corresponding value for MINT is 50% (Fig. 4b). In our experience the advantages achieved by SHARP in facilitating analysis far outweigh any detrimental effects of the non-linear behaviour.

Combining NUS with SHARP parameter selection for maximum entropy reconstruction and the co-addition approach described, we have successfully been able to assign the resonances of MAL^{TIRC116A} in the structured regions of the protein. The above approach was in our experience the only way to access this data, and for very challenging systems it may indeed be the only viable approach. For the general case, the manual assessment of improvements is tedious and would require automation, but should under such circumstances provide a very

time-efficient approach to acquisition of lengthy multidimensional experiments.

Concluding remarks

Structural studies of proteins by NMR spectroscopy requires acquisition of multiple high resolution, multidimensional experiments, often requiring weeks of acquisition time. This places proteins that are poorly soluble or unstable outside the scope of traditional NMR analysis. Here, we have combined non-uniform sampling with iterative data acquisition, to minimise the time required for individual experiments. NUS allows a dataset to be acquired with sufficient resolution along all indirect dimensions in a very short timeframe (a few hours). The resultant spectrum will, for dilute samples, be of insufficient sensitivity for resonance assignment or structural characterisation. However, multiple such NUS datasets may be acquired and co-added until sufficient sensitivity is achieved for analysis. This will allow the analysis of proteins that are poorly stable even at low concentrations.

We find that for decaying signals, co-addition is particularly beneficial for increasing S/N for samples at long evolution times, whereas for short evolution times, signal averaging provides similar gains in S/N as the acquisition of additional sample points. At short evolution times, the gains in S/N achieved by signal averaging compared to the acquisition of additional sample points are far outweighed by the beneficial effect of acquiring these additional

samples, which in linear sampling serve to increase resolution and in NUS serve to reduce sampling artifacts. In contrast, at long evolution times signal averaging is the only way to increase S/N. These findings suggest that non-uniform signal averaging approaches may be best applied where longer evolution times are signal-averaged more than early time points. We, further, introduce a new automated parameter selection procedure for maximum entropy reconstruction that can significantly aid analysis of noisy spectra and implement its use in the previously described automated script generator (Mobli et al. 2007).

Finally, we note that the example shown here required a total ~50 h of acquisition time to yield sufficient sensitivity for assignment of the sidechain H α /H β atoms of MAL^{TIRC116A}. This is a significant time burden and indeed a total ~400 h (2.5 weeks) of instrument time was required to achieve resonance assignment in the structured regions of the protein. This is despite using the very time efficient approach described. In the case of MAL, the long time-frame is justified given the important role of the protein and the unique role of NMR in providing insights into the redox state and solution structure of the protein. The implications of our work for the structure and redox state of MAL^{TIRC116A} will be reported elsewhere.

Acknowledgements The authors would like to thank Dimitri Maziuk for assistance in analysis of the BMRB. This work was supported by the Australian Research Council through Discovery Grant DP140101098 to MM and a Future Fellowship to MM (FT110100925), and the National Health and Medical Research Council through a project grant to MM and BK (APP1081409). BK is an NHMRC Principal Research Fellow (APP1003325 and APP1110971).

References

- Barna J, Laue E, Mayger M, Skilling J, Worrall S (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments *J Magn Reson* (1969) 73:69–77
- Daniell GJ, Hore PJ (1989) Maximum entropy and NMR—A new approach *J Magn Reson* (1969) 84:515–536
- Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K (2002) NMR analysis of a 900 K GroEL GroES complex. *Nature* 418:207–211
- Havel T, Wüthrich K (1984) A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ¹H-¹H proximities in solution. *Bull Math Biol* 46:673–698
- Hodgkinson P, Holmes KJ, Hore PJ (1996) Selective data acquisition in NMR. The quantification of anti-phase scalar couplings. *J Magn Reson Ser A* 120:18–30
- Holland DJ, Bostock MJ, Gladden LF, Nietlispach D (2011) Fast multidimensional NMR spectroscopy using compressed sensing. *Angew Chem Int Ed* 50:6548–6551
- Hybets SG, Takeuchi K, Wagner G (2010) Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* 132:2145–2147
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of proton, carbon-13, and nitrogen-15 spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29:4659–4667
- Jaravine VA, Orekhov VY (2006) Targeted acquisition for real-time NMR spectroscopy. *J Am Chem Soc* 128:13421–13426
- Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. *Angew Chem Int Ed* 50:5556–5559
- Kazimierczuk K, Zawadzka A, Kozminski W (2008) Optimization of random time domain sampling in multidimensional NMR. *J Magn Reson* 192:123–130
- Lin Z, Lu J, Zhou W, Shen Y (2012) Structural insights into TIR domain specificity of the bridging adaptor Mal in TLR4 signaling. *PLoS ONE* 7:e34202
- Lustig M, Donoho D, Pauly JM (2007) Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med* 58:1182–1195
- Mobli M (2015) Reducing seed dependent variability of non-uniformly sampled multidimensional NMR data. *J Magn Reson* 256:60–69
- Mobli M, Hoch JC (2014) Nonuniform sampling and non-fourier signal processing methods in multidimensional NMR. *Prog Nucl Magn Reson Spectrosc* 83:21–41
- Mobli M, Stern AS, Hoch JC (2006) Spectral reconstruction methods in fast NMR: Reduced dimensionality, random sampling and maximum entropy. *J Magn Reson* 182:96–105
- Mobli M, Maciejewski MW, Gryk MR, Hoch JC (2007) An automated tool for maximum entropy reconstruction of biomolecular NMR spectra. *Nat Methods* 4:467–468
- Mobli M, Stern AS, Bermel W, King GF, Hoch JC (2010) A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. *J Magn Reson* 204:160–164
- Paramasivam S et al (2012) Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *J Phys Chem B* 116:7416–7427
- Rovnyak D, Frueh DP, Sastry M, Sun Z-YJ, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson* 170:15–21
- Rovnyak D, Sarcone M, Jiang Z (2011) Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. *Magn Reson Chem* 49:483–491
- Schuyler AD, Maciejewski MW, Arthanari H, Hoch JC (2011) Knowledge-based nonuniform sampling in multidimensional NMR. *J Biomol NMR* 50:247–262
- Snyder GA et al (2014) Crystal structures of the Toll/Interleukin-1 receptor (TIR) domains from the Brucella protein TcpB and host adaptor TIRAP reveal mechanisms of molecular mimicry. *J Biol Chem* 289:669–679
- Stern AS, Hoch JC (2002) Rowland NMR Toolkit
- Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l1-norm reconstruction. *J Magn Reson* 188:295–300
- Thomas V, Nicholas JG, Ashley M, Bostjan K, Stuart K (2012) Adaptors in Toll-Like Receptor Signaling and their Potential as Therapeutic Targets. *Curr Drug Targets* 13:1360–1374
- Valkov E et al (2011) Crystal structure of Toll-like receptor adaptor MAL/TIRAP reveals the molecular basis for signal transduction and disease protection. *Proc Natl Acad Sci USA* 108:14879–14884
- Woo JR, Kim S, Shoelson SE, Park S (2012) X-ray crystallographic structure of TIR-domain from the human TIR-domain containing adaptor protein/MyD88-adaptor-Like protein (TIRAP/MAL). *Bull Korean Chem Soc* 33:3091–3094
- Worley B (2016) Subrandom methods for multidimensional nonuniform sampling. *J Magn Reson* 269:128–137