

## Non-uniform sampling in biomolecular NMR

Martin Billeter<sup>1</sup> 

Received: 4 May 2017 / Accepted: 11 May 2017 / Published online: 15 June 2017  
© Springer Science+Business Media Dordrecht 2017

Proteins are nature's tools for everything; their remarkable diversity of biological roles is tightly related to their amazing structural variability, with optimized combinations of stability and flexibility, the presence of minor forms, or underlying mechanisms of folding. NMR is a highly suited analysis tool for many protein characteristics, include properties of structure, dynamics and interaction. Nonetheless, lack of defined 3D structure as in partly or fully intrinsically disordered proteins, lifetimes on the order of the lengths of NMR experiments or shorter, as well as the presence of multiple states, where a minor state may well be the most interesting one, pose complex and contradicting challenges to the design of NMR experiments and in turn to the processing of the measured raw data.

Three entities are central to the design of NMR experiments. (1) Spectral width is related to the repetition rate of individual measurements; with uniform sampling, the dwell time should correspond to the inverse of twice the spectral width, which for indirect dimensions strongly affects the total experiment time. (2) Resolution depends on the maximal evolution times, which again represent time-determining parameters for the experiment. (3) Sensitivity is naturally related to the number of observations, and too extensive reduction of the number of measurements may compromise signal detection. Whereas time is not a problem for one-dimensional NMR studies, experiment time for multi-dimensional, uniformly sampled spectra is proportional to  $m^{n-1}$ , where  $m$  is the number of points recorded along each of the  $n-1$  indirect dimensions (in this text,  $m$

is for simplicity assumed to be identical in all dimensions). For four dimensions, and with 64 (complex) points along each indirect dimension, this factor exceeds  $2 \times 10^6$ . Recording of such data sets would result in excessive experiment times.

However, the number of individual observations defined by the above factor is in sharp contrast to the number of parameters that one hopes to extract from such a spectrum. Consider a 4D HCC(CO)NH-TOCSY (with resonances observed for HN, N, aliphatic carbons and their hydrogens) recorded for a 200 residue protein. Let's assume that there are seven relevant resonances per residue (i.e., seven observable nuclei as in AMX spin systems). Each resonance can be described by two numbers: chemical shift and line width. The number of parameters to extract is thus 2800, which could be determined in an ideal case by the same number of measurements (in this context: number of FIDs times number of direct points). Therefore, assuming acceptable sensitivity, the above factor of  $2 \times 10^6$  can often be greatly reduced, while still acquiring sufficient data for an overdetermined system. (For example, for a 4D spectrum with 64 complex points along all indirect dimensions, the 13 projections with angles to the spectral axes of  $0^\circ$  or  $45^\circ$  requires recording of 1664 FIDs; this is sufficient for side chain assignments of very small proteins. Note that each point in the direct dimension of an FID represents a measurement.) This time reduction comes with a concomitant lowering of memory demands, which may be highly necessary in the case of 5D spectra.

The contradicting requests on resolution and experiment duration (while not compromising on spectral widths) can be met by reducing the number of FID acquisitions, randomly or with a selected strategy. Data size can be kept in acceptable limits by focusing on interesting regions. The above 4D TOCSY, if stored in full, may contain several

✉ Martin Billeter  
martin.billeter@chem.gu.se

<sup>1</sup> Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

100 million numbers. It contains fewer than 1000 signals, which therefore are very sparsely distributed over the spectrum. For a 5D experiment, a further dramatic increase of the data size happens, together with even higher levels of sparsity. Focusing on subspaces (e.g., on 2D planes) that do contain signals may significantly reduce the overall data size. However, reconstruction of the full spectrum reintroduces these data size problems; thus, one should consider also in this situation only subspaces—or completely avoid reconstructions.

So called “fast” NMR approaches can be classified by experiment design or by processing concepts. NUS-type (non-uniform sampling) acquisition typically consists of randomly selecting acquisition points in the indirect dimensions. The “random” aspect is often restricted by applying a certain distribution type (e.g. fully random, exponential, Poisson distribution ...); additional minor restrictions may be applied, such as each acquisition time in each indirect dimension (on a preselected regular grid) should be chosen at least once. Non-random selection of reduced acquisition, for example projections, may be associated with very low sampling factors (often well below 1% for 4D spectra). Approaches for processing of the recorded data can be classified in several ways. The sparse time-domain data may be directly transformed using various forms of Fourier-transform algorithms, or the missing data may first added by interpolation, followed by conventional Fourier-transform. A different distinction concerns the intermediate reconstruction of the spectra followed by their conventional analysis (i.e., peak picking etc.), or alternatively by extracting the relevant data (shifts, linewidths, peak volumes) directly from the raw data.

The present special issue on “Non-uniform sampling in biomolecular NMR” addresses a wide number of aspects, illustrating the still ongoing discussion about what is the “best” method. New implementations of processing tools are presented (Ying et al. 2017). Choices of NUS-type acquisition are addressed: What kind of sampling should be chosen, e.g. in terms of sampling extent, distribution of points, maximal acquisition times, non-uniform selection within quadrature detection (e.g., Bostock et al. 2017)? Various post-processing procedures are discussed: How can NUS-related noise be characterized (and eliminated), and how should the raw data be processed, for example regarding extrapolation (Hyberts et al. 2017)? What assumptions

on sparsity are warranted, and what problems are connected to reconstructions (Shchukina et al. 2017)? Other topics covered are: How can one best extract information from very short lived samples (Miljenović et al. 2017)? How does one focus on relevant information in 5D data sets (Kosiński et al. 2017)?

Finally, one contribution introduces novel aspects of NUS in NMR (Urbańczyk et al. 2017). Most NUS usage focuses on data for assignments and structures, where all dimensions in the full (or reconstructed) spectra correspond to frequency axes. In studies of dynamics often a stack of low dimensional spectra is assembled, where the “stack axis” reports relaxation times. This stack can be considered a single,  $n + 1$  dimensional data set, with  $n$  frequency axes and one relaxation time axis, and treated with NUS along all axes (except for the direct acquisition dimension). Can this type of NUS application be generalized to data sets with axes that, besides frequencies, describe intermolecular interactions, e.g. in drug screening projects, or the time course of chemical reactions?

## References

- Bostock MJ, Holland DJ, Nietlispach D (2017) Improving resolution in multidimensional NMR using random quadrature detection with compressed sensing reconstruction. *J Biomol NMR*. doi:[10.1007/s10858-016-0062-9](https://doi.org/10.1007/s10858-016-0062-9)
- Hyberts SG, Robson SA, Wagner G (2017) Interpolating and extrapolating with hmsIST—seeking a tmax for optimal sensitivity, resolution and frequency accuracy. *J Biomol NMR*. doi:[10.1007/s10858-017-0103-z](https://doi.org/10.1007/s10858-017-0103-z)
- Kosiński K, Stanek J, Górka MJ, Żerko S, Koźmiński W (2017) Reconstruction of non-uniformly sampled five-dimensional NMR spectra by signal separation algorithm. *J Biomol NMR*. doi:[10.1007/s10858-017-0095-8](https://doi.org/10.1007/s10858-017-0095-8)
- Miljenović T, Jia X, Lavrencic P, Kobe B, Mobli M (2017) A non-uniform sampling approach enables studies of dilute and unstable proteins. *J Biomol NMR*. doi:[10.1007/s10858-017-0091-z](https://doi.org/10.1007/s10858-017-0091-z)
- Shchukina A, Kasprzak P, Dass R, Nowakowski M, Kazimierczuk K (2017) Pitfalls in compressed sensing reconstruction and how to avoid them. *J Biomol NMR*. doi:[10.1007/s10858-016-0068-3](https://doi.org/10.1007/s10858-016-0068-3)
- Urbańczyk M, Nowakowski M, Koźmiński W, Kazimierczuk K (2017) Joint non-uniform sampling of all incremented time delays for quicker acquisition in protein relaxation studies. *J Biomol NMR*. doi:[10.1007/s10858-017-0115-8](https://doi.org/10.1007/s10858-017-0115-8)
- Ying J, Delaglio F, Torchia DA, Bax A (2017) Multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR*. doi:[10.1007/s10858-016-0072-7](https://doi.org/10.1007/s10858-016-0072-7)