G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment

Hanudatta S. Atreya and Thomas Szyperski*

Departments of Chemistry and Structural Biology, State University of New York, Northeast Structural Genomics Consortium, Buffalo, NY 14260

Communicated by Herbert Hauptman, Hauptman–Woodward Medical Research Institute, Buffalo, NY, May 18, 2004 (received for review December 14, 2003)

A G-matrix Fourier transform (GFT) NMR spectroscopy-based strategy for resonance assignment of proteins is described. Each of the GFT NMR experiments presented here rapidly affords four-, five-, or six-dimensional spectral information in combination with precise measurements of chemical shifts. The resulting high information content enables one to obtain nearly complete assignments by using only four NMR experiments. For the backbone amide proton detected ''out-and-back'' experiments, data collection was further accelerated up to ≈2.5-fold by use of longitudinal ^{1}H relaxation optimization. The GFT NMR experiments were acquired for three proteins with molecular masses ranging from 8.6 to 17 kDa, demonstrating that the proposed strategy is of key interest for automated resonance assignment in structural genomics.

In an era of "big science," efforts to establish NMR-based structural genomics (1–3) paralleled with improvements in spectrometer sensitivity (e.g., ref. 4), fast collection of multidimensional spectra has emerged as a subject of broader scientific interest in structural biology and pharmacology (5–8). Recently, we introduced G-matrix Fourier transform (GFT) NMR spectroscopy to meet this demand (9, 10).

GFT NMR allows one to acquire multidimensional FT NMR spectral information rapidly, thus avoiding "sampling limitations" (3) without compromising on the precision of chemical-shift measurements. Sampling limitations arise in higher-dimensional FT NMR because measurement times increase steeply with the number of spectral dimensions: typical 2D, 3D, and 4D spectra can be acquired within minutes, hours, or days, respectively, whereas recording 5D and 6D spectra would take too long to be feasible. As a result, the signal-to-noise ratios registered in higher-dimensional FT NMR may be exceedingly large; that is, instrument time is "wasted" to sample indirect dimensions. This is aggravated further when protein structures are determined, because this requires recording of several multidimensional spectra. Moreover, high dimensionality in FT NMR is generally associated with low spectral resolution in the indirect dimensions, which severely limits the precision of the chemical shift measurements and hampers automated data analysis (11). GFT NMR affords increased precision for shift measurements, thus enabling both fast and precise acquisition of high-dimensional information. This opens opportunities to establish rapid and automated protein structure determination (11) and accurately investigate dynamic phenomena with unprecedented time resolution.

GFT NMR is based on phase-sensitive joint sampling of several indirect dimensions of a multidimensional FT NMR experiment (9). Therefore, the dimensionality of an *N*D experiment can be reduced to $N - K$ by sampling of $K + 1$ chemical shifts in a single "GFT dimension." The components of the resulting chemical shift multiplets (9) are separated into different spectra through G-matrix transformation, resulting in $2^{K+1} - 1$ $(N - K)$D FT NMR spectra. These constitute an $(N, N - K)$D GFT experiment providing the same information as the *N*D experiment. The overdetermination associated with $2^{K+1} - 1$ peaks encoding linear combinations of the $K + 1$ chemical shifts warrants the increased precision of the shift measurements (9, 10).

Nearly complete resonance assignments are generally considered a necessity for NMR-based protein structure determination (e.g., refs. 11 and 12). Here we describe a strategy for complete protein resonance assignment based on GFT NMR experiments affording accurate 4D, 5D, and 6D spectral information. Applications are presented for proteins with molecular masses ranging from 8.6 to 17 kDa.

Materials and Methods

NMR Spectrometer and Protein Samples. All measurements were performed at 25°C on Varian INOVA 600 and 750 MHz spectrometers, equipped with conventional $^1H/^{13}C/^{15}N$ triple-resonance probes, by using ≈1 mM solutions in 95% H_2O/5% ^2H_2O [20 mM 2-(*N*-morpholino)-ethanesulfonic acid/100 mM NaCl/10 mM DTT/5 mM $CaCl_2$/0.02% NaN_3, pH 6.5] of two protein samples of the Northeast Structural Genomics Consortium pipeline, i.e., the proteins encoded in *Escherichia coli* gene *YgdK* (17 kDa; isotropic rotational correlation time measured as described (3), $\tau_{rot} \approx 8.5$ ns; Northeast Structural Genomics code, "ER75") and *Pyrococcus furiosus* gene *PF0455* (13 kDa; $\tau_{rot} \approx 8$ ns; "PfR13"), as well as a 2 mM solution of ubiquitin (8.6 kDa; $\tau_{rot} \approx 4.5$ ns) in 95% H_2O/5% ^2H_2O (50 mM K-PO_4, pH 5.8).

GFT NMR Experiments. Three groups of GFT NMR experiments were implemented as described in the following (underlined letters indicate nuclei for which the chemical shifts are jointly sampled in the GFT dimension).

Group I. (4,3)D $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN/HNN(CO)$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$, (5,3)D $\underline{H}^{\alpha\beta}\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN, and (6,3)D $\underline{H}^{\alpha\beta}\underline{C}^{\alpha\beta}\underline{C}^{\alpha}\underline{CO}$NHN sequentially correlate the chemical shifts of C'–C^{α}H–C^{β}H moieties of residue $i - 1$ and the NH group of residue i.

Group II. (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$ provides intraresidue correlations of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^1H^N$ shifts of residue i.

Group III. (5,3)D $\underline{HCC}$-CH and (4,2)D $\underline{HCC}$H correlate two proton and two carbon shifts. Combinations of experiments selected from groups I–III allow one to obtain nearly complete protein resonance assignments.

$(N, N - K)$D GFT experiments yield 2^K "basic" spectra delineating the linear combinations of $K + 1$ chemical shifts (Table 1). The measurement of combinations involving only $K, K - 1, \ldots, 1$ shifts has been named "central peak detection" (9, 13) and is generally required to retain the full information of the parent *N*D spectrum. To facilitate data analysis, experiments presented here are designed so as to provide matching peak patterns along the GFT dimension of basic and/or central peak spectra of different experiments (Table 1).

Radio-Frequency Pulse Sequence Design. The radio-frequency pulse schemes for the GFT NMR experiments are shown in Figs. 7–11, which are published as supporting information on the PNAS web site. Except for $C^{\alpha\beta}$ shift evolution in HNN(CO)$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$/

Abbreviations: FT, Fourier transform; GFT, G-matrix FT.

*To whom correspondence should be addressed. E-mail: szypersk@chem.buffalo.edu.

© 2004 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0403529101

PNAS PNAS PNAS PNAS