Journal of Biomolecular NMR **29:** 467–476, 2004.

(3,2)D GFT-NMR experiments for fast data collection from proteins

Youlin Xia[a], Guang Zhu[a,b], Sudha Veeraraghavan[c] & Xiaolian Gao[a,*]
[a]*Department of Chemistry, University of Houston, Houston, TX 77004-5003, U.S.A.;* [b]*Department of Biochemistry, Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, SAR, China;* [c]*Structural Biology Research Center, Department of Biochemistry & Molecular Biology, University of Texas – Houston Medical School, U.S.A.*

Received 19 November 2003; Accepted 9 April 2004

Key words: GFT-NMR, multidimensional NMR experiments, protein structure determination, RD-NMR

Abstract

High throughput structure determination of proteins will contribute to the success of proteomics investigations. The G-Matrix Fourier Transformation NMR (GFT-NMR) method significantly shortens experimental time by reducing the number of the dimensions of data acquisition for isotopically labeled proteins (Kim, S. and Szyperski, T. (2003) *J. Am. Chem. Soc.* **125**, 1385). We demonstrate herein a suite of ten 3D→2D or (3,2)D GFT-NMR experiments using $^{13}C/^{15}N$-labeled ubiquitin. These experiments were completed within 18 hours, representing a 4- to 18-fold reduction in data acquisition time compared to the corresponding conventional 3D experiments. A subset of the GFT-NMR experiments, (3,2)D HNCO, HNCACB, HN(CO)CACB, and 2D ^{1}H-^{15}N HSQC, which are necessary for backbone assignments, were carried out within 6 hours. To facilitate the analysis of the GFT-NMR spectra, we developed automated procedures for viewing and analyzing the GFT-NMR spectra. Our overall strategy allows (3,2)D GFT-NMR experiments to be readily performed and analyzed. Nevertheless, the increase in spectral overlap and the reduction in signal sensitivity in these fast NMR experiments presently limit their application to relatively small proteins.

Introduction

The sequencing of whole genomes and the desire to understand biological systems have presented unprecedented challenges to those who are in the field of structural biology. A large gap exists between the need for structural information of biomolecules, especially proteins, and the time required to obtain this information. Intensive efforts in protein biology, NMR technology, and the development of methods for structure determination have been enlisted to narrow this gap. Conventional NMR experiments required for protein structure determination take from days to weeks of instrument time, limiting the overall speed of structure determination. Additionally, some proteins in solution tend to precipitate in a matter of days, thereby reducing the time available to record NMR data. Several methods have been proposed to reduce the duration of NMR data collection, such as time-sharing (TS) NMR (Pascal et al., 1994; Farmer and Mueller, 1994; Xia et al., 2001, 2003), reduced dimensionality NMR (RD-NMR) (Brutscher et al., 1994, 1995; Löhr and Rüterjans, 1995; Ding and Gronenborn, 2002; Szyperski et al., 2002; Xia et al., 2002), projection-reconstruction NMR (PR-NMR) (Kupče and Freeman, 2003), ZQ/DQ/MQ-NMR (Simorre et al., 1994; Bersch et al., 2003), and G-Matrix FT NMR (GFT-NMR) (Service, 2003; Kim and Szyperski, 2003; Koźmiński and Zhukov, 2003). In the following discussion, we refer to the dimension in which multiple nuclei evolve as the 'shared' dimension and the dimension whose frequency is encoded in the shared dimension as the 'projecting' dimension. The nuclei in the shared dimension will be underlined in the name of the experiment (e.g., HACACONHN).

*To whom correspondence should be addressed. E-mail: xgao@uh.edu

A common scheme in these methods of fast data collection is time-sharing or simultaneous incrementing of time, such as in t_1 and t_2, so that the data in multiple dimensions are collectively recorded in a shared dimension. Consequently, these methods can significantly shorten experimental time. In the RD-NMR spectroscopy, three or four chemical shifts are encoded in two- or three-dimensional spectra, i.e., (3,2)D or (4,3)D RD-NMR, using cosine modulation. Using this method, 4D or 3D correlations of ^{1}H, ^{13}C and ^{15}N in reduced 3D or 2D data sets can be obtained more quickly. For example, ten (4,3)D data sets can be recorded in 65 hours compared to about two weeks required for the corresponding conventional experiments (Szyperski et al., 2002). 2D versions of the triple resonance experiments HN(CO)CACB and HN(COCA)CACB and the corresponding versions for CAHA (Ding and Gronenborn, 2002) have been used to assign backbone chemical shifts; the inter-residue correlations are removed in the 2D HN(COCA)CACB experiment to simplify the spectrum. A 2D triple resonance experiment can be run in less than ten hours. Given the attractive features of the above experiments, a few points concerning practical experimental aspects should be noted. In the above RD-NMR experiments, the cosine modulation doubles the number of cross peaks, which may introduce peak overlapping (Koźmiński and Zhukov, 2003). These RD-NMR experiments also use a shift of the carrier frequency or the TPPI method to obtain an artificial frequency offset for the projecting dimension; this results in a large spectral width in a shared dimension. Although use of a larger spectral width does not give lower signal sensitivity (S/N) as long as the digital resolution and experiment time remain the same (compared to those of a normal experiment), the larger spectral width requires that more FIDs be acquired. There is still a need for improving pulse sequences, for example: two 54° ^{15}N pulses are used to provide the central peak in either 2D HN(CO)CACB or HN(COCA)CACB (Ding and Gronenborn, 2002), which decreases the intensity of the doublet by 35% ($\sin^2 54° = 0.65$). Additional steps of magnetization transfer in 2D HN(COCA)CACB and lack of ^{1}H-decoupling during the periods of ^{15}N evolution in 2D HN(CO)CACB and HN(COCA)CACB further reduce the signal sensitivity.

PR-NMR represents a different way reducing the number of dimensions, which is achieved by cross peak reconstruction based on two 2D orthogonal projections at $t_1 = 0$ and $t_2 = 0$, respectively, and a 2D tilted projection with t_1 and t_2 incremented simultaneously and in a fixed ratio (Kupče and Freeman, 2003). Their HNCO experiment performed with a ubiquitin sample required a total of 29 min, whereas the corresponding conventional 3D would require 39-fold more time. The sensitivity and spectral resolution of the 2D PR-NMR experiment are comparable to its 3D counterpart.

G-matrix Fourier Transformation NMR (GFT-NMR), an extension of RD-NMR, is a new approach for rapidly acquiring multidimensional NMR data. A smaller spectral width can be used in GFT-NMR compared to RD-NMR, by simultaneously acquiring cosine- and sine-modulated components for each projecting dimension, thereby allowing positive and negative frequencies to be separated. By reducing data acquisition from *n* dimensions to fewer dimensions, GFT-NMR can considerably shorten the time for data acquisition (Kim and Szyperski, 2003). For example, in a (5,2)D HACACONHN GFT-NMR experiment, four frequencies (HA, CA, CO, and N) are encoded in a shared dimension to give 2D correlation spectra, achieving a time saving of up to 125-fold compared to a conventional 5D HACACONHN experiment. However, the signal sensitivity in the GFT-NMR spectra is significantly lower than that of conventional FT-NMR spectra because of peak splitting in each projecting dimension. In addition, the encoding of chemical shifts of nuclei of multiple types in one dimension in GFT-NMR spectra complicates the data processing and analysis. In contrast, the recently published suite of seven 2D experiments (Bersch et al., 2003) encodes chemical shifts of nuclei of two types into a shared dimension to simplify data processing and analysis. The use of long delays ($T + \Delta = 43$ ms) in HNCA, HN(CA)CB, and HN(CA)HA results in further decay of the signal strength.

The use of GFT-NMR to reduce a 3D to a 2D experiment represents a good compromise of sensitivity (for mM protein samples) and time-saving. Although a few (3,2)D GFT-NMR experiments (Koźmiński and Zhukov, 2003; Bersch et al., 2003) have been reported, there is still a need for a robust, integrated process for chemical shift assignments of backbone and side chains at low sample concentrations. In this work, we demonstrate a suite of ten important and frequently used 3D NMR experiments (Cavanagh et al., 1996), namely (3,2)D HNCO, HNCACB, HN(CO)CACB, HN(CA)CO, HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, C(CCO)NH, and H(CCCO)NH, for complete assign-

ments of doubly labelled proteins. In these experiments, the chemical shifts of ^{13}C or $^{1}H_{aliph}$ are encoded in the ^{15}N dimension. The first eight experiments in the preceding list are used for backbone assignments and the last two experiments are used for side chain chemical shifts. These GFT-NMR experiments require considerably less time than that of conventional experiments. In 18 h, all ten (3,2)D GFT-NMR spectra were acquired using a 1.4 mM ubiquitin sample on a Bruker Avance 800 spectrometer, whereas 135 h were needed for conventional 3D experiments. The sensitivities of these experiments were reduced by 2.0–2.8-fold compared to the corresponding conventional 3D FT-NMR experiments as shown in Table 1. Since GFT-NMR spectra encode chemical shifts in a non-conventional way in the shared dimension, a new interface of the NMR analysis software SPARKY (Goddard and Kneller) and additional data analysis scripts have been developed in order to facilitate spectral interpretation and to take full advantage of the GFT-NMR method.

Materials and methods

The sample used was 1.4 mM $^{13}C/^{15}N$ uniformly labeled human ubiquitin (0.5 ml in 50 mM potassium phosphate, 95% H_2O-5% D_2O, pH 5.8, VLI, Research Inc., PA). Experiments were performed at 25 °C on a Bruker Avance 800 spectrometer equipped with a triple-resonance, z-axis pulsed field gradient probehead. The pulse sequence of the (3,2)D HNCO GFT-NMR experiment is modified from that of the normal 3D HNCO experiment (Kay et al., 1994) as follows: the evolution times of $^{13}C'$ (C' denotes CO) and of ^{15}N are incremented simultaneously; two 2D data matrices with the phase of the first $^{13}C'$ pulse being 0° and 90°, respectively, are acquired in an interleaved mode and stored separately (which corresponds to quadrature detection for the projecting $^{13}C'$ dimension). The pulse sequences of other (3,2)D GFT NMR experiments were obtained by modifying the corresponding normal 3D NMR experiments in a way similar to that used in (3,2)D HNCO. The acquisition parameters of the (3,2)D GFT-NMR and their conventional counterparts (Clubb et al., 1992; Grzesiek and Bax, 1992, 1993; Schleucher et al., 1993; Wittekind and Mueller, 1993; Kay et al., 1994; Muhandiram and Kay, 1994; Yamazaki et al., 1994; Carlomagno et al., 1996) are given in Table 1. The spectral widths in F1 dimensions of all (3,2)D GFT-NMR experiments were set to 10–15% less than the sums of F1 and F2 spectral widths without causing any spectral folding. In order to reduce the spectral width and increase digital resolution of the F1 dimension of (3,2)D H(CCCO)NH, the carrier frequency of the aliphatic proton was shifted from 4.7 to 3.0 ppm for the first INEPT frequency-labeling of ^{1}H and the transferring of magnetization from ^{1}H to ^{13}C. All data were processed using NMRPIPE (Delaglio, 1995). Linear prediction was used to double the data size in the t_1 dimension (and t_2 dimension for all 3D spectra) in order to increase spectral resolution (Zhu and Bax, 1990, 1992). Cosine-squared bell window functions were applied to both t_1 and t_2 dimensions.

Results and discussion

GFT-NMR data acquisition

The (3,2)D HNCO GFT NMR experiment is used herein as an example to illustrate the general strategy of the data acquisition. In this experiment quadrature detection is used to distinguish the sign of the frequencies of the projecting dimension as described in the literature (Kim and Szyperski, 2003; Koźmiński and Zhukov, 2003). Two 2D data sets were acquired by altering the phase of the first $^{13}C'$ pulse by 90°: one corresponds to the cosine modulation, represented by $(\cos(\Omega_{C'}t_1))(= 0.5(\exp(i\Omega_{C'}t_1)+\exp(-i\Omega_{C'}t_1)))$, where $\Omega_{C'}$ is chemical shift of $^{13}C'$; the other corresponds to the sine modulation, represented by $\sin(\Omega_{C'}t_1)(= -0.5\ i\ (\exp(i\Omega_{C'}t_1) - \exp(-i\Omega_{C'}t_1)))$. The phase of the second data set is shifted by 90° in the F_1 dimension so that it becomes $i\sin(\Omega_{C'}t_1)$ $(= 0.5\ (\exp(i\Omega_{C'}t_1) - \exp(-i\Omega_{C'}t_1)))$. With the cosine and sine modulation, the form of the two 2D data sets of (3,2)D HNCO is such that at each ^{1}H frequency (F2 axis) there are at least two connectivity peaks symmetrical to Ω_N (Hz) and separated by $2\times\Omega_{C'}$ (Hz). The two peaks in the first 2D spectrum are positive, while the two peaks in the second 2D spectrum are positive and negative in intensity. The addition and subtraction of the two 2D spectra result in two new 2D spectra, referred to as ADD and SUB spectrum, respectively. The described procedure for data processing can be easily performed with NMR software. For a given HNCO connectivity, one peak located at $\Omega_N + \Omega_{C'}$ appears in the spectrum ADD, and another peak located at $\Omega_N - \Omega_{C'}$ appears in the spectrum SUB. The chemical shift $\Omega_{C'}$, encoded in Ω_N, can then be derived by referencing to the central position Ω_N determined from

Table 1. Acquisition parameters of (3,2)D GFT-NMR and conventional 3D NMR experiments

(3, 2)D GFT-NMR experiments						3D conventional NMR experiments							
Exp. type	Num. of scans	Data size (t1×2)	Spectral width (F1), (Hz)	S/N	Exp. time (min/h)	Exp. type	Num. of scans	Data size (t1×t2)	Spectral width (F1,F2), (Hz)	S/N	Exp. time (min/h)	Time saving factor	S/N reducing factor
HNCO[1]	2	64×2	4865	120	10/0.2	HNCO5	2	35×32	3018, 2757	1272	180/3.0	18	1/2.5
HNCACB2	16	128×2	15003	49	165/2.8	HNCACB6	8	70×32	14084, 2757	414	1444/24.1	8.8	1/2.8
HN(CO)CACB[2]	16	128×2	15003	61	168/2.8	HN(CO)CACB6	8	70×32	14084, 2757	505	1468/24.5	8.7	1/2.8
HN(CA)CO[3]	8	64×2	4865	48	42/0.7	HN(CA)CO5	2	35×32	3018, 2757	229	182/3.0	4.3	1/2.3
HNCA[4]	4	128×2	7298	78	41/0.7	HNCA5	2	40×32	6037, 2757	500	204/3.4	5.0	1/2.8
HN(CO)CA[4]	4	128×2	7298	124	41/0.7	HN(CO)CA5	2	40×32	6037, 2757	730	181/3.0	4.4	1/2.8
CBCANH	16	80×2	15003	38	112/1.9	CBCANH6	8	70×32	14084, 2757	320	1468/24.5	13.1	1/2.3
CBCA(CO)NH	16	80×2	15003	45	112/1.9	CBCA(CO)NH6	8	70×32	14084, 2757	333	1484/24.7	13.3	1/2.0
C(CCO)NH	16	128×2	17029	40	174/2.9	C(CCO)NH7	4	70×32	14487, 2757	206	752/12.5	4.3	1/2.4
H(CCCO)NH	16	128×2	8108	23	188/3.1	H(CCCO)NH7	4	70×32	4801, 2757	108	753/12.6	4.0	1/2.4
Total exp. hours					17.6						135.3	7.7	

Data sizes are given in number of complex points; '×2' in the column of Data size (t1) for the (3,2)D GFT-NMR means two 2D data sets were acquired; S/N is an average value for all isolated peaks. Time saving factor is the ratio of experimental time of the conventional 3D FT-NMR over that of its (3,2)D GFT-NMR. S/N reducing factor is the ratio of S/N of the (3,2)D GFT-NMR over that of the conventional 3D FT-NMR if they are carried out with the same experimental times. A 2D ^{15}N-^{1}H HSQC was acquired in 10 min to provide central positions. Related references: 1. Simorre et al. (1994); Brutscher et al. (1994); 2. Ding and Gronenborn (2002); Bersch et al. (2003); 3. Brutscher et al. (1994); 4. Simorre et al. (1994); Brutscher et al. (1994); Koźmiński and Zhukov (2003); Bersch et al. (2003); 5. Clubb et al. (1992); Grzesiek and Bax (1992); Schleucher et al. (1993); Kay et al. (1994); 6. Grzesiek and Bax (1992, 1993); Wittekind and Mueller (1993); Muhandiram and Kay (1994); Yamazaki et al. (1994); 7. Carlomagno et al. (1996).

the 2D ^{15}N-^{1}H HSQC spectrum of the same sample and multiplying by the ratio of gyromagnetic ratios k ($k = \gamma_N/\gamma_C$ or γ_N/γ_H). In another example shown in Figure 1, $\Omega_N + \Omega_{CA}$ and $\Omega_N + \Omega_{CB}$ appear in an ADD spectrum and $\Omega_N - \Omega_{CA}$ and $\Omega_N - \Omega_{CB}$ appear in a SUB spectrum. CA and CB in ADD and SUB spectra are easily recognized by their antiphase peaks. Therefore, complete assignments of ^{1}H, ^{13}C, and ^{15}N chemical shifts are feasible based on the analysis of the two 2D spectra simultaneously acquired in the GFT-NMR experiments and a 2D ^{15}N-^{1}H HSQC reference spectrum.

The data acquisition and processing of other (3,2)D GFT-NMR experiments (Table 1) followed methods similar to what is described for the (3,2)D HNCO GFT-NMR experiment.

Sensitivity and time-saving of the (3,2)D GFT-NMR experiments

The experimental times and the observed sensitivities for all (3,2)D GFT-NMR experiments are given in Table 1 along with that of their conventional counterparts. The various (3,2)D GFT-NMR data can be acquired in 10 min to three hours with sufficient sensitivity. The S/Ns of all isolated peaks in the ADD and SUB spectra of the (3,2)D HNCO recorded in 10 min (5 min per 2D dataset) range from 44 to 409, with an average S/N of 120. In the less sensitive (3,2)D experiments for backbone assignments, such as (3,2)D HNCACB, the S/Ns of all isolated peaks for intra-residue correlations in the ADD and SUB spectra range from 19 to 122 with an average S/N of 49, which means that weak peaks are strong enough for making assignments. The GFT-NMR experiments for backbone assignment can reduce the experimental time by more than 7.7-fold, and the sensitivity of (3,2)D GFT-NMR is acceptable for a protein sample of regular concentration (1 mM) and small molecular weight (10 kDa or less). One would expect further gaining in time-saving and/or sensitivity using CryoProbe technologies.

If S/N is not limiting, the overall time required for NMR experiments is determined by the number of scans per FID and the size of the data. These parameters for the (3,2)D GFT-NMR or the conventional 3D NMR experiments are given in Table 1. For 3D experiments, it is desirable to record at least two scans for each FID because this can provide better solvent suppression than using just one scan. Thus, when possible the number of scans of 3D HNCO, HN(CA)CO, HNCA, and HN(CO)CA was set to two, which also minimized experimental time (with the shortest 3D experiment taking three hours (Table 1)). Eight scans were needed with each FID for 3D HNCACB, HN(CO)CACB, CBCANH and CBCA(CO)NH to obtain cleaner spectra with fewer artifacts. This requirement lengthens experimental time, with an unnecessarily high sensitivity. GFT-NMR on the other hand, by encoding multinuclear frequencies in a shared dimension, reduces a 3D data set into two 2D data sets plus a reference 2D spectrum (Table 1, columns for data size). Therefore, (3,2)D GFT-NMR considerably shortens experimental time compared with conventional 3D NMR experiments by an average factor of 7.7 for the ten critical experiments. Several 3D experiments which require a day to complete can be accomplished using (3,2)D GFT-NMR experiments in 2–3 h, and with good quality (Table 1). As an example, a (3,2)D HN(CO)CACB spectrum and its 1D traces are shown in Figure 1. This data set was used in combination with (3,2)D HNCACB and 2D ^{15}N-^{1}H HSQC spectra to obtain chemical shift assignments for Ω_H, Ω_N, Ω_{CA-}, Ω_{CB-}, Ω_{CA} and Ω_{CB} (CA-, CB-, CA and CB are $^{13}C_\alpha$ and $^{13}C_\beta$ of residues i-1 and i, respectively).

The (3,2)D GFT-NMR experiments shown in Table 1 show sensitivity reduction by 2–3 fold from the corresponding conventional 3D experiments. Several reasons may account for the observed lower apparent sensitivity in the GFT-NMR data recorded under the given conditions (Table 1, last column): (1) Peak splitting in the shared dimension. In this case the sensitivity of the (3,2)D GFT-NMR experiments is reduced by a factor of squareroot of two (Kim and Szyperski, 2003). (2) Further reduction in sensitivity due to the difference in data processing procedures. For instance, linear prediction (Zhu and Bax, 1990, 1992) cannot only improve resolution of the ^{15}N dimension in 3D HNCO, but also enhance signal sensitivity by more than a factor of squareroot of two. (3) Differences in other experimental parameters. For example, longer $t_{1\,max}$ used in (3,2)D GFT-NMR data collection causes additional loss of sensitivity. Different experimental and data processing conditions were used for optimization of the performance of each type of experiment. In a comparison data set of the regular 3D HNCO and (3,2)D GFT HNCO acquired using comparable parameters and no linear prediction, the ratio of the S/N of the regular 3D HNCO *versus* (3,2)D GFT HNCO was calculated for all assigned and reasonably resolved cross peaks (71 peaks in the ADD

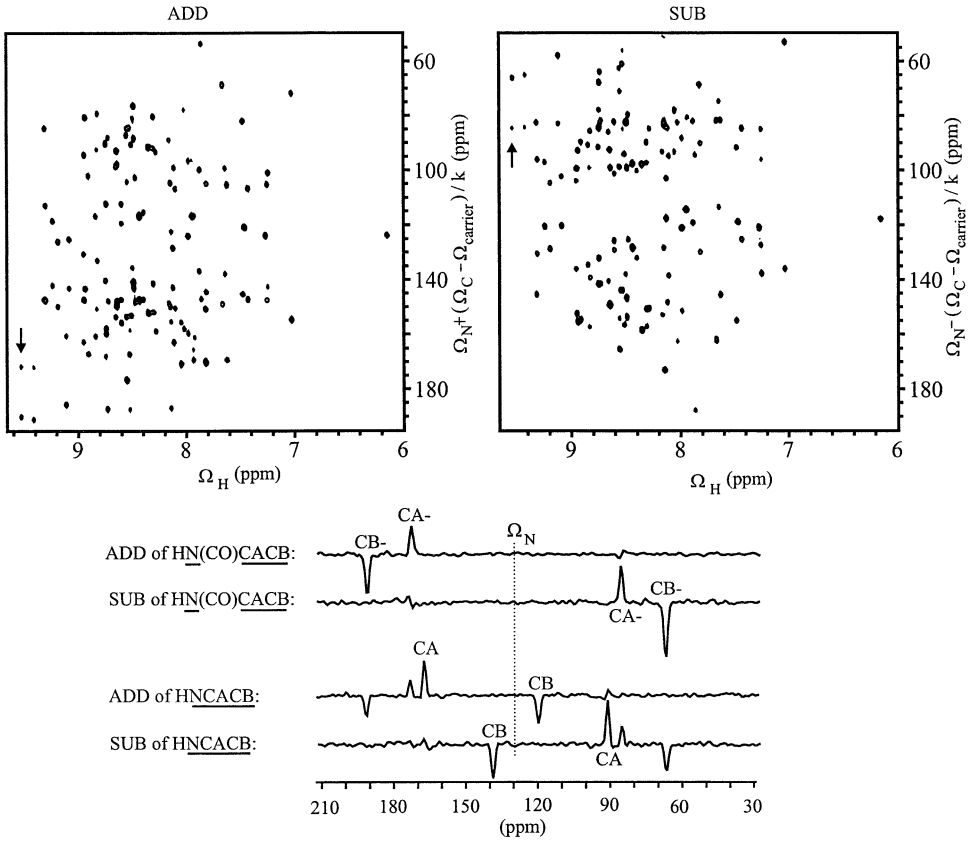

Figure 1. 2D plot of (3,2)D HN(CO)CACB and 1D traces taken from this spectrum and from the HNCACB spectrum at Ω_H indicated by arrows in 2D plot. Spectra ADD and SUB are from the addition and subtraction of the two 2D data sets acquired in the (3,2)D HN(CO)CACB experiment. Here, $k = \gamma_C/\gamma_N = 2.48$ and $\Omega_{carrier} = 44.7$ ppm. Peaks in ADD and SUB are of positive and negative intensities, from $^{13}C_\alpha$ and $^{13}C_\beta$, respectively. Ω_N is derived from the ^{15}N-^1H HSQC spectrum. The doublets (the two peaks symmetrical to Ω_N) can be assigned to give the chemical shifts of CA, CA-, CB and CB-. The chemical shifts in the F1 dimension of the 2D plot are scaled according to the ^{15}N carrier frequency.

and 74 peaks in the SUB spectrum) to give an average value of 1.45 ± 0.14, which is close to squareroot of two as discussed in the literature (Kim and Szyperski, 2003) (Figure S1, Tables S1 and S2, Supplementary material).

GFT-NMR spectral analysis using SPARKY

The shared dimension of the ADD and SUB GFT-NMR contain signals of $K + 1$ nuclei, such as ^{13}C and ^{15}N or $^1H_{aliph}$ and ^{15}N. The resonance frequencies of these signals are characterized by a symmetrical pattern (ADD: $\Omega_N + \Omega_X$, X = ^{13}C or ^1H; SUB: $\Omega_N - \Omega_x$) and their positive or negative phase (Figure 1). These spectral patterns and the match of the reference ^1H-^{15}N HSQC provide a basis for complete assignments of the backbone and side chain resonances in protein molecules. Clearly, the analysis of GFT-NMR spectra is not the same as that of the conventional 3D spectra.

SPARKY is a versatile and interactive software package widely used for spectral analysis of conventional NMR experiments. To effectively analyze the GFT-NMR experiments, two new extensions in Python, cross_hair.py and gft_shift.py, were created and made available to us for interactive analysis of GFT-NMR data in SPARKY. These patches activate a new user interface for resonance assignments in (3,2)D GFT-NMR spectra. One function of the new interface allows simultaneous display of three cursors: one locates the central peak in ^{15}N-^1H HSQC and two others locate the two corresponding symmetrical peaks in ADD and SUB. Once the peaks are located, SPARKY displays the converted chemical shifts of the cross peak and the central position in a new dialog box according to $\Omega_x = \Omega_{carrier} + k \times \Delta\Omega_x$, where

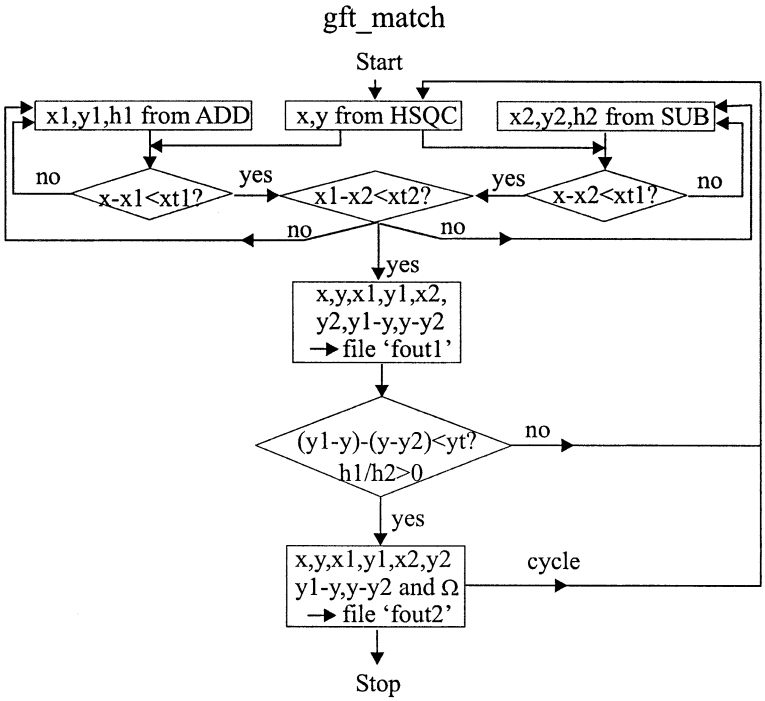

Figure 2. The flowchart of gft_match.awk to automatically extract deconvoluted chemical shifts from data acquired in the (3,2)D GFT-NMR. x, x1, and x2 are chemical shifts of direct acquisition dimensions (Ω_H); y, y1, and y2 are chemical shifts of indirect dimensions; h1 and h2 are peak heights; xt1, xt2, and yt are matching tolerances: xt1 = 0.02 ppm, xt2 = 0.01 ppm, yt = 0.4 to 0.7 ppm, which depends on the spectral resolution in the indirect dimension; fout1 and fout2 are two output files; $\Omega_{C'} = \Omega_{carrier} + k * (y1 - y2)/2$, is the deconvoluted chemical shift in the indirect dimension of the matched peaks in ADD and SUB.

$\Delta\Omega_x$ is the chemical shift separation encoded in and centered at Ω_N ($\Omega_{carrier}$ is the frequency carrier of nuclear ^{13}C or $^{1}H_{aliph}$ and $k = \gamma_N/\gamma_C$ or γ_N/γ_H). These functions make the visualization and analysis of the GFT-NMR spectra a much easier task. The development of the software tool-kit is an important step towards widespread applications of the more efficient GFT-NMR.

Automated GFT-NMR data analysis

Although non-conventional in appearance, the well-defined spectral characteristics of the GFT-NMR spectra allows analysis by using an automated program, gft_match, developed as awk scripts to produce assignments for chemical shifts and connectivities. The gft_match program uses the peak picking files generated from SPARKY as input. Three peak lists for the reference (^{15}N-^{1}H HSQC), ADD and SUB spectra define the peak positions as (x, y), (x1, y1) and (x2, y2), respectively, and peak heights as h, h1 and h2, respectively. The logic of the data analysis is given in Figure 2. There are four criteria embedded in the program: (1) The tolerance, $xt1 = |x - x1|$ or $xt1 = |x - x2|$, in chemical shifts of Ω_H (F2, direct data acquisition dimension) is for matching symmetrical peaks in ADD or SUB with HSQC. (2) The tolerance, $xt2 = |x1 - x2|$, in chemical shifts of Ω_H is for matching the peaks in ADD with SUB or *vice versa*. The tolerance of xt2 (0.01 ppm) is smaller than that of xt1 (0.02 ppm) because the spectra ADD and SUB are from the two data sets acquired in an interleaved manner but HSQC is acquired separately. xt2 further reduces the number of possible assignments in a set of peaks after xt1 filtering. (3) The tolerance, $yt = |(y1 - y) - (y - y2)|$, in chemical shifts of Ω_{F1} (F1, indirect data acquisition dimension) is used to match the symmetrical doublet. If criteria 1–3 are satisfied, an output file *fout1* is generated to give chemical shifts of all dimensions and the connectivities. (4) The comparable heights, h1 and h2, of a pair of peaks can be a criterion of matching in ADD and SUB. These are used as loose constraints due to possible spectral overlap. In gft_match, if $h1/h2 < 0$ (these are negative and positive peaks in intensity), the two peaks are not considered as a symmetrical pair. In addition, the

ratios of heights are written to a file, *fout2*, which can be a reference to confirm the matching. The gft_match program was used to deconvolute chemical shifts of all ten GFT-NMR spectra. As an example, 94% of deconvoluted chemical shifts were correctly obtained with the use of this approach for (3,2)D HNCACB and an additional 6% from redundant matches will be obtained according to the strategy described in the following paragraph.

The symmetrical patterns of the peaks in the shared dimension can be used to resolve overlapping cross peaks due to frequency convolution in GFT-NMR (Figure 3). GFT-NMR does not provide an advantage over conventional FT-NMR for resolving overlapping cross peaks. In one case shown in Figure 3, two Ω_H resonances are degenerate; in the shared dimension of the SUB spectrum, there are two overlapping peaks (Figure 3, the left part of lower spectrum, peak 1,2) due to the following equality of frequency differences: $\Omega_{N1} - \Omega_{C1} = \Omega_{N2} - \Omega_{C2}$. However, the overlapping may be resolved by examining the corresponding ADD peaks where $\Omega_{N1}+\Omega_{C1}$ and $\Omega_{N2}+\Omega_{C2}$ are well resolved and each can be matched with a SUB peak symmetrical to Ω_{N1} and Ω_{N2}, respectively. This was the case for the assignments of CO chemical shifts of Q2 (Ω_H, Ω_N and Ω_{CO} are 8.94, 122.96, and 170.44 ppm, respectively) and A46 (Ω_H, Ω_N and Ω_{CO-} are 8.95, 132.97 and 174.47 ppm, respectively) in (3,2)D HNCO. In another case shown in Figure 3 (the right part), three Ω_H resonances are overlapped. Further complications occur when in the shared dimensions of the SUB spectrum $\Omega_{N3}-\Omega_{C3} = \Omega_{N5} - \Omega_{C5}$ and of the ADD spectrum $\Omega_{N4} + \Omega_{C4} = \Omega_{N5}+\Omega_{C5}$. These ambiguities can be removed by disallowing one pair of symmetrical and doubly matched peaks (Figure 3) if the root peak at (Ω_H, Ω_N) in HSQC already has one pair of symmetrical previously matched peaks. This situation can be found for the assignments of CB chemical shifts of Q31 (Ω_H, Ω_N and Ω_{CB} are 8.55, 123.65, and 27.76 ppm, respectively) and K27 (Ω_H, Ω_N and Ω_{CB} are 8.55, 119.02 and 33.61 ppm, respectively) in (3,2)D HNCACB. These rules for the assignments of overlapping peaks are easily programmable. Furthermore, there is a flag in the file *fout2* indicating the number of times a peak matched such as 2 for peaks labeled 1,2, 3,5 or 4,5 in Figure 3. The scripts list2cacb.awk and cacb_match.awk were used to sort and match the chemical shifts from a pair of experiments, such as (3,2)D HNCACB and HN(CO)CACB. Each root peak located at (Ω_H, Ω_N) of HSQC should correspond to

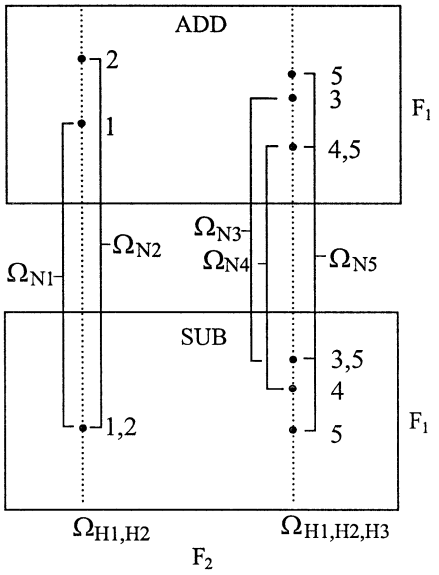

Figure 3. The schematic diagram of overlapped peaks in (3,2)D GFT NMR spectra. Chemical shifts Ω_{H1} to Ω_{H5} and Ω_{N1} to Ω_{N5}, which provide the central positions of peak pairs (or doublets) are from ^{15}N-^{1}H HSQC and $\Omega_{H1} = \Omega_{H2}$, $\Omega_{H3} = \Omega_{H4} = \Omega_{H5}$. Two peaks in a peak pair are labeled with same number.

four ^{13}C chemical shifts (or three chemical shifts if the residue i or $i-1$ is GLY) Ω_{CA}, Ω_{CA-}, Ω_{CB} and Ω_{CB-} for residues i and $i-1$ if it is not overlapped with other peaks. If the number of the corresponding chemical shifts is more than four, some chemical shifts matching more than two times or with ratios of heights being significantly different than 1 (unity) may be removed.

Backbone chemical shifts

The backbone assignments were completed using three GFT-NMR (3,2)D data sets, HNCO, HNCACB, and HN(CO)CACB (Table 1) and a 2D ^{1}H-^{15}N HSQC. The four experiments took a total of less than six hours for the ^{13}C,^{15}N-labeled ubiquitin. The basic procedure of the assignment is as follows: (1) deconvolute chemical shifts from (3,2)D GFT-NMR experiments using gft_match.awk; (2) sort and match the chemical shifts from the pair of experiments using list2cacb.awk and cacb_match.awk to obtain Ω_{CA}, Ω_{CA-}, Ω_{CB} and Ω_{CB-} as discussed above; (3) check peak-picking and identify overlapped peaks so that there are four or three ^{13}C chemical shifts for each root peak located at (Ω_H, Ω_N) of HSQC; and (4) make backbone assignments using AUTOASSIGN (Zimmerman et al., 1997). The assignments were confirmed by iteration of the use of gft_match.awk to check the corres-

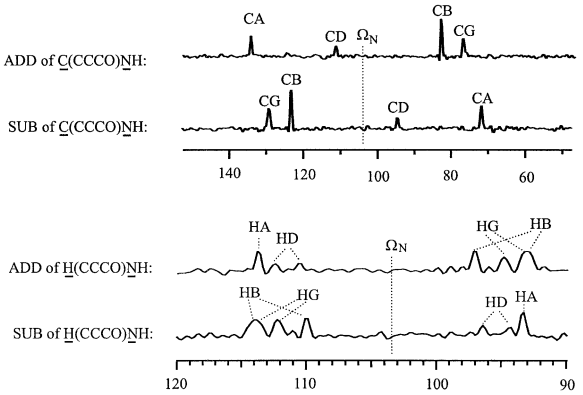

Figure 4. 1D traces taken at $\Omega_H = 7.02$ ppm from the ADD and SUB spectra obtained from (3,2)D C(CCCO)NH and H(CCCO)NH, respectively. The related root (or central) peak at (7.02, 103.45) ppm in HSQC is assigned to residue S20 in the backbone assignment. Therefore the peaks represent the resonances of aliphatic ^{13}C and ^{1}H of residue P19, the residue preceding S20. The deconvoluted chemical shifts are 65.25 (CA), 31.83 (CB), 27.92 (CG), 50.40 (CD), 4.13 (HA), 2.44 and 2.03 (HB), 2.21 and 2.08 (HG), and 4.02 and 3.81 (HD) ppm using $\Omega_{carrier} = 44.70$ and k = 0.403 for ^{13}C and $\Omega_{carrier} = 3.10$ and k = 0.101 for ^{1}H, respectively.

pondences of peaks and to provide higher quality chemical shift input. This combined use of awk (for GFT-NMR spectral peak matching and sorting) and AUTOASSIGN led to the assignments of nearly all backbone resonances except for the chemical shifts of proline residues and these results are given in Table S3 (Supplementary material). This result shows that it is possible to completely automate the GFT-NMR spectral assignments, which is a key step for routine use of this efficient but non-conventional experimental method.

Sidechain chemical shifts

Two GFT-NMR experiments, (3,2)D C(CCCO)NH and H(CCCO)NH, were developed to be used in combination with 3D HCCH-COSY and HCCH-TOCSY to obtain the ^{13}C and ^{1}H chemical shifts of sidechains. Representative spectra and the chemical shift assignments of these two GFT-NMR experiments are shown in Figure 4. The GFT-NMR data were analyzed in a way similar to what has been described above, with each root peak of (Ω_H, Ω_N) in HSQC matching Ω_{CA}, Ω_{CB}, $\Omega_{C'}$, etc. The analysis of (3,2)D H(CCCO)NH indicates that its spectral sensitivity is low. In general, four 3D experiments, C(CCCO)NH, H(CCCO)NH, HCCH-TOCSY and HCCH-COSY (Kay et al., 1993), are used for the sidechain assignments. Among these, 3D C(CCCO)NH, H(CCCO)NH can be readily modified to (3,2)D C(CCCO)NH and H(CCCO)NH GFT-NMR experiments as shown herein. However, for (3,2)D HCCH-TOCSY and HCCH-COSY in the GFT-NMR scheme, the reference positions cannot be obtained from conventional 2D ^{1}H-^{13}C HSQC because the frequency-labeled ^{13}C (in HCCH) is not the same ^{13}C as required by the GFT-NMR frequency-labeling. It is possible to obtain the central peaks if the steady state magnetization of ^{13}C is employed in (3,2)D HCCH-TOCSY and HCCH-COSY (Szyperski et al., 2002). In this case, the overlapping in the $C_\beta H_2$ region of (3,2)D HCCH-TOCSY and HCCH-COSY would be quite extensive, perhaps rendering the GFT-version of these two experiments less useful. Our analysis of the (3,2)D C(CCCO)NH and H(CCCO)NH GFT-NMR data shows the presence of the majority of the ^{13}C and ^{1}H chemical shifts of the sidechains. The (3,2)D C(CCCO)NH experiment gives 219 ^{13}C chemical shifts, of which only five (CE of M1, CG2 of T22, CG of E24, CE of K29, CG of R42) are missing. The (3,2)D H(CCCO)NH gives 288 ^{1}H chemical shifts, of which only 24 are missing. The missing resonances may be recovered from the experiments using different mixing times since the signal strength of the two experiments depends on the mixing time of ^{13}C-^{13}C TOCSY. The mixing time may be varied to obtain more complete peaks.

Conclusion

A set of ten (3,2)D GFT-NMR experiments, which are important for protein chemical shift assignments, is shown herein for rapid data collection of ^{13}C,^{15}N-labeled proteins. The ten GFT-NMR experiments were completed within 18 hours using a ^{13}C,^{15}N -labeled ubiquitin sample at a 1.4 mM concentration. A subset of the GFT-NMR experiments necessary for the backbone assignments, (3,2)D HNCO, HNCACB, HN(CO)CACB, and 2D ^{15}N-^{1}H HSQC, were acquired in less than 6 hours. These results represent significant reduction in data collection time, by at least 4-fold and sometimes, up to 18-fold compared to the time required for conventional 3D experiments. These data sets were analyzed using SPARKY with a combination of new software for GFT-NMR data visualization and extraction and AUTOASSIGN to obtain complete assignment of the backbone chemical shifts. Using (3,2)D C(CCCO)NH and H(CCCO)NH spectra more than 90% of the chemical shifts of side chains are identified. These experiments and methods are designed

to provide convenient solutions for GFT-NMR data acquisition and spectral assignments and should help to accelerate the application of the newly developed, powerful GFT-NMR methods to routine 3D experiments of proteins. Our overall strategy allows (3,2)D GFT-NMR experiments to be readily performed and analyzed. Combined with time-sharing 3D NOESY-HSQC and HSQC-NOESY-HSQC experiments (Xia et al., 2003), which allow one to obtain total NOE correlations for ^1H_C and ^1H_N in a quarter to half of the time of conventional multidimensional experiments, this set of (3,2)D GFT-NMR experiments should contribute to the overall goal of an integrated approach for fast and automated structure determination of proteins using NMR.

Software availability

The software package containing pulse sequences, macros and scripts for (1) data acquisition using a Bruker NMR spectrometer; (2) data processing using XWINNMR and NMRPIPE; and (3) data analysis using SPARKY is available at http://www.chem.uh.edu/Faculty/Gao/ResearchWebpages/index.html

Supplementary material

Figure S1 shows the comparison of the S/N between the regular 3D NHCO and (3,2)D GFT NHCO and the data were collected back to back in sequence. Table S1 lists the experimental parameters used for 3D HNCO. Table S2 lists the experimental parameters ued for (3,2)D GFT HNCO. Table S3 shows the residue-specific chemical shifts from (3,2)D GFT-NMR experiments. Supplementary material available on: http://kluweronline.com/issn/0925-2738.

Acknowledgements

The 800 MHz NMR spectrometer at the University of Houston is funded by the W.M. Keck Foundation and the University of Houston. This research was supported by grants from NIH (GM49957) and the Robert A. Welch Foundation (E-1027). Dr Guang Zhu thanks the Hong Kong Research Grant Council for partial financial support. The authors are grateful to Dr Thomas D. Goddard (University of California, San Francisco) for the new improvements in SPARKY, and Dr David Cohen (University of Houston) for helpful discussions.

References

Bersch, B., Rossy, E., Covès, J. and Brutscher, B. (2003) *J. Biomol. NMR*, **27**, 57–67.
Brutscher, B., Cordier, F., Simorre, J.P., Caffrey, M.S. and Marion, D. (1995) *J. Biomol. NMR*, **5**, 202–206.
Brutscher, B., Simorre, J.P., Caffrey, M.S. and Marion, D. (1994) *J. Magn. Reson.*, **B105**, 77–82.
Carlomagno, T., Maurer, M., Sattler, M., Schwendinger, M.G., Glaser, S.J. and Griesinger, C. (1996) *J. Biomol. NMR*, **8**, 161–170.
Cavanagh, J., Fairbrother, W.J., Palmer III, A.G. and Skelton, N.J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press.
Clubb, R.T., Thanabal V. and Wagner, G. (1992) *J. Magn. Reson.*, **97**, 213–217.
Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
Ding, K. and Gronenborn, A.M. (2002) *J. Magn. Reson.*, **156**, 262–268.
Farmer II, B.T. and Mueller, L. (1994) *J. Biomol. NMR*, **4**, 673–687.
Goddard, T.D. and Kneller, D.G. *SPARKY 3*, University of California, San Francisco.
Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.
Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.
Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) *J. Magn. Reson.*, **B101**, 333–337.
Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) *J. Magn. Reson.*, **A109**, 129–133.
Kim, S. and Szyperski, T. (2003) *J. Am. Chem. Soc.*, **125**, 1385–1393.
Koźmiński, W. and Zhukov, I. (2003) *J. Biomol. NMR*, **26**, 157–166.
Kupče, $\overline{\text{E}}$. and Freeman, R. (2003) *J. Biomol. NMR*, **27**, 383–387.
Löhr, F. and Rüterjans, H. (1995) *J. Biomol. NMR*, **6**, 189–197.
Muhandiram, D.R. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 203–216.
Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1994) *J. Magn. Reson.*, **103**, 197–201.
Schleucher, J., Sattler, M. and Griesinger, C. (1993) *Angew. Chem. Int. Ed.*, **32**, 1489–1491.
Service, R.S. (2003) *Science*, **299**, 503–503.
Simorre, J.P., Brutscher, B., Caffrey, M.S. and Marion, D. (1994) *J. Biomol. NMR*, **4**, 325–333.
Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N.B. and Montelione, G.T. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 8009–8014.
Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **B101**, 201–205.
Xia, Y., Arrowsmith, C.H. and Szyperski, T. (2002) *J. Biomol. NMR*, **24**, 41–50.
Xia, Y., Man D. and Zhu, G. (2001) *J. Biomol. NMR*, **19**, 355–360.
Xia, Y., Yee, A, Arrowsmith, C.H. and Gao, X. (2003) *J. Biomol. NMR*, **27**, 193–203.
Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.
Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.
Zhu, G. and Bax, A. (1992) *J.Magn.Reson.*, **100**, 202–207.
Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., Montelione, G.T. (1997) *J. Mol. Biol.*, **269**, 592–610.