Biochemical and Biophysical Research Communications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Evaluation of the reliability of the maximum entropy method for reconstructing 3D and 4D NOESY-type NMR spectra of proteins

Yoshiki Shigemitsu [a], Teppei Ikeya [a,b], Akihiro Yamamoto [a], Yuusuke Tsuchie [a], Masaki Mishima [a,b], Brian O. Smith [c], Peter Güntert [a,d,e], Yutaka Ito [a,b,*]

[a] Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo 192-0397, Japan
[b] CREST/Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
[c] Institute of Molecular, Cell and Systems Biology & School of Life Sciences, College of Medical, Veterinary & Life Sciences, Joseph Black Building, University of Glasgow, Glasgow G12 8QQ, UK
[d] Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
[e] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 16 December 2014
Available online xxxx

Keywords:
Protein NMR
4D NMR
NOESY
Spectrum reconstruction
Maximum entropy
Nonlinear sampling

ABSTRACT

Despite their advantages in analysis, 4D NMR experiments are still infrequently used as a routine tool in protein NMR projects due to the long duration of the measurement and limited digital resolution. Recently, new acquisition techniques for speeding up multidimensional NMR experiments, such as nonlinear sampling, in combination with non-Fourier transform data processing methods have been proposed to be beneficial for 4D NMR experiments. Maximum entropy (MaxEnt) methods have been utilised for reconstructing nonlinearly sampled multi-dimensional NMR data. However, the artefacts arising from MaxEnt processing, particularly, in NOESY spectra have not yet been clearly assessed in comparison with other methods, such as quantitative maximum entropy, multidimensional decomposition, and compressed sensing.

We compared MaxEnt with other methods in reconstructing 3D NOESY data acquired with variously reduced sparse sampling schedules and found that MaxEnt is robust, quick and competitive with other methods. Next, nonlinear sampling and MaxEnt processing were applied to 4D NOESY experiments, and the effect of the artefacts of MaxEnt was evaluated by calculating 3D structures from the NOE-derived distance restraints. Our results demonstrated that sufficiently converged and accurate structures (RMSD of 0.91 Å to the mean and 1.36 Å to the reference structures) were obtained even with NOESY spectra reconstructed from 1.6% randomly selected sampling points for indirect dimensions. This suggests that 3D MaxEnt processing in combination with nonlinear sampling schedules is still a useful and advantageous option for rapid acquisition of high-resolution 4D NOESY spectra of proteins.

1. Introduction

For the structure determination of small to medium size proteins by solution NMR, 3D NOESY experiments are usually analysed for the collection of NOE-derived distance restraints, even though assignment ambiguity due to degeneracy of ^{1}H resonances remains for one of the ^{1}H dimensions. Further separation of the ambiguous ^{1}H dimension with the chemical shifts of directly bound ^{13}C or ^{15}N nuclei in 4D spectra is a straightforward way to resolve the degeneracy. However, 4D NOESY experiments are less commonly used due to the long duration of the measurement and limited digital resolution in indirectly observed dimensions because generally insufficient data points are acquired in order to keep the measurement time manageable. It would therefore be advantageous to be able to measure 4D NOESY spectra with good digital resolution in affordable measurement time.

NMR spectroscopy is an inherently insensitive technique, thus new acquisition schemes for speeding up multidimensional NMR experiments are demanded for dramatic improvements in both sensitivity and resolution. Among the various approaches, nonlinear

Abbreviations: CS, compressed sensing; FT, Fourier transform; IRLS, iteratively reweighted least-squares; IST, iterative soft thresholding; MaxEnt, maximum entropy; MDD, multi-dimensional decomposition; QME, quantitative maximum entropy; RMSE, root mean square error; RMSD, root mean square deviation.

* Corresponding author at: Department of Chemistry, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo 192-0397, Japan. Fax: +81 42 677 2525.

E-mail address: ito-yutaka@tmu.ac.jp (Y. Ito).

http://dx.doi.org/10.1016/j.bbrc.2014.12.088
0006-291X/

Please cite this article in press as: Y. Shigemitsu et al., Evaluation of the reliability of the maximum entropy method for reconstructing 3D and 4D NOESY-type NMR spectra of proteins, Biochem. Biophys. Res. Commun. (2015), http://dx.doi.org/10.1016/j.bbrc.2014.12.088

sampling for indirectly acquired dimensions (also called non-uniform sampling or sparse sampling) [1–3] has been shown to be a robust technique. The effect of reduced sampling schemes on structure determination has also been assessed [4].

Since discrete Fourier transform (FT) cannot be used for processing sparsely sampled data, maximum entropy (MaxEnt) [5,6] has been used as an alternative for more than 20 years. Recently, multi-dimensional decomposition (MDD) [7], non-uniform Fourier transform [8,9] and forward maximum-entropy reconstruction [10,11] have been proposed. More recently, l_p-norm ($0 < p \leqslant 1$) minimisation referred to as compressed sensing (CS) was introduced to the NMR field [12,13]. We have reported an extended version of MaxEnt based on the MemSys5 package [14], quantitative maximum entropy (QME) [15]. For the reconstruction of 4D NMR spectra, MDD [16], CLEAN [17,18], MaxEnt [19,20], and iterative soft thresholding (IST) [21] have been applied.

One of the major criticisms to non-FT methods is their questionable reliability in reproducing cross peaks with proper signal intensity, especially in the case of signals with a wide dynamic range as in NOESY-type experiments. However, it has not yet been clearly compared the quality of "classical" MaxEnt processing in NOESY spectra with that of other methods.

In this report, we applied MaxEnt processing to 3D ^{15}N-separated and ^{13}C-separated NOESY of a small protein, the *Thermus thermophilus* HB8 TTHA1718 gene product, and compared its reliability in reproducing accurate signal intensity from nonlinearly sampled data with the alternative approaches MDD, CS, and QME. In addition, we employed a nonlinear sampling scheme in 4D ^{13}C/^{15}N-separated and ^{13}C/^{13}C-separated NOESY of TTHA1718 and assessed the quality of MaxEnt processing on these 4D NOESY data by calculating 3D structures from the NOE-derived distance restraints obtained from the reconstructed spectra.

2. Materials and methods

2.1. Sample preparation and NMR spectroscopy

The expression and purification of ^{13}C/^{15}N-labelled TTHA1718 were performed as described previously [22]. The final ^{13}C/^{15}N-TTHA1718 fractions were concentrated to approximately 1.0 mM and dissolved in M9 medium containing 10% D_2O for NMR lock.

NMR experiments were performed at 37 °C probe temperature in a triple-resonance cryoprobe fitted with a *z*-axis pulsed field gradient coil, using a Bruker Avance 600 MHz spectrometer. The 3D ^{15}N-separated and ^{13}C-separated NOESY experiments were measured with 80 ms NOE mixing time and a total of 512 (t_3, ^{1}H acquisition) × 128 (t_1, ^{1}H) × 32 (t_2, ^{15}N or ^{13}C) complex points. The 4D ^{13}C/^{15}N-separated and ^{13}C/^{13}C-separated NOESY experiments were measured with 8 transients, a 200 ms NOE mixing time and a total of 512 (t_4, ^{1}H^N acquisition) × 32 (t_1, ^{1}H) × 24 (t_2, ^{13}C) × 8 (t_3, ^{15}N) and 512 (t_4, ^{1}H acquisition) × 24 (t_1, ^{1}H) × 20 (t_2 ^{13}C) × 18 (t_3 ^{13}C) complex points, respectively. The total measurement times for the 4D ^{13}C/^{15}N-separated and 4D ^{13}C/^{13}C-separated NOESY experiments were 5.7 and 6.4 days, respectively. These 3D and 4D NOESY data are henceforth referred to as "reference" data.

In order to achieve nonlinear sampling, the pulse sequences were modified according to the procedure reported by Rovnyak et al. [3]. The pulse programs with conventional and nonlinear sampling and the VC list generator program are available from the corresponding author.

2.2. Preparation of various data sets with conventional and nonlinear sampling

For the evaluation of the artefacts arising from the employment of nonlinear sampling and MaxEnt processing, data sets with various randomly sampled points in the indirect dimensions were prepared from the reference 3D and 4D NOESY data. For the nonlinearly sampled data, sampling schemes were generated using an in-house program. Six steps for random reduction of sampling points, 1/2, 1/4, 1/8, 1/16, 1/32 and 1/64, were generated for 3D NOESY experiments, while seven steps of sampling points, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64 and 1/128 were generated for 4D NOESY experiments. In order to assess the deviation due to the selected sampling points, three different sampling schedules were generated from different random seeds for each random reduction step. Next, new data sets were concatenated by rearranging the raw data based on the schedules. These 3D and 4D NOESY data are henceforth referred to as "nonlinearly sampled" data.

For comparison, conventionally (linearly) sampled 4D ^{13}C/^{15}N-separated and 4D ^{13}C/^{13}C-separated NOESY data sets with reduced numbers of data points were also prepared corresponding to approximately 1/2, 1/4, 1/8, 1/16, 1/32 and 1/128 of the reference data sets. These 4D NOESY data are henceforth referred to as "linearly sampled" data. The parameters, e.g. total number of data points for all indirect dimensions, are described in Supplementary Table S1.

2.3. Data processing and spectral analysis

The reference, nonlinearly sampled and linearly sampled 3D and 4D NOESY data were processed with 2D and 3D MaxEnt, respectively, on LINUX-PCs using the AZARA 2.7/2.8 software suite (W. Boucher, http://www.bio.cam.ac.uk/azara/). Consequently, 3D NOESY spectra were produced with 512 (F_3, ^{1}H) × 512 (F_1, ^{1}H) × 128 (F_2, ^{13}C or ^{15}N) data points, and 4D ^{13}C/^{15}N-separated NOESY and 4D ^{13}C/^{13}C-separated NOESY spectra were produced with 512 (F_4, ^{1}H^N) × 128 (F_1, ^{1}H) × 128 (F_2, ^{13}C) × 64 (F_3, ^{15}N), and 480 (F_4, ^{1}H) × 128 (F_1, ^{1}H) × 128 (F_2, ^{13}C) × 128 (F_3, ^{13}C) data points, respectively. The duration of 3D MaxEnt processing depends upon the number of iterations, the sizes of both input and output data, etc. Typically the processing took 10–20 min for 3D ^{13}C/^{15}N-separated NOESY and 6–16 h for 4D ^{13}C/^{13}C-separated NOESY data using a LINUX-PC with a 3.4 GHz Intel Core i7-4770 CPU. In addition, the "reference" 3D and 4D spectra were also processed with conventional FT on Azara for all dimensions. For 4D Fourier transform, linear prediction was utilised for the indirect dimensions.

In order to assess the quality of MaxEnt-processed 3D NOESY data, MDD, CS [with IST and iteratively reweighted least-squares (IRLS) algorithms], and QME processing were employed for comparison.

The MDD and CS processings were performed by the MDDNMR software [23] on the nmrPipe [24] platform. After processing the directly acquired dimension (t_3) by FT using nmrPipe, MDD and CS calculations were performed by employing the standard parameters used in the example scripts of the software, and 3D interferograms were reconstructed with 256 (t_1) × 64 (t_2) complex points for the indirect dimensions. The indirect dimensions were then apodised, zero-filled (×2) and processed with FT.

The QME processing was performed by a C-language program after processing the directly acquired dimension (t_3) by FT using Azara.

All spectra were visualised and analysed on LINUX-PCs with the combination of customised macro programs on the OpenGL-version of ANSIG 3.3 software [25,26] and the CcpNmr Analysis 2.2.2 software [27]. Peak positions were identified using the automated peak picking algorithm of Azara.

2.4. Structure calculation

The structure calculations were performed with the program CYANA [28] version 3.0 using automated NOE assignment [29] and torsion angle dynamics for the structure calculation [30].

Please cite this article in press as: Y. Shigemitsu et al., Evaluation of the reliability of the maximum entropy method for reconstructing 3D and 4D NOESY-type NMR spectra of proteins, Biochem. Biophys. Res. Commun. (2015), http://dx.doi.org/10.1016/j.bbrc.2014.12.088