

## Resonance Assignment of Proteins with High Shift Degeneracy Based on 5D Spectral Information Encoded in G<sup>2</sup>FT NMR Experiments

Hanudatta S. Atreya, Alexander Eletsy, and Thomas Szyperski\*

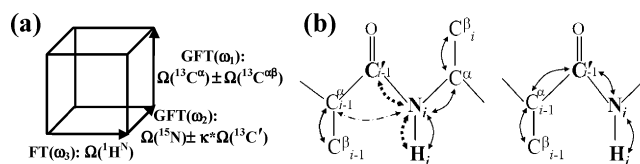
Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260

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

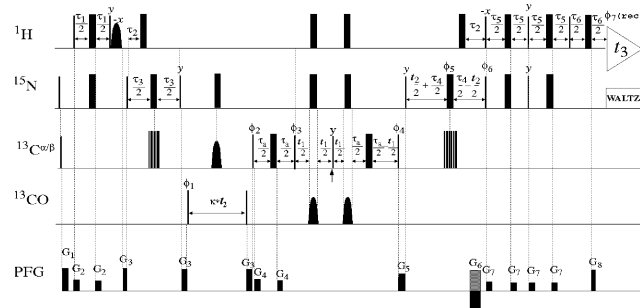
NMR assignments of proteins are obtained by combining several multidimensional experiments.<sup>1</sup> <sup>15</sup>N, <sup>1</sup>H<sup>N</sup>-resolved triple resonance experiments sequentially linking <sup>13</sup>C<sup>αβ</sup> and/or <sup>1</sup>H<sup>α</sup> shifts are the most widely used.<sup>1</sup> 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. <sup>15</sup>N, <sup>1</sup>H<sup>N</sup> degeneracy can be largely removed in <sup>13</sup>C<sup>′</sup><sub>*i*-1</sub>, <sup>15</sup>N<sub>*i*</sub>, <sup>1</sup>H<sup>N</sup><sub>*i*</sub>-resolved experiments<sup>2</sup> (*i* is a residue number). When using correlated <sup>13</sup>C<sup>αβ</sup> or <sup>13</sup>C<sup>α</sup>/<sup>1</sup>H<sup>α</sup> shifts to establish connectivities, conventional NMR<sup>1</sup> would require the recording of 5D spectra. Measurement times for such spectra are prohibitively long or lead to “sampling-limited” data collection.<sup>3</sup> G-matrix FT (GFT) NMR,<sup>4</sup> rooted in reduced-dimensionality NMR<sup>5</sup> and related to accordion spectroscopy,<sup>6a,b</sup> can (i) rapidly provide precise high-dimensional spectral information and (ii) serve to reconstruct higher-dimensional spectra.<sup>6c,d</sup> Previously published (4,3)D GFT experiments<sup>4c</sup> 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 “G<sup>2</sup>FT 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<sup>4c</sup> for sequential assignment are retained in the second GFT dimension. First, <sup>13</sup>C<sup>′</sup><sub>*i*-1</sub>, <sup>15</sup>N<sub>*i*</sub>, <sup>1</sup>H<sup>N</sup><sub>*i*</sub>-resolved experiments were implemented (Figure 1). <sup>15</sup>N<sub>*i*</sub> and <sup>13</sup>C<sup>′</sup><sub>*i*-1</sub> shifts are jointly sampled for breaking <sup>15</sup>N, <sup>1</sup>H<sup>N</sup> shift degeneracy,<sup>2</sup> and <sup>13</sup>C<sup>αβ</sup> and <sup>13</sup>C<sup>α</sup> shifts are jointly sampled for sequentially linking spin systems.<sup>4c</sup> Resulting (5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>} and HN{NCO}-{C<sup>αβ</sup>C<sup>α</sup>} provide, respectively, intraresidue and sequential connectivities via one-bond scalar couplings (Figure 1) based on  $2^* \Omega(^{13}\text{C}^\alpha)$ ,  $\Omega(^{13}\text{C}^\alpha) + \Omega(^{13}\text{C}^\beta)$ , and  $\Omega(^{13}\text{C}^\alpha) - \Omega(^{13}\text{C}^\beta)$ .<sup>4c</sup> The brackets group jointly sampled shifts represented by underlined letters,<sup>4a</sup> and in (5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>}, the comma indicates a bifurcated <sup>13</sup>C<sup>′</sup><sub>*i*-1</sub> ← <sup>15</sup>N<sub>*i*</sub> → <sup>13</sup>C<sup>α</sup><sub>*i*</sub> transfer.<sup>2,5c,e</sup>

The r.f. pulse schemes of (5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>} (Figures 2, S1) and HN{NCO}-{C<sup>αβ</sup>C<sup>α</sup>} (Figure S2) yield “out-and-back” transfers. This allows employment of GFT<sup>4a</sup>-TROSY<sup>7a</sup> for (large) deuterated<sup>8</sup> proteins (embedded in membrane mimics) and enables longitudinal <sup>1</sup>H relaxation (L-) optimization.<sup>7b</sup> (5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>} and HN{NCO}-{C<sup>αβ</sup>C<sup>α</sup>} L-G<sup>2</sup>FT NMR experiments were performed (Table S1), respectively, in 24 and 20 h for a ~0.8 mM solution of <sup>15</sup>N, <sup>13</sup>C doubly labeled 17 kDa protein yqgG, 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  $\omega_1$ :  $\Omega(^{13}\text{C}^\alpha) \pm \Omega(^{13}\text{C}^\beta)$ / $\omega_2$ :  $\Omega(^{15}\text{N}) \pm \kappa\Omega(^{13}\text{C}^\alpha)$ . Assignments are accomplished in three steps. (i) Peak pairs at  $\omega_1$ :  $\Omega(^{15}\text{N}) \pm \kappa\Omega(^{13}\text{C}^\alpha)$  in (3,2)D HNNCO (Figure S3; Table S1) are



**Figure 1.** (a) <sup>13</sup>C<sup>′</sup><sub>*i*-1</sub>, <sup>15</sup>N, <sup>1</sup>H<sup>N</sup>-resolved (5,3)D G<sup>2</sup>FT NMR. (b) Magnetization transfers of HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>} (left) and HN{NCO}-{C<sup>αβ</sup>C<sup>α</sup>} (right). Double arrows indicate bidirectional transfers via one-bond scalar couplings. The “intraresidue” experiment (left) also yields sequential connectivities via smaller <sup>2</sup>J<sub>CN</sub> couplings.



**Figure 2.** Radio frequency (r.f.) pulse scheme of G<sup>2</sup>FT L-(5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>}. The 90 and 180° pulses are represented by thin and thick vertical bars. Composite 180° pulses (hatched bars) are used to simultaneously invert <sup>13</sup>C<sup>α</sup> and <sup>13</sup>C<sup>β</sup> polarization. The scaling<sup>4a,5a</sup> factor  $\kappa$  for <sup>13</sup>C<sup>′</sup> shift evolution was set to 0.25 because (i) polarization transfer in the sequential counterpart (i.e., (5,3)D HN{NCO}-{C<sup>αβ</sup>C<sup>α</sup>}) (Figures 1, S2)) limits  $t_{\text{max}}(^{13}\text{C}^\alpha)$  to ~6 ms; (ii) a short  $t_{\text{max}}(^{13}\text{C}^\alpha)$  limits  $T_2(^{13}\text{C}^\alpha)$  losses in (5,3)D HN{N,CO}-{C<sup>αβ</sup>C<sup>α</sup>} with non-constant time<sup>1</sup> <sup>13</sup>C<sup>′</sup> shift evolution; and (iii)  $t_{\text{max}}(^{15}\text{N}) \approx 24$  ms ensures high spectral resolution in  $\omega_2$ . A 90° selective pulse after the 2nd “hard” 90° <sup>1</sup>H pulse enables water flip-back<sup>1</sup> and L-optimization.<sup>4c,7b</sup> For definition of rectangular z-field gradients, delays, and phase cycling, see legend of Figure S1. Quadrature detection in  $t_2(^{15}\text{N}; ^{13}\text{C}^\alpha)$  is achieved with sensitivity enhancement<sup>1</sup> ( $G_6$  is inverted with a 180° shift for  $\phi_6$ ; in  $t_1(^{13}\text{C}^\alpha; ^{13}\text{C}^\beta)$ ,  $\phi_4$  is altered according to States-TPP1.<sup>1</sup> GFT NMR phase cycle:  $\phi_1 = x, y$ ;  $\phi_2 = 2x, 2y$ ;  $\phi_3 = 2y, 2x$ . A description of G<sup>2</sup>FT theory and data processing is provided as Supporting Information.

centered about peaks in 2D [<sup>15</sup>N, <sup>1</sup>H]-HSQC and provide spin system identification. (ii) Peak pair positions are transferred to (5,3)D G<sup>2</sup>FT subspectra, where the same  $\omega_1$  pattern is observed at  $\omega_2$ :  $\Omega(^{15}\text{N}) \pm \kappa\Omega(^{13}\text{C}^\alpha)$ . Signals at  $\omega_1$ :  $2^* \Omega(^{13}\text{C}^\alpha)$  are “central peaks” for pair identification at  $\omega_1$ :  $\Omega(^{13}\text{C}^\alpha) \pm \Omega(^{13}\text{C}^\beta)$ ,<sup>4c</sup> which profits from increased dispersion along  $\omega_1$  due to C<sup>αβ</sup>C<sup>α</sup> frequency labeling.<sup>4c</sup> (iii) “Sequential walks” at  $\omega_2$ :  $\Omega(^{15}\text{N}) \pm \kappa\Omega(^{13}\text{C}^\alpha)$  in two sets of subspectra yield three connectivities each, that is, a total of six.

α-Helical protein yqgG exhibits <sup>15</sup>N, <sup>1</sup>H<sup>N</sup> shift degeneracy in 2D [<sup>15</sup>N, <sup>1</sup>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  $\omega_2$ :  $\Omega(^{15}\text{N}) \pm \kappa\Omega(^{13}\text{C}^\alpha)$  (Figures 3c,d and S4c,d) is resolved for *all* residues. This allows efficient sequential