

G-Matrix Fourier Transform NOESY-Based Protocol for High-Quality Protein Structure Determination

Yang Shen, Hanudatta S. Atreya, Gaohua Liu, and Thomas Szyperski*

Contribution from the Departments of Chemistry and Structural Biology, The State University of New York at Buffalo, Buffalo, New York 14260

Received January 11, 2005; E-mail: szypersk@chem.buffalo.edu

Abstract: A protocol for high-quality structure determination based on G-matrix Fourier transform (GFT) NMR is presented. Five through-bond chemical shift correlation experiments providing 4D and 5D spectral information at high digital resolution are performed for resonance assignment. These are combined with a newly implemented (4,3)D GFT NOESY experiment which encodes information of 4D $^{15}\text{N}/^{15}\text{N}$ -, $^{13}\text{C}_{\text{aliphatic}}/^{15}\text{N}$ -, and $^{13}\text{C}_{\text{aliphatic}}/^{13}\text{C}_{\text{aliphatic}}$ -resolved $^1\text{H}, ^1\text{H}$ -NOESY in two subspectra, each containing one component of chemical shift doublets arising from 4D \rightarrow 3D projection at $\omega_1:\Omega(^1\text{H}) \pm \Omega(\text{X})$ [$\text{X} = ^{15}\text{N}, ^{13}\text{C}_{\text{aliphatic}}$]. The peaks located at the centers of the doublets are obtained from simultaneous 3D $^{15}\text{N}/^{13}\text{C}_{\text{aliphatic}}/^{13}\text{C}_{\text{aromatic}}$ -resolved $^1\text{H}, ^1\text{H}$ -NOESY, wherein NOEs detected on aromatic protons are also obtained. The protocol was applied for determining a high-quality structure of the 14 kDa Northeast Structural Genomics consortium target protein, YqfB (PDB ID 1TE7). Through-bond correlation and NOESY spectra were acquired, respectively, in 16.9 and 39 h (30 h for shift doublets, 9 h for central peaks) on a 600 MHz spectrometer equipped with a cryogenic probe. The rapidly collected highly resolved 4D NOESY information allows one to assign the majority of NOEs directly from chemical shifts, which yields accurate initial structures “within” ~ 2 Å of the final structure. Information theoretical “QUEEN” analysis of initial distance limit constraint networks revealed that, in contrast to structure-based protocols, such NOE assignment is not biased toward identifying additional constraints that tend to be redundant with respect to the available constraint network. The protocol enables rapid NMR data collection for robust high-quality structure determination of proteins up to ~ 20 – 25 kDa in high-throughput.

Introduction

Efficient NMR-based protein structure determination¹ relies on measurement of nuclear Overhauser effects (NOEs), which yield ^1H – ^1H upper distance limit constraints. The assignment of NOEs quite generally depends on having (nearly) complete resonance assignments.^{1,2} However, due to degeneracy of chemical shifts, the NOE assignment remains a nontrivial task even when complete resonance assignments are available. Nowadays, two approaches are routinely used to solve this “NOE assignment problem”. First, proteins are $^{15}\text{N}/^{13}\text{C}$ double labeled³ so that NOEs can be measured in 3D ^{15}N - or ^{13}C -resolved $^1\text{H}, ^1\text{H}$ -NOE spectroscopy (NOESY).² Dispersing NOE signals in a third dimension, which encodes a ^{13}C or a ^{15}N shift, typically allows one to assign for medium-sized proteins ~ 15 – 25% of the NOEs directly based on chemical shift data (compared to only few percent in 2D $^1\text{H}, ^1\text{H}$ -NOESY^{1,2}). Second, an *initial* structure is calculated which is used in conjunction with the chemical shifts to assign additional NOEs. Several such cycles of structure calculation and NOE assignment are usually performed iteratively until a refined structure is obtained.

Importantly, inaccuracies in the initial fold arising from incorrectly assigned NOEs may result in the mis-assignment of additional NOEs. Hence, proper convergence of the NMR structure determination depends on obtaining an appropriately accurate initial structure; that is, it is advantageous if the bundle of conformers representing the initial solution structure covers a conformational subspace which overlaps with that of the refined ensemble of conformers. This requirement constitutes a key challenge for reliable automated NOE assignment⁴ and thus also for the development of a robust and scalable platform for high-throughput structure determination in structural genomics.⁵ Several programs have been established to automatically obtain accurate initial folds.⁴ Among those are AutoStructure⁶ and CYANA,⁷ both of which are widely used. Conceptually,

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(2) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy*; Academic Press: San Diego, CA, 1996.
(3) (a) Kainosho, M. *Nat. Struct. Biol.* **1997**, *4*, 858–861. (b) Acton, T. B. et al. *Methods Enzymol.* **2005**, *394*, 210–243.

(4) (a) Güntert, P. *Prog. NMR Spectrosc.* **2003**, *43*, 105–125. (b) Baran, M. C.; Huang, Y. J.; Moseley, H. N. B.; Montelione, G. T. *Chem. Rev.* **2004**, *104*, 3451–3555. (c) Huang, Y. J.; Moseley, H.; Baran, M. C.; Arrowsmith, C. H.; Powers, R.; Tejero, R.; Szyperski, T.; Montelione, G. T. *Methods Enzymol.* **2005**, *394*, 111–141.
(5) (a) Montelione, G. T.; Zheng, D.; Huang, Y.; Gunsalus, C.; Szyperski, T. *Nat. Struct. Biol.* **2000**, *7*, 982–984. (b) Yee A. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1825–1830.
(6) (a) Moseley, H. N. B.; Monleon, D.; Montelione, G. T. *Methods Enzymol.* **2001**, *339*, 91–108. (b) Huang, Y. J.; Swapna, G. V.; Rajan, P. K.; Ke, H.; Xia, B.; Shukla, K.; Inouye, M.; Montelione, G. T. *J. Mol. Biol.* **2003**, *327*, 521–536. (c) Huang, Y. J.; Powers, R.; Montelione, G. T. *J. Am. Chem. Soc.* **2005**, *127*, 1665–1674.
(7) (a) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–298. (b) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209–227. (c) Güntert, P. *Methods Mol. Biol.* **2004**, *278*, 347–372.