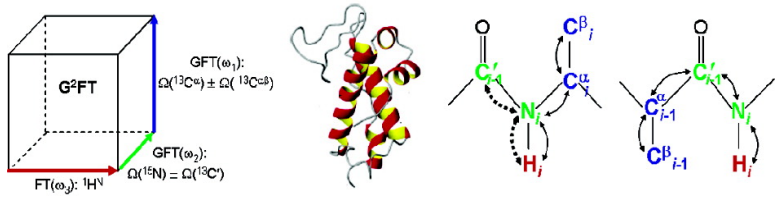
Communication

Resonance Assignment of Proteins with High Shift Degeneracy Based on 5D Spectral Information Encoded in GFT NMR Experiments

Hanudatta S. Atreya, Alexander Eletsky, and Thomas Szyperski

J. Am. Chem. Soc., **2005**, 127 (13), 4554-4555• DOI: 10.1021/ja042562e • Publication Date (Web): 12 March 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 6 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 03/12/2005

Resonance Assignment of Proteins with High Shift Degeneracy Based on 5D Spectral Information Encoded in G²FT NMR Experiments

Hanudatta S. Atreya, Alexander Eletsky, and Thomas Szyperski*

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260

Received December 10, 2004; E-mail: szypersk@chem.buffalo.edu

NMR assignments of proteins are obtained by combining several multidimensional experiments.[1] $^{15}N,^{1}H^{N}$-resolved triple resonance experiments sequentially linking $^{13}C^{\alpha/\beta}$ and/or $^{1}H^{\alpha}$ shifts are the most widely used.[1] For (partly) unfolded or α-helical (membrane) proteins, spectral analysis is, however, impeded by very high shift degeneracy, so that novel methodology for their efficient assignment is required. $^{15}N,^{1}H^{N}$ degeneracy can be largely removed in $^{13}C'_{i-1},^{15}N_i,^{1}H^{N}_i$-resolved experiments[2] (i is a residue number). When using correlated $^{13}C^{\alpha/\beta}$ or $^{13}C^{\alpha}/^{1}H^{\alpha}$ shifts to establish connectivities, conventional NMR[1] would require the recording of 5D spectra. Measurement times for such spectra are prohibitively long or lead to "sampling-limited" data collection.[3] G-matrix FT (GFT) NMR,[4] rooted in reduced-dimensionality NMR[5] and related to accordion spectroscopy,[6a,b] can (i) rapidly provide precise high-dimensional spectral information and (ii) serve to reconstruct higher-dimensional spectra.[6c,d] Previously published (4,3)D GFT experiments[4c] have greatly increased the speed of NMR structure determination, but are not optimally tailored for proteins with very high shift degeneracy.

Here, we present novel "G²FT NMR experiments" in which two G-matrix transformations are applied. This allows one to jointly sample shifts solely serving to provide increased resolution *separately* from those also providing sequential connectivities. As a result, one obtains data sets in which spin system identification can be based on (3,2)D GFT NMR in the first GFT dimension, while the previously described peak patterns[4c] for sequential assignment are retained in the second GFT dimension. First, $^{13}C'_{i-1},^{15}N_i,$ $^{1}H^{N}_i$-resolved experiments were implemented (Figure 1). $^{15}N_i$ and $^{13}C'_{i-1}$ shifts are jointly sampled for breaking $^{15}N,^{1}H^{N}$ shift degeneracy,[2] and $^{13}C^{\alpha/\beta}$ and $^{13}C^{\alpha}$ shifts are jointly sampled for sequentially linking spin systems.[4c] Resulting (5,3)D HN{N,CO}-{$C^{\alpha/\beta}C^{\alpha}$} and HN{NCO}{$C^{\alpha/\beta}C^{\alpha}$} provide, respectively, intraresidue and sequential connectivities via one-bond scalar couplings (Figure 1) based on $2*\Omega(^{13}C^{\alpha})$, $\Omega(^{13}C^{\alpha}) + \Omega(^{13}C^{\beta})$, and $\Omega(^{13}C^{\alpha}) - \Omega(^{13}C^{\beta})$.[4c] The brackets group jointly sampled shifts represented by underlined letters,[4a] and in (5,3)D HN{N,CO}{$C^{\alpha/\beta}C^{\alpha}$}, the comma indicates a bifurcated $^{13}C'_{i-1} \leftarrow {}^{15}N_i \rightarrow {}^{13}C^{\alpha}_i$ transfer.[2,5c,e]

The r.f. pulse schemes of (5,3)D HN{N,CO}{$C^{\alpha/\beta}C^{\alpha}$} (Figures 2, S1) and HN{NCO}{$C^{\alpha/\beta}C^{\alpha}$} (Figure S2) yield "out-and-back" transfers. This allows employment of GFT[4a]-TROSY[7a] for (large) deuterated[8] proteins (embedded in membrane mimics) and enables longitudinal ^{1}H relaxation (L-) optimization.[7b] (5,3)D HN{N,CO}-{$C^{\alpha/\beta}C^{\alpha}$} and HN{NCO}{$C^{\alpha/\beta}C^{\alpha}$} L-G²FT NMR experiments were performed (Table S1), respectively, in 24 and 20 h for a ~0.8 mM solution of $^{15}N,^{13}C$ doubly labeled 17 kDa protein yqbG, target of the Northeast Structural Genomics consortium, at 25 °C on a VARIAN INOVA 600 spectrometer equipped with a cryogenic probe. Processing yields four subspectra. Each contains one peak of a quartet at ω_1: $\Omega(^{13}C^{\alpha}) \pm \Omega(^{13}C^{\alpha/\beta})/\omega_2$: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$. Assignments are accomplished in three steps. (i) Peak pairs at ω_1: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$ in (3,2)D HNNCO (Figure S3; Table S1) are centered about peaks in 2D [$^{15}N,^{1}H$]-HSQC and provide spin system identification. (ii) Peak pair positions are transferred to (5,3)D G²FT subspectra, where the same ω_1 pattern is observed at ω_2: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$. Signals at ω_1: $2*\Omega(^{13}C^{\alpha})$ are "central peaks" for pair identification at ω_1: $\Omega(^{13}C^{\alpha}) \pm \Omega(^{13}C^{\beta})$,[4c] which profits from increased dispersion along ω_1 due to $C^{\alpha/\beta}C^{\alpha}$ frequency labeling.[4c] (iii) "Sequential walks" at ω_2: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$ in two sets of subspectra yield three connectivities each, that is, a total of six.

α-Helical protein yqbG exhibits $^{15}N,^{1}H^{N}$ shift degeneracy in 2D [$^{15}N,^{1}H$]-HSQC (Figures 3a, S4a). This is aggravated at the lower resolution of 3D spectra (Figures 3b, S4b), where complete degeneracy is observed for eight residues. In contrast, at least one of the two peaks at ω_2: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$ (Figures 3c,d and S4c,d) is resolved for *all* residues. This allows efficient sequential

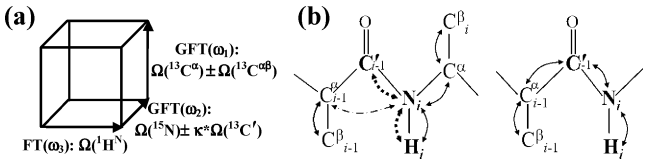


Figure 1. (a) $^{13}C'_{i-1},^{15}N,^{1}H^{N}$-resolved (5,3)D G²FT NMR. (b) Magnetization transfers of HN{N,CO}{$C^{\alpha/\beta}C^{\alpha}$} (left) and HN{NCO}{$C^{\alpha/\beta}C^{\alpha}$} (right). Double arrows indicate bidirectional transfers via one-bond scalar couplings. The "intraresidue" experiment (left) also yields sequential connectivities via smaller[1] two-bond $^{2}J_{CN}$ couplings.

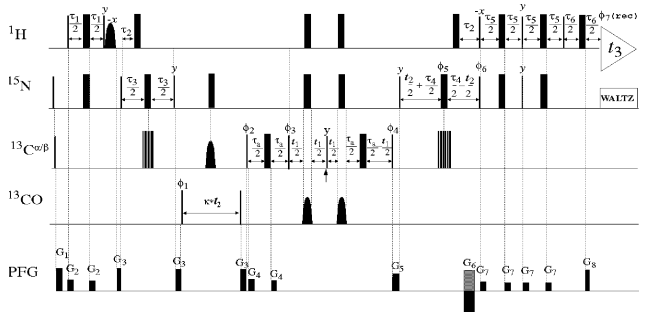


Figure 2. Radio frequency (r.f.) pulse scheme of G²FT L-(5,3)D HN{N,CO}{$C^{\alpha/\beta}C^{\alpha}$}. The 90 and 180° pulses are represented by thin and thick vertical bars. Composite 180° pulses (hatched bars) are used to simultaneously invert $^{13}C^{\alpha}$ and $^{13}C'$ polarization. The scaling[4a,5a] factor κ for $^{13}C'$ shift evolution was set to 0.25 because (i) polarization transfer in the sequential counterpart (i.e., (5,3)D HN{NCO}{$C^{\alpha/\beta}\mathbf{C}^{\alpha}$} (Figures 1, S2)) limits $t_{max}(^{13}C')$ to ~6 ms; (ii) a short $t_{max}(^{13}C')$ limits $T_2(^{13}C')$ losses in (5,3)D HN{N,CO}{$C^{\alpha/\beta}C^{\alpha}$} with non-constant time[1] $^{13}C'$ shift evolution; and (iii) $t_{max}(^{15}N) \approx 24$ ms ensures high spectral resolution in ω_2. A 90° selective pulse after the 2nd "hard" 90° ^{1}H pulse enables water flip-back[1] and L-optimization.[4c,7b] For definition of rectangular *z*-field gradients, delays, and phase cycling, see legend of Figure S1. Quadrature detection in $t_2(^{15}N;^{13}C')$ is achieved with sensitivity enhancement[1] (G_6 is inverted with a 180° shift for ϕ_6); in $t_1(^{13}C^{\alpha};^{13}C^{\alpha/\beta})$, ϕ_4 is altered according to States-TPPI.[1] GFT NMR phase cycle: $\phi_1 = x, y$; $\phi_2 = 2x, 2y$; $\phi_3 = 2y, 2x$. A description of G²FT theory and data processing is provided as Supporting Information.

10.1021/ja042562e CCC: $30.25

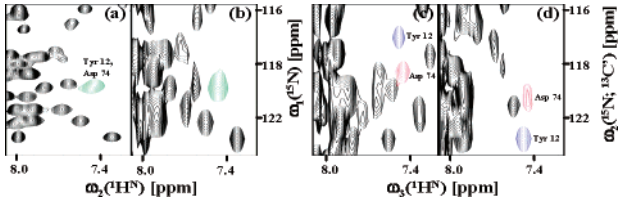


Figure 3. Plots of (a) 2D [^{15}N,^{1}H]-HSQC,[1] (b) [$\omega_1(^{15}N),\omega_3(^1H^N)$] projection of ^{15}N-resolved 3D spectra recorded for protein yqbG (95% H_2O/5% ^2H_2O; 20 mM MES, pH 6.5, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3). (c and d) First, [$\omega_2(^{15}N;^{13}C')$,$\omega_3(^1H^N)$] planes from 3D NCO-resolved experiments ($\kappa = 0.25$) showing signals at $\Omega(^{15}N) + \kappa\Omega(^{13}C')$ (left) and $\Omega(^{15}N) - \kappa\Omega(^{13}C')$ (right). The green signal in (a) and (b) arises from Tyr12 and Asp74 having degenerate ^{15}N and $^1H^N$ shifts; peaks are resolved in (c) and (d) due to nondegenerate $^{13}C'_{i-1}$ shifts.

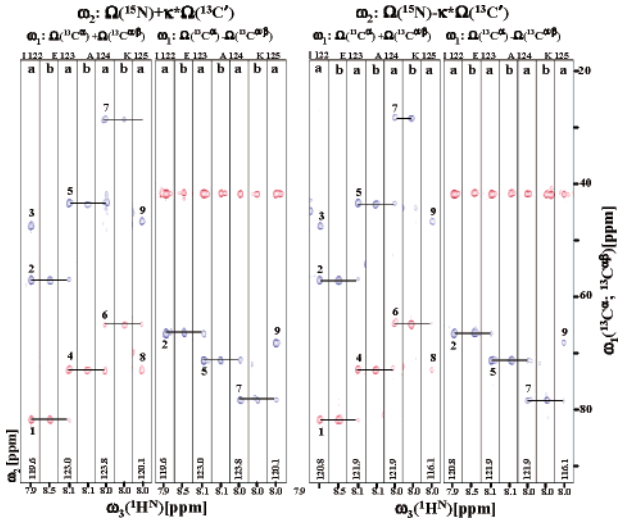


Figure 4. [$\omega_1(^{13}C^\alpha;^{13}C^{\alpha\beta}),\omega_3(^1H^N)$] strips from G^{2}FT (5,3)D HN{N,CO}-{$C^{\alpha\beta}C^\alpha$} ("a") and HN{NCO}{$C^{\alpha\beta}C^\alpha$} ("b") recorded for 17 kDa protein yqbG. Strips were taken at $\omega_2(^{15}N;^{13}C')$ of residues Ile122 to Lys125 (chemical shifts indicated at bottom) and comprise peaks at $\Omega_0(^{13}C^\alpha) \pm \Omega_1(^{13}C^\alpha)$ (red) and $\Omega_0(^{13}C^\alpha) \pm \Omega_1(^{13}C^\beta)$ (blue). Peaks "1"−"9" are assigned to Gly (one letter code: G) 121 (3); Ile (I) 122 (1,2); Glu (E) 123 (4,5); Ala (A) 124 (6,7); Lys (K) 125 (8,9). Connectivities are indicated by dashed lines; six sequential "walks" are established.

assignment using the (5,3)D HN{N,CO}{$C^{\alpha\beta}C^\alpha$} (peak detection yield 93%)/HN{NCO}{$C^{\alpha\beta}C^\alpha$}[9] (yield 95%) experiments (Figure 4). In general, high assignment efficiency is predicted even when considering shifts of proteins with more than 80% α-helical content (Figure S7).

L-Optimization[7b] (Figure 2) can increase sampling speed of out-and-back experiments without loss of intrinsic sensitivity,[4c] yielding minimal measurement times of ∼7 h for (5,3)D HN{N,CO}-{$C^{\alpha\beta}C^\alpha$}/HN{NCO}{$C^{\alpha\beta}C^\alpha$} (Table S1). A further reduction of measurement time can be achieved by maximum entropy reconstruction of nonlinearly sampled data,[6b,10] as is demonstrated here for 13.5 kDa protein rps24e, target of the Northeast Structural Genomics consortium. L-(5,3)D G^{2}FT HN{NCO}{$C^{\alpha\beta}C^\alpha$} data were recorded in 3.5 h (Figure S8), making L-(5,3)D G^{2}FT NMR a viable option for high-throughput data collection in structural genomics.[11]

To enable assignment of systems with very high shift degeneracy, additional G^{2}FT NMR experiments were implemented (Table S1). We recorded for 8 kDa protein Z-domain[12] (5,3)D HN{N,CO}-{$C^\alpha H^\alpha$} (Figure S9; measurement time 16 h; peak detection yield 100%) and (5,3)D {$H^\alpha C^\alpha$}{CON}HN (Figure S10; 13 h; 100%), which allows $^{13}C'_{i-1}$,$^{15}N_i$,$^1H^N_i$-resolved sequential assignment based on $\Omega(^{13}C^\alpha)$ and $\Omega(^1H^\alpha)$ (Figure S11), and for 9 kDa protein ubiquitin, we recorded (6,3)D {$H^{\alpha\beta}C^{\alpha\beta}C^\alpha$}{CON}HN (Figure S12, S13; 24 h; 100%) for assignment of $^1H^{\alpha\beta}$. As was shown for (4,3)D GFT congeners,[4c] $C^{\alpha\beta}C^\alpha$-type G^{2}FT experiments can be combined with (5,3)D HCC-CH for aliphatic side chain assignment. Experiments profiting from large $^{13}C^\alpha$ shift dispersion[1] were recorded for 13.5 kDa protein rps24e. (5,3)D G^{2}FT HN{NC$^\alpha$}{$C^{\alpha\beta}C^\alpha$} (Figure S14; 13 h; 100%) and (5,3)D HN{N(CO)C^α}{$C^{\alpha\beta}C^\alpha$} (Figures S15; 13 h; 100%) enable $^{13}C^\alpha$,^{15}N,$^1H^N$-resolved sequential assignment (Figure S16). The (5,3)D G^{2}FT NMR experiments can be combined with ^{15}N-resolved (4,3)D counterparts.[4c,13] This enables one to establish sequential walks at ω_2: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C')$ or ω_2: $\Omega(^{15}N) \pm \kappa\Omega(^{13}C^\alpha)$ as well as ω_2: $\Omega(^{15}N)$.[14] Taken together, the novel (5,3)D G^{2}FT NMR experiments are powerful for efficiently assigning proteins with high shift degeneracy and promise to pave the way for NMR-based structural genomics of membrane proteins.[15]

Acknowledgment. This work was supported by NIH (P50 GM62413-01) and NSF (MCB 00075773 and 0416899). We thank Drs. G. Montelione and T. Acton for providing samples of proteins Z-domain, yqbG, and rps24e, and Dr. A. Stern for helpful discussions regarding MER of G^{2}FT NMR spectra.

Supporting Information Available: G^{2}FT design principles, theory, and data processing. Table S1 with acquisition parameters. Radio frequency pulse schemes and additional contour plots. This material is available free of charge via the Internet http://pubs.acs.org.

References

(1) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy*; Academic Press: San Diego, CA, 1996.
(2) (a) Szyperski, T.; Braun, D.; Fernandez, C.; Bartels, C.; Wüthrich, K. *J. Magn. Reson. B* **1995**, *109*, 229−233. (b) Konrat, R.; Yang, D.; Kay, L. E. *J. Biomol. NMR* **1999**, *15*, 309−313.
(3) Szyperski, T.; Yeh, D. C.; Sukumaran, D. K.; Moseley, H. N. B.; Montelione, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8009−8014.
(4) (a) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393. (b) Kim, S.; Szyperski, T. *J. Biomol. NMR* **2004**, *28*, 117−130. (c) Atreya, H. S.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9642−9647.
(5) (a) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wüthrich, K. *J. Am. Chem. Soc.* **1993**, *115*, 9307−9308. (b) Brutscher, B.; Simorre, J. P.; Caffrey, M. S.; Marion, D. *J. Magn. Reson. B* **1995**, *109*, 238−242. (c) Szyperski, T.; Braun, D.; Banecki, B.; Wüthrich, K. *J. Am. Chem. Soc.* **1996**, *118*, 8147−8148. (d) Löhr, F.; Rüterjans, H *J. Biomol. NMR* **1995**, *6*, 189−195. (e) Szyperski, T.; Banecki, B.; Braun, D.; Glaser, R. W. *J. Biomol. NMR* **1998**, *11*, 387−405.
(6) (a) Bodenhausen, G.; Ernst, R. R. *J. Magn. Reson.* **1981**, *45*, 367−373. (b) Atreya, H. S.; Szyperski, T. *Methods Enzymol.* **2005**, *394*, 78−108. (c) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2003**, *125*, 13958−13959. (d) Coggins, B. E.; Venters, R. A.; Zhou, P. *J. Am. Chem. Soc.* **2004**, *126*, 1000−1001.
(7) (a) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371. (b) Pervushin, K.; Vogeli, B.; Eletski, A. *J. Am. Chem. Soc.* **2002**, *124*, 12898−12902.
(8) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 307−318.
(9) For nondeuterated proteins, (5,3)D {$C^{\alpha\beta}C^\alpha$}{CON}HN (Figures S5 and S6) is often more sensitive then the out-and-back implementation.
(10) (a) Hoch, J. C.; Stern, A. S. *NMR Data Processing*; Wiley-Liss: New York, 1996. (b) Rovnyak, D.; Frueh, D. P.; Sastry, M.; Sun, Z.-Y. J.; Stern, A. S.; Hich, J. C.; Wagner, G. *J. Magn. Reson.* **2004**, *170*, 15−21.
(11) Montelione, G. T.; Zheng, D.; Huang, Y.; Gunsalus, C.; Szyperski, T. *Nat. Struct. Biol.* **2002**, *7*, 982−984.
(12) Tashiro, M.; Tejero, R.; Zimmerman, D. E.; Celda, B.; Nilsson, B.; Montelione, G. T. *J. Mol. Biol.* **1997**, *272*, 573−590.
(13) ^{15}N-resolved (4,3)D GFT subspectra can be created by symmetrizing[2a] pairs of corresponding (5,3)D G^{2}FT subspectra along $\omega_1(^{15}N;^{13}C')$ about the position of $\Omega(^{15}N)$, accurately defined in (3,2)D HNNCO.
(14) In the three sets of subspectra constituting (5,3)D HN{N,CO}{$C^{\alpha\beta}C^\alpha$}, (5,3)D HN{NCO}{$C^{\alpha\beta}C^\alpha$}, and (4,3)D HNN$C^{\alpha\beta}C^\alpha$, nine sequential walks are established (Figure S3). Hence, combination of all three pairs of (5,3)D G^{2}FT experiments described in this publication can provide a total of 24 sequential walks.
(15) Sorgen P. L.; Hu, Y.; Guan, L.; Kaback, H. R.; Girvin, M. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14037−14040.

JA042562E