Journal of Biomolecular NMR, **15:** 227–239, 1999.
KLUWER/ESCOM

Efficient side-chain and backbone assignment in large proteins: Application to tGCN5

Yingxi Lin[a,b] & Gerhard Wagner[a,*]
[a]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.;* [b]*Physics Department, Harvard University, Cambridge, MA 02138, U.S.A.*

Received 5 August 1999; Accepted 22 September 1999

Key words: backbone assignment, deuteration, histone acetyltransferase, side chain assignment, tGCN5, TOCSY

Abstract

In determining the structure of large proteins by NMR, it would be desirable to obtain complete backbone, side-chain, and NOE assignments efficiently, with a minimum number of experiments and samples. Although new strategies have made backbone assignment highly efficient, side-chain assignment has remained more difficult. Faced with the task of assigning side-chains in a protein with poor relaxation properties, the Tetrahymena histone acetyltransferase tGCN5, we have developed an assignment strategy that would provide complete side-chain assignments in cases where fast ^{13}C transverse relaxation causes HCCH-TOCSY experiments to fail. Using the strategy presented here, the majority of aliphatic side-chain proton and carbon resonances can be efficiently obtained using optimized H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments on a partially deuterated protein sample. Assignments can be completed readily using additional information from a ^{13}C-dispersed NOESY-HSQC spectrum. Combination of these experiments with H(CC)NH-TOCSY and (H)C(C)NH-TOCSY may provide complete backbone and side-chain assignments for large proteins using only one or two samples.

Introduction

Optimized triple resonance experiments and protein deuteration have made protein backbone assignment very efficient (Muhandiram and Kay, 1994; Yamazaki et al., 1994; Matsuo et al., 1996a, b, c; Shan et al., 1996). However, in order to determine protein structures, a majority of the proton and carbon resonances of the side-chains must also be assigned. This has remained a more difficult task, at least for large proteins. So far, side-chain assignments have typically been obtained with a number of different experiments, including ^{15}N dispersed proton TOCSY (Marion et al., 1989a; Fesik and Zuiderweg, 1990; Bax and Grzesiek, 1993), HCCH-TOCSY (Kay et al., 1993), HCCH-COSY (Kay et al., 1990), HCCH-NOESY (Fischer et al., 1996), and ^{1}H-^{1}H NOESY spectra. Unfortunately, most of the isotropic mixing experiments, and in particular the HCCH-TOCSY, perform poorly on large proteins, and it is tedious and difficult to obtain unambiguous resonance assignments from NOESY information alone. Thus, side-chain assignment has been a bottleneck in the structure determination process.

Poor performance of carbon-carbon TOCSY experiments can be attributed to the fast transverse relaxation of protonated carbons. Deuteration and deuterium decoupling have previously been shown to alleviate this problem (LeMaster and Richards, 1988; Grzesiek et al., 1993b; LeMaster, 1994; Yamazaki et al., 1994; Nietlispach et al., 1996; Kupce et al., 1999). Because the gyromagnetic ratio, γ, of deuterium is more than six times smaller than that of hydrogen, the spin-spin relaxation rates of ^{13}C nuclei are approximately 40 times slower in deuterated proteins than in their protonated counterparts. However, HCCH-TOCSY experiments do not benefit from full or partial deuteration for two reasons. First, the carbons on both ends of the TOCSY transfer pathway

*To whom correspondence should be addressed. E-mail: wagner@hms.harvard.edu

must have directly attached protons, so they still suffer from rapid transverse relaxation; second, the intensities of cross peaks are scaled down as the product of the proton populations on both carbons. The latter problem can be alleviated by transferring coherence from side-chain CH groups to backbone NH spins, such as described by several groups (Grzesiek et al., 1993a; Logan et al., 1993; Lyons and Montelione, 1993; Farmer II and Venters, 1995), when used with deuterated proteins.

Here we present a strategy to obtain nearly complete assignments of NMR spectra of large proteins based on an optimized suite of such experiments. We show that side-chain proton and carbon resonances can be efficiently assigned in a partially deuterated protein by a combination of H(CC-CO)NH-TOCSY, (H)C(C-CO)NH-TOCSY and 3D ^{13}C-dispersed NOESY-HSQC. In contrast to HCCH-TOCSY, the gain in sensitivity from partial deuteration in the first two spectra is dramatic, since the amide groups are fully protonated and all carbons along the transfer pathway can be deuterated, except the one from which the coherence pathway originates. In addition, the N-H spin pair used for detection has more favorable relaxation properties than a C-H pair due to the lower gyromagnetic ratio of ^{15}N. Furthermore, the good dispersion of the amide nitrogens and protons makes analysis of the spectra straightforward. Assignments not obtained from these spectra can be reliably filled in using the ^{13}C-NOESY-HSQC experiment. With this approach, we have obtained virtually complete aliphatic side-chain assignments for the 20 kDa catalytic domain of Tetrahymena GCN5 (tGCN5), a protein for which almost no side-chain assignments could be obtained from HCCH-TOCSY. Despite the relatively small size of this protein, it has unfavorable relaxation properties at room temperature under conditions suitable for NMR studies. Below room temperature it aggregates; at moderately elevated temperatures it unfolds. This required the development of the strategy we describe here. This strategy can be extended by including the (H)C(C)NH-TOCSY and/or H(CC)NH-TOCSY experiments, which yield sequential connectivities as well as side-chain assignments, and could potentially replace multiple pairs of conventional triple resonance experiments typically used for sequential backbone assignments. This reduces the number of experiments needed for protein structure determination.

Materials and methods

Sample preparation

A 100% ^{13}C and ^{15}N labeled, and 65% deuterated tGCN5 sample was expressed in inclusion bodies in *E. coli* BL21(DE3) using M9 minimal media containing 65% D_2O and supplemented with ^{13}C glucose and ^{15}N NH_4Cl. Cell pellets were resuspended in 50 mM sodium phosphate, pH 7.0, 100 mM NaCl, 5 mM EDTA, 10 mM DTT, 1 mM PMSF and 0.25 mg/ml lysozyme. After sonication and centrifugation, the pellets were washed three times with 100 mM Tris.HCl, pH 7.0, 5 mM EDTA, 10 mM DTT, 2 M urea and 2% Triton X-100 and once with the same buffer without urea and Triton X-100. Inclusion body pellets from 1 liter of cell culture were dissolved in 6 M guanidine.HCl in 200 ml buffer A (50 mM Tris.HCl, pH 7.0, 5 mM EDTA and 10 mM DTT) and dialyzed against 1 liter 8 M urea in buffer A at room temperature. The protein was refolded at 4 °C by serial dialysis against 1 liter each of 4 M, 2 M and 0 M urea in 50 mM Tris.HCl, pH 7.0, 0.5 M NaCl, 5 mM EDTA, 0.04% Tween 40 and 10 mM DTT followed by two 1 liter volumes of buffer C (50 mM NaH_2PO_4/Na_2HPO_4, pH 7.0, 100 mM NaCl, 5 mM EDTA and 10 mM DTT). Refolded tGCN5 was purified on an SP sepharose cation exchange column using a 0.1–1 M NaCl gradient in buffer C, eluting at 325 mM NaCl, followed by gel filtration on a Pharmacia HiLoad 16/60, superdex 75 column. NMR samples contained 0.4–0.5 mM protein in 50 mM NaH_2PO_4/Na_2HPO_4, pH 6.5, 100 mM NaCl, 10 mM DTT, 0.8–1.0 mM unlabeled CoA, 0.1 mM NaN_3 and 0.1 mM EDTA.

NMR experiments

A 100% ^{15}N and ^{13}C labeled, 65% deuterated sample in H_2O was used for all triple-resonance experiments, at a concentration of 0.4 mM. The spectra were measured at 300 K, on a 400 MHz Varian UnityPlus spectrometer without deuterium decoupling or a 500 MHz Varian INOVA spectrometer with deuterium decoupling. Each of the 3D triple-resonance spectra were acquired with $64(\omega_1) \times 32(\omega_2) \times 512(\omega_3)$ complex points, 32 scans per increment and 1.8 s recycling delays in about 131 h and processed to $256(\omega_1) \times 128(\omega_2) \times 512(\omega_3)$ points. The TOCSY mixing period consisted of 2 DIPSI-3 cycles (14–15 ms total mixing time). A 3D ^{13}C-dispersed NOESY-HSQC spectrum was acquired on a Varian INOVA 750 MHz spectrometer on a 0.4 mM fully pro-

tonated and 100% ^{13}C labeled samples in D_2O using $150(\omega_1) \times 72(\omega_2) \times 512(\omega_3)$ complex points, 8 scans per increment, a 1 s recycling delay and a total time of 96 h. The mixing time was 100 ms. The spectrum was processed to $512(\omega_1) \times 256(\omega_2) \times 1024(\omega_3)$ points. A 0.4 mM fully protonated and 100% ^{13}C labeled sample in D_2O was used to acquire an HCCH-TOCSY spectrum on a 400 MHz Varian UnityPlus spectrometer with 32 scans, 1 s recycling time, and 14 ms TOCSY mixing time. There are $64(\omega_1) \times 64(\omega_2) \times 512(\omega_3)$ complex points in the acquired spectrum and $256(\omega_1) \times 256(\omega_2) \times 512(\omega_3)$ points in the processed spectrum. The spectral widths are: 10 (ω_1, ^1H) × 32 (ω_2, ^{15}N) × 15 (ω_3, ^1H) ppm for H(CC-CO)NH-TOCSY, 81.6 (ω_1, ^{13}C) × 31.6 (ω_2, ^{15}N) × 14 (ω_3, ^1H) ppm for (H)C(C-CO)NH-TOCSY and (H)C(C)NH-TOCSY, 10 (ω_1, ^1H) × 47.2 (ω_2, ^{13}C) × 14 (ω_3, ^1H) ppm for ^{13}C-dispersed NOESY-HSQC, and 6 (ω_1, ^1H) × 39.7 (ω_2; ^{13}C) × 10 (ω_3, ^1H) ppm for HCCH-TOCSY. All spectra were processed using FELIX (Molecular Simulations, San Diego, CA). Data were apodized using a 90°-shifted, sinebell-squared window function. In the direct detection dimension (ω_3), data were doubled by zero-filling, but the upper-field halves of the spectra were discarded. In the indirect dimensions (ω_1 and ω_2), data were first doubled by linear prediction and then zero-filled to the next nearest power of two.

Results

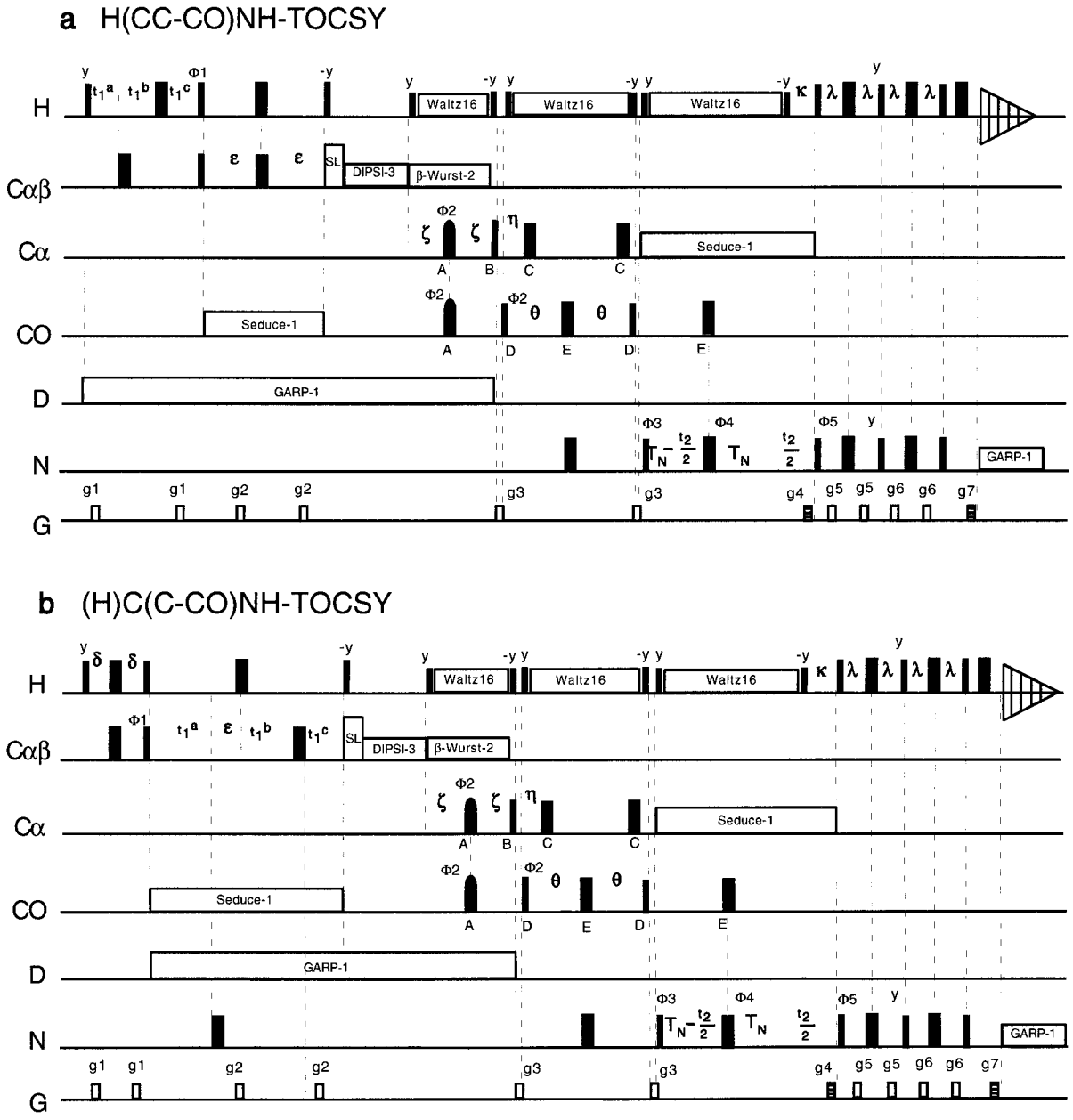
Optimization of isotropic mixing experiments to correlate side-chain CH with backbone NH signals in deuterated proteins

Four pulse sequences were optimized to obtain efficient correlation of side-chain resonances with backbone NH groups. The pulse sequences of the H(CC-CO)NH-TOCSY, (H)C(C-CO)NH-TOCSY, H(CC)NH-TOCSY and (H)C(C)NH-TOCSY are shown in Figure 1. In these experiments, the coherence from one of the rare side-chain protons is first transferred to the attached carbon via INEPT (Morris and Freeman, 1979), and then to the C^α by isotropic mixing. From there it is transferred by INEPT steps to the backbone amide group of the following residue (Figure 1a and 1b) or to the NH of both the same and following residues (Figure 1c and 1d). To derive the experiments of Figure 1a and 1b we started from sequences contained in ProteinPack (Varian Instruments, Palo Alto, CA). These pulse sequences were originally derived from experiments of Grzesiek et al. (1993a) for non-deuterated proteins. The sequences of Figure 1c and 1d use the same coherence transfer pathway as originally proposed by Lyons and Montelione (1993) for non-deuterated proteins. We have found that the following changes improved the performance of these sequences. During the t_1 evolution periods, hyperbolic delays (ProteinPack) are used (Figure 1) to concatenate part or all of the INEPT delay with the t_1 evolution time. This approach is more efficient than the semi-constant time approach originally used by Grzesiek and Bax (1993), because the total duration of the t_1 period is shorter. Deuterium decoupling was added where appropriate, and the placement of pulsed field gradients was optimized so as to achieve coherence selection (Muhandiram and Kay, 1994) for sensitivity enhancement and optimum water suppression. $^{13}C^\beta$ decoupling was added during the 2ζ period to minimize signal loss from carbon-carbon scalar coupling (Figure 1a and 1b). In addition, ^{13}CO decoupling was applied during the 2ζ period in the sequences of Figure 1c and 1d. To optimize solvent suppression, the water magnetization was flipped to the z-axis for most of the sequence except during periods of proton decoupling when it was aligned along the decoupling field. ^{13}C decoupling pulses were used instead of selective 180° pulses to minimize off-resonance effects. Whenever possible, ^{13}C hard pulses were used instead of selective ones to optimize coherence transfers and facilitate carbon pulse calibration.

Sensitivity of the experiments on a partially deuterated protein

Figure 2 shows a comparison of the first plane of an H(CC-CO)NH-TOCSY spectrum on a fully protonated (a) and a 65% deuterated sample (b). The concentration of both samples is the same, and the lowest contours plotted are just above noise level. As expected, partial deuteration yields a dramatic improvement in sensitivity for these spectra, despite the lower proton concentration. For the sake of comparison, the acquisition parameters were kept identical. Although for a fully protonated protein a shorter recycling time (1 s) and roughly twice as many scans per FID could be used to achieve a better signal to noise ratio in the same total measuring time, the gain in sensitivity from partial deuteration, as shown in Figure 2, is more than 1.4-fold.

Figure 3 compares strips from the H(CC-CO)NH-TOCSY with the equivalent strips from an HCCH-TOCSY recorded on a fully protonated sample, show-



ing that the NH correlated experiments are dramatically better. The HCCH-TOCSY strip of Lys35 shown in Figure 3 is one of the best for tGCN5; most residues gave neither diagonal nor cross peaks, as is the case for Pro137 (Figure 3).

Figure 4 shows some typical strips from the 3D H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY spectra. The figure includes residues with long side-chains, such as lysine and arginine, residues with methyl groups, such as isoleucine, leucine, valine and threonine, and glutamine and proline residues. These experiments exhibit signals for the majority of side-chain resonances, even in this poorly behaved protein at the low concentration of 0.4 mM. Figure 4 also shows that the H^{α} and C^{α} signals are weaker than the other side-chain signals. This is due to faster relaxation of C^{α} and a higher level of C^{α} deuteration, as discussed below. However, the main point of

Figure 1. Pulse sequences for optimized versions of experiments where coherence is transferred from side-chain protons of partially deuterated proteins to backbone NH groups. Side-chain carbon magnetization is transferred using carbon-carbon TOCSY to the backbone C^{α} and further transferred by INEPT steps to the NH of the following residue, as in (a) and (b), or to the NH of the same and following residues, as in (c) and (d). The experiments are (a) H(CC-CO)NH-TOCSY, (b) (H)C(C-CO)NH-TOCSY, (c) H(CC)NH-TOCSY, and (d) (H)C(C)NH-TOCSY. Narrow and wide bars correspond to 90° and 180° pulses. Unless otherwise indicated, pulses are along the x-axis. In all four sequences, the ^{1}H carrier is set to the H_2O frequency at 300 K, and the ^{15}N carrier to 119 ppm. In (a) and (b), the ^{13}C carrier is set to 46 ppm until the end of the pulse labeled B, and is then moved to 175 ppm. In (c) and (d), the ^{13}C carrier remains at 46 ppm throughout the sequences. The following field strengths for pulses are appropriate for a 500 MHz spectrometer. The hard pulses for ^{1}H, ^{15}N, and ^{13}C have the respective field strengths of 35 kHz, 5.2 kHz, and 21 kHz. The ^{1}H decoupling is achieved by WALTZ-16 and flanking 90° pulses with a 6.8 kHz field. Decoupling pulses with a 770 Hz field and a GARP-1 profile are applied for ^{2}H decoupling. A 872 kHz field was used for ^{15}N decoupling during acquisition. The rf field for DIPSI mixing is 12 kHz. $^{13}C^{\alpha}$ and ^{13}CO decoupling are achieved by a 20 ppm field centered at 56 ppm and 175 ppm, respectively, using a SEDUCE-1 profile (McCoy and Mueller, 1992). (a) The C^{β} decoupling is achieved by using a WURST-2 profile (Kupce and Freeman 1996a, b; Kupce and Wagner, 1995, 1996) with 6 ppm, 20 ppm, and 4 ppm decoupling bandwidths centered at 20.5 ppm, 35 ppm, and 70 ppm, respectively. The two ^{13}C 180° pulses labeled A in the middle of the the 2ζ period are achieved by a six-pulsed composite 'Shaka 6' pulse (Shaka, 1985). The low power $^{13}C^{\alpha}$ and ^{13}CO pulses (B, C, D and E) have the field strengths of $\Delta/(15)^{1/2}$ and $\Delta/(3)^{1/2}$ for 90° and 180° pulses, respectively, where Δ is the difference in Hz between the $^{13}C^{\alpha}$ (56 ppm) and ^{13}CO (175 ppm) frequencies. The phase cycle is $\Phi1 = (x)$; $\Phi2 = (x, -x)$; $\Phi3 = (x)$; $\Phi4 = (x, x, y, y, -x, -x, -y, -y)$; $\Phi5 = (x)$; rec = (x, −x, −x, x). Quadrature detection in F1 is achieved via States-TPPI of $\Phi1$ (Marion et al., 1989b). Quadrature detection in F2 is achieved using the sensitivity enhanced gradient scheme, where for each value of t_2 N- and P-type coherences are obtained by alternating the signs of g4 and $\Phi5$ (Kay et al., 1992). The delays used are as follows: $\delta = 1/4J_{CH}$, $\epsilon = 1.05$ ms, $\zeta = 3$ ms, $\eta = 4.6$ ms, $\theta = 14$ ms, $\kappa = 5.4$ ms, $\lambda = 2.4$ ms, $T_N = 11$ ms. During the t_1 evolution time, ${t_1}^{a} = t_1/2 + \delta$, ${t_1}^{b} = t_1/2 - \xi$, ${t_1}^{c} = \delta - \xi$, where ξ is a hyperbolic function: $\xi = \delta$ when $t_1 \geq 4\delta$, $\xi = 1/(2/t_1 + 1/2\delta)$ when $AT > 4\delta$ and $t_1 < 4\delta$, $\xi = 1/(2/t_1 + 1/\delta - 2/AT)$ when AT (total Acquisition Time) $< 4\delta$. The durations and strengths of the gradients (rectangular) are: g1 = 0.5 ms, 17.5 G/cm; g2 = 0.25 ms, 28.4 G/cm; g3 = 1.0 ms, 21.9 G/cm; g4 = 2.5 ms, 28.4 G/cm; g5 = 0.5 ms, 4.4 G/cm; g6 = 0.5 ms, 6.6 G/cm; g7 = 0.25 ms, 28.4 G/cm. Gradients g4 and g7 (hatched) should be adjusted to achieve optimized coherence selection during experiments. Delays of 50 μs are used for Eddy current compensation for g1, g2, g5, and g6, and delays of 200 μs for g3, g4, and g7. (b) All the parameters are the same as in (a), except that during the t_1 evolution time, ${t_1}^{a} = t_1/2$, ${t_1}^{b} = t_1/2 - \xi$, ${t_1}^{c} = \epsilon - \xi$, where ξ is a hyperbolic function: $\xi = \epsilon$ when $t_1 \geq 4\epsilon$, $\xi = 1/2(t_1 + 1/2\epsilon)$ when $AT > 4\epsilon$ and $t_1 < 4\epsilon$, $\xi = 1/(2/t_1 + 1/\epsilon - 2/AT)$ when $AT < 4\epsilon$. (c) The parameters are the same as in (a) with the following differences. All the carbon pulses are on-resonance hard pulses. Delay $\zeta = 12.4$ ms. $^{13}C^{\beta}$ and ^{13}CO decoupling during the 2ζ period is achieved by a WURST-2 profile (Kupce and Freeman, 1996a, b; Kupce and Wagner, 1995, 1996) with 6 ppm, 20 ppm, 4 ppm and 20 ppm decoupling bandwidths centered at 20.5 ppm, 35 ppm, 70 ppm and 175 ppm, respectively. (d) The same as in (c), except that the delays during t_1 evolution ${t_1}^{a}$, ${t_1}^{b}$, ${t_1}^{c}$, are the same as those in (b).

the experiments presented here is to obtain efficient side-chain assignments while $C^{\alpha}H$ assignments can be obtained easily from ^{15}N NOESY-HSQC and ^{15}N TOCSY-HSQC with short mixing time. Another reason for the low intensity of the $C^{\alpha}H$ signals could be that most of the coherence originating from the H^{α} is distributed to the side-chain carbons, away from C^{α}, so that it is not transferred to C' at the end of the isotropic mixing period.

Use of H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY with a ^{13}C-dispersed NOESY-HSQC for efficient side-chain assignments

HCCH-TOCSY spectra are often difficult to assign because of overlap in the carbon chemical shifts. Transfer of signals from the whole side-chain to the better dispersed $C^{\alpha}H$ or methyl groups is needed to obtain reliable assignments, but this is difficult in proteins with fast ^{13}C and ^{1}H transverse relaxation rates. In addition to greatly improved sensitivity for large proteins, the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments measured on a partially deuterated sample correlate side-chain resonances with the well-dispersed backbone amide moiety. This makes side-chain assignment reliable and convenient, since the spectra closely resemble the amide correlated spectra used for backbone assignment. In tGCN5 the majority of the side-chain proton and carbon signals were assigned using this pair of experiments.

We record the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY sequences as 3D experiments. They could be recorded as a single 4D experiment (Nietlispach et al., 1996); however, we prefer the 3D versions because they can be recorded at higher resolution and are easier to analyze. The correlation between the side-chain carbon and proton signals can be easily obtained from the diagonal peaks of a 3D ^{13}C-dispersed NOESY as shown in Figure 5. Since the 3D ^{13}C-dispersed NOESY is generally recorded in the course of a structure determination anyway, this approach requires no additional spectrometer time. This experiment also provides reliable assignments of side-chain signals that are not observed in the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments. This is illustrated in Figure 5 for assignment of the missing $H^{\gamma1}$ signals of Ile145 in tGCN5. The $\gamma1$ protons were missing in the H(CC-CO)NH-TOCSY spectrum, but the $\gamma1$ carbon chemical shift

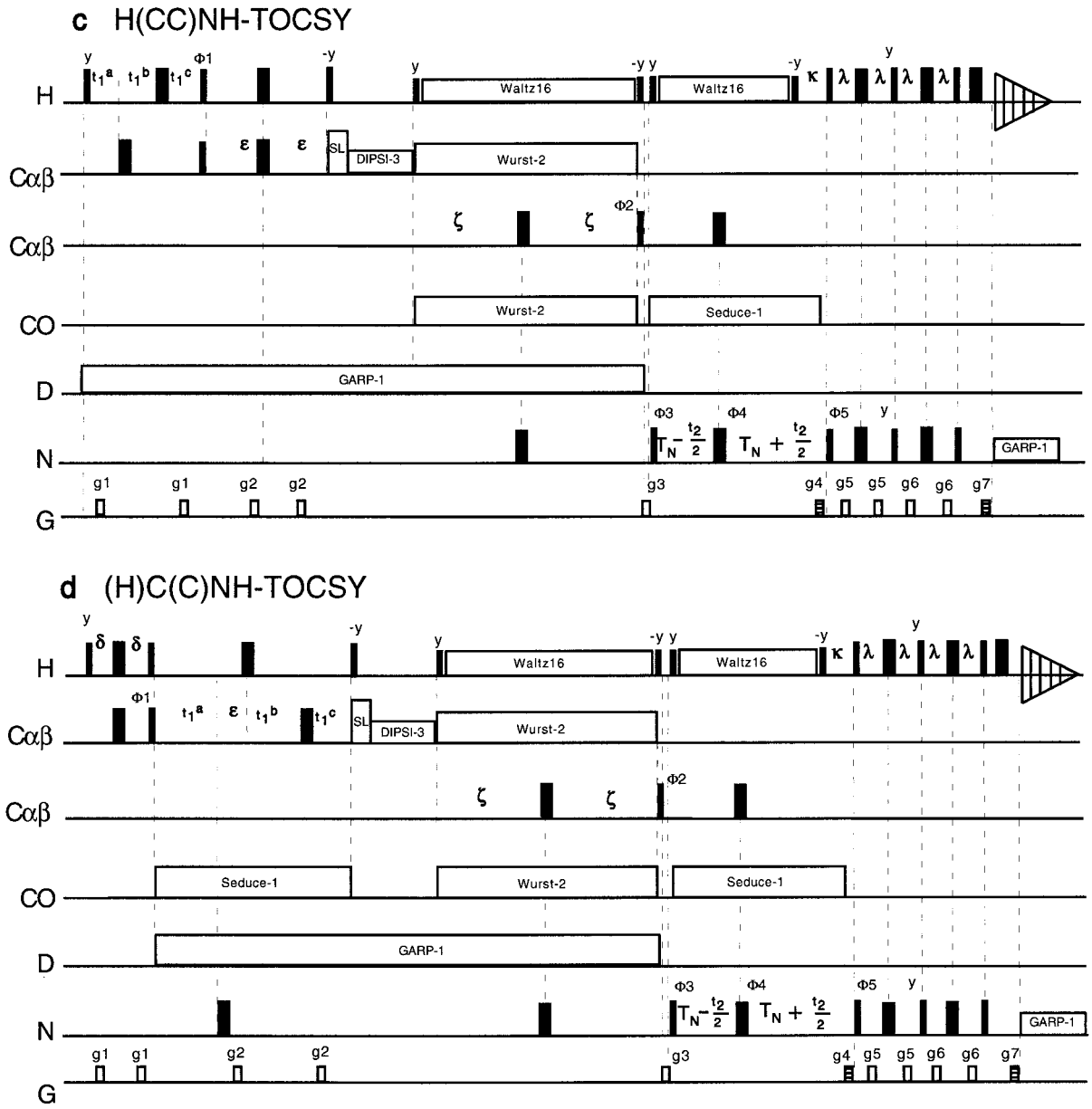


Figure 1. (continued).

was obtained from the (H)C(C-CO)NH-TOCSY. First, the strips corresponding to C^{β}, $C^{\gamma 2}$ and $C^{\delta 1}$ from the ^{13}C-NOESY-HSQC show two strong signals at 1.50 and 1.33 ppm. Second, the ^{13}C plane corresponding to $C^{\gamma 1}$ shows that a pair of protons at 1.50 and 1.33 ppm are attached to a carbon with this chemical shift. Finally, this proton pair shows NOEs to the β, γ2, and δ1 frequencies of Ile145. By combining all of the evidence, these two signals were confidently identified as the missing γ1 CH_2 protons. Missing resonances can be assigned in the same way, even when both proton and carbon signals are missing. One searches the carbon planes in the appropriate range of chemical shifts for that residue type, looking for the appropriate pattern of intra-residue NOEs.

The catalytic domain of tGCN5 has 166 amino residues, including 19 lysines and 46 residues with methyl groups (excluding methionine). A rough mea-

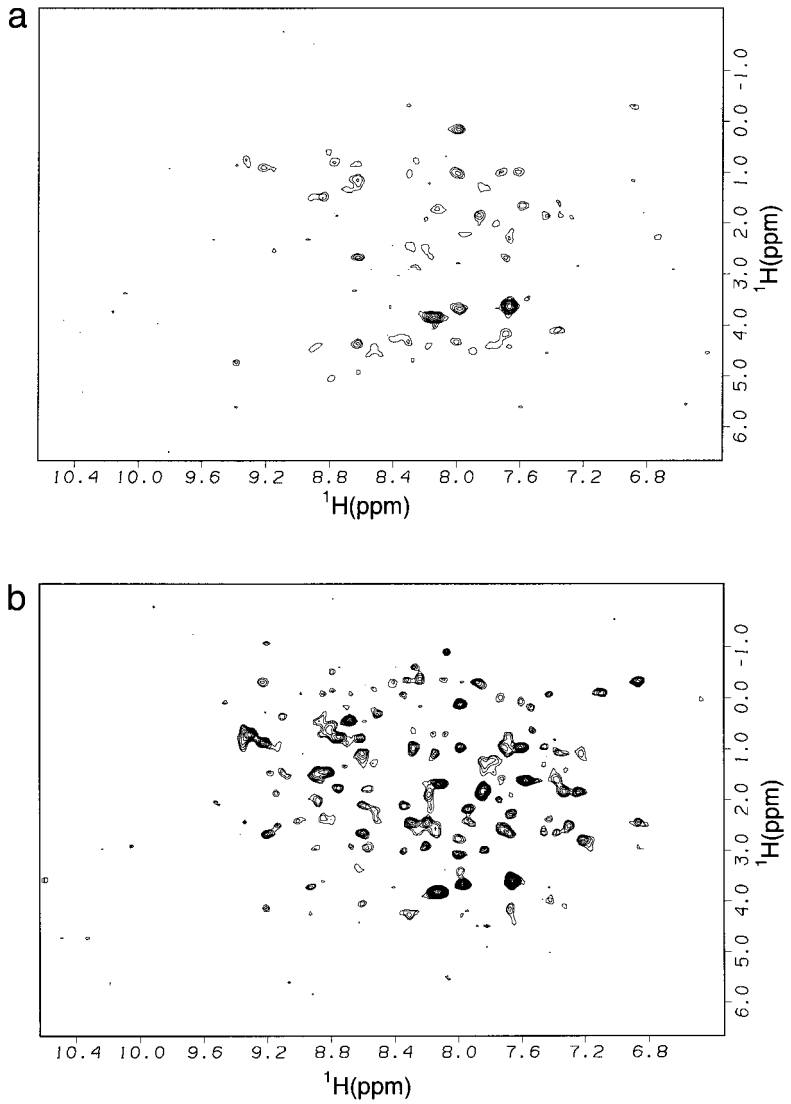


Figure 2. Comparison of the first ω_1-ω_3 planes of the H(CC-CO)NH-TOCSY spectrum of (a) a 100% protonated sample and (b) a 65% deuterated sample of tGCN5. Both planes were acquired on samples of the same concentration using the same parameters: 64(ω_1, 10 ppm) $\times$ 512(ω_3, 15 ppm) complex points, 128 scans per increment and a 1.8 s recycling delay. The Fourier transformed spectra have 256(ω_1) $\times$ 512(ω_3) points and are plotted at the same contour level. The spectra were acquired on a Varian UnityPlus 400 MHz spectrometer. Each experiment used 8.1 h of instrument time.

surement of aliphatic proton linewidths yielded 60–70 Hz for alpha protons and 35–45 Hz for methyl group protons. As stated earlier, almost no side-chain assignments could be obtained from HCCH-TOCSY; however, using the approach described here, complete aliphatic side-chain assignments were obtained in one month, including the time required to acquire three 3D spectra. The backbone amide signals from His50, Arg135 and His159 were missing from all of the spectra, but the side-chain signals of His50 and Arg135 were identified by connectivities to the following residues in the sequence. No side-chain assignments were obtained for Met136 and His159 because they are followed by prolines. This is a significant limitation of the assignment procedure described above, but may be overcome by use of the related HCCNH experiments that correlate side-chain signals to the backbone amide group of the same residue and the following one (see below). Otherwise, only five side-chain resonances of four residues (two lysines, one argi-

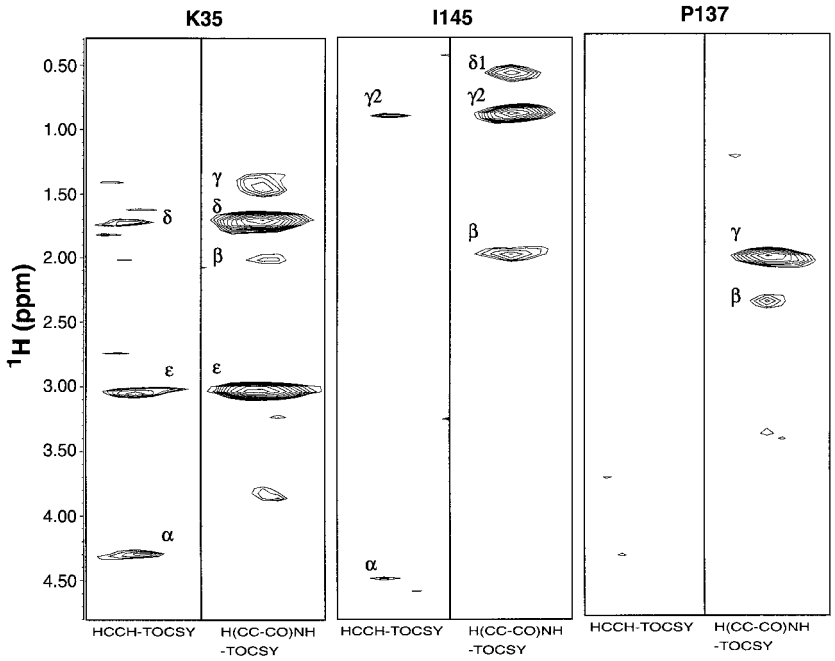


Figure 3. Comparison of equivalent slices from the HCCH-TOCSY (left-hand slice of each pair) and H(CC-CO)NH-TOCSY spectra (right-hand slice of each pair). A 100% ^{15}N and ^{13}C labeled, fully protonated sample of tGCN5 in D_2O was used for the HCCH-TOCSY. A 100% ^{15}N and ^{13}C labeled, 65% deuterated sample of tGCN5 in H_2O was used for the H(CC-CO)NH-TOCSY. Side-chain cross peaks for three representative residues (K35, I145 and P137) are shown. Both spectra were acquired on a Varian 400 MHz UnityPlus spectrometer. The sample concentrations were approximately the same (0.4 mM). The side-chain group nomenclature is the same as that used in the XEASY (Bartels et al., 1995) and XPLOR (Brünger, 1992) libraries. For example, the only δ group in an Ile residue is called δ1.

nine and one methionine) could not be assigned due to spectral overlap. The assignments have been deposited in the BioMagResBank, http://www.bmrb.wisc.edu, with the access code BMRB-1271.

The H(CC)NH-TOCSY and (H)C(C)NH-TOCSY experiments for side-chain and backbone assignments

The optimized pulse sequences of Figure 1c and 1d can be used to obtain side-chain assignments for residues followed by a proline. They are less sensitive than the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY sequences because they utilize the small one-bond and two-bond C^{α}-N couplings, $^1J_{NC^{\alpha}}$ (∼7–10 Hz) and $^2J_{NC^{\alpha}}$ (∼4–9 Hz), instead of the larger C′-N coupling (∼15 Hz). In addition, each C^{α} is coupled to two nitrogens; this creates two pathways that coherence may follow and leads to multiple quantum coherence, both of which attenuate the signal further. Nevertheless, the (H)C(C)NH-TOCSY experiment provided the missing side-chain assignments for Met136 in tGCN5. No assignments for His159 were obtained, since the backbone NH is absent from the spectra.

A combination of H(CC)NH-TOCSY and (H)C(C)NH-TOCSY and the H(CC-CO)NH and (H)C(C-CO)NH experiments can provide sequential backbone assignments as well as side-chain assignments, replacing the multiple pairs of triple resonance experiments that are commonly used to obtain backbone sequential assignments.To save measuring time, the H(CC)NH-TOCSY experiment can be omitted. Sequential and carbon assignments can be obtained from the (H)C(C-CO)NH-TOCSY and (H)C(C)NH-TOCSY pair of experiments, while the side-chain proton resonances can be provided by the more sensitive H(CC-CO)NH-TOCSY experiment. Figure 6 shows a typical series of strips with intra- and inter-residue cross peaks from an (H)C(C)NH-TOCSY experiment.

Discussion

Limitations of HCCH-TOCSY experiments for side-chain assignments in large proteins

While pursuing the structure determination of tGCN5, it became clear that this protein gave very poor TOCSY spectra. The conventional HCCH-TOCSY

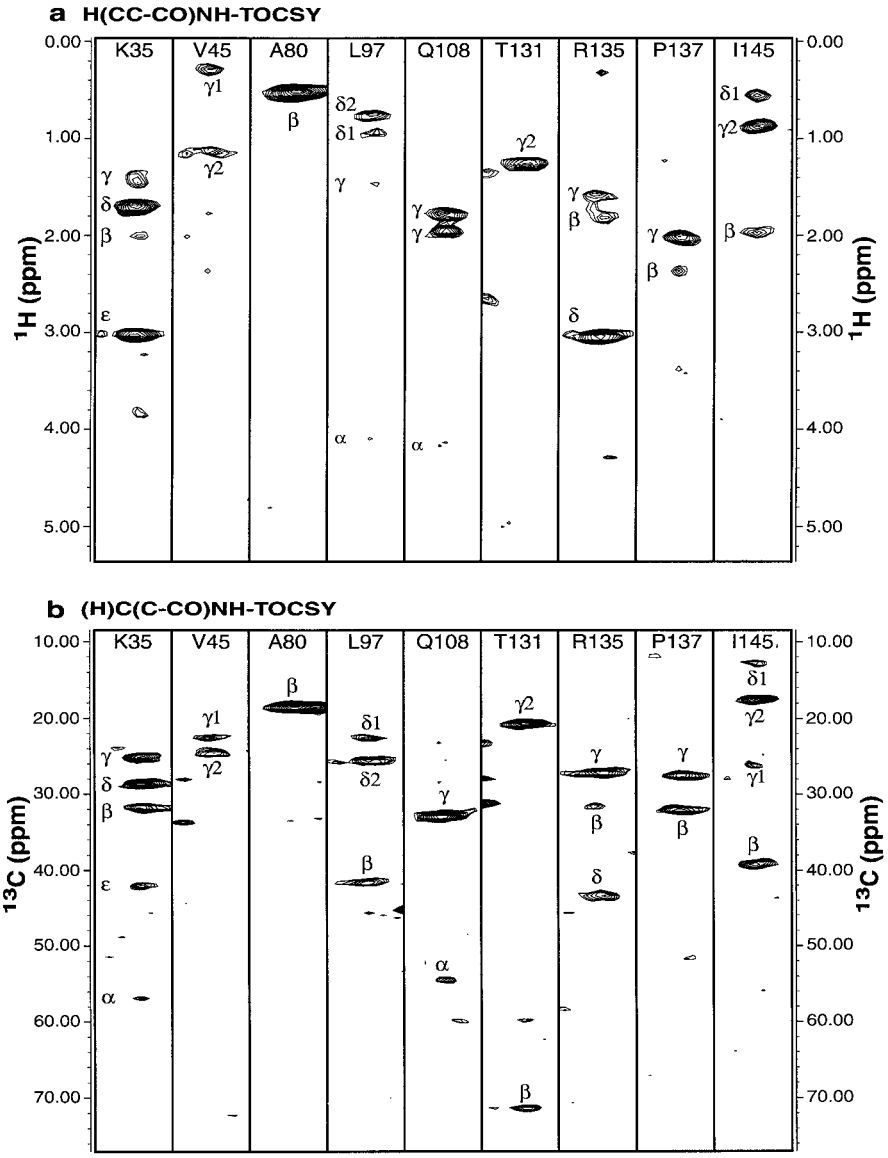


Figure 4. Strips from (a) H(CC-CO)NH-TOCSY and (b) (H)C(C-CO)NH-TOCSY spectra of 65% deuterated tGCN5. The experiment in (a) was recorded on a 400 MHz Varian UnityPlus spectrometer without deuterium decoupling. The spectrum in (b) was recorded on a 500 MHz Varian Inova spectrometer. The strips are identified with the residue type and number of the side-chain signals revealed. The strips are plotted at the nitrogen (ω_2) and amide proton (ω_3) frequencies of the following residue. The full implementation of the sequences of Figure 1a and 1b was used for other proteins and exhibited even better spectra.

provided almost no side-chain assignments, despite extensive variation of sample conditions and examining different isotropic mixing schemes. We attributed this to fast ^{13}C transverse relaxation. Furthermore, due to the overlap in carbon chemical shifts, the most useful correlations potentially available in the HCCH-TOCSY spectrum are those to the well-dispersed $C^{\alpha}H$ signals, but α-carbons relax even more rapidly than the side-chain carbons since the backbone is usually more rigid than the side-chains. Because ^{13}C relaxation is dominated by the dipolar interaction with the directly attached hydrogen, partial deuteration is an obvious way to reduce the ^{13}C transverse relaxation rate. However, as was discussed above, HCCH-TOCSY gains little from partial deuteration. Consistent with this, the ω_1-ω_3 2D slices of HCCH-TOCSY we measured on a

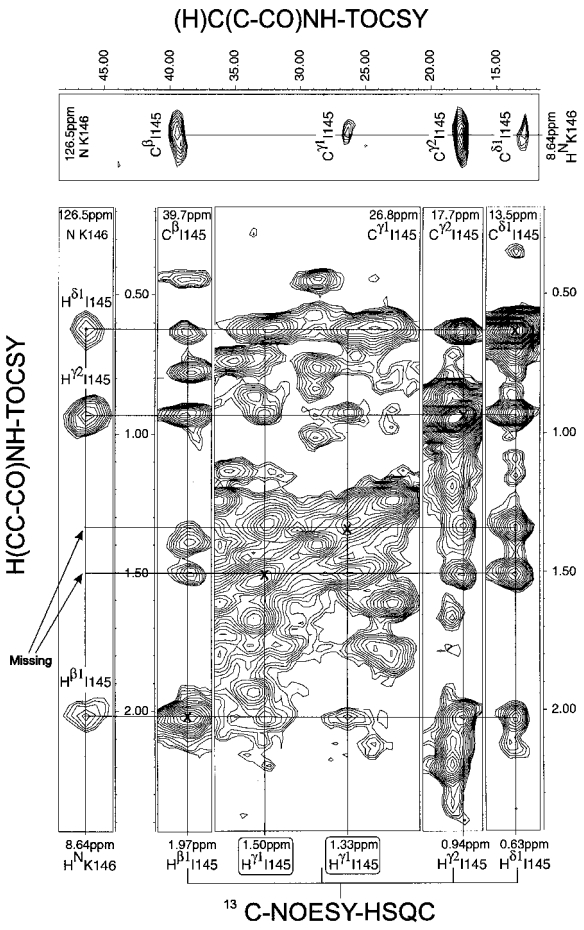


Figure 5. Use of the ^{13}C-NOESY-HSQC spectrum to assign side-chain resonances missing from the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY spectra. The case of Ile145, one of the most difficult spin systems to assign completely, is shown. The diagonal peaks of the 3D ^{13}C-NOESY-HSQC, marked as crosses (X), were used to correlate the carbon and proton resonances observed in the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY strips. Some side-chain protons that do not show up in the H(CC-CO)NH-TOCSY experiment can readily be identified from cross peaks in the ^{13}C-NOESY-HSQC appearing at most of the carbon planes of this residue. The new assignments obtained from the ^{13}C-NOESY-HSQC are highlighted in the boxes at the bottom of the figure.

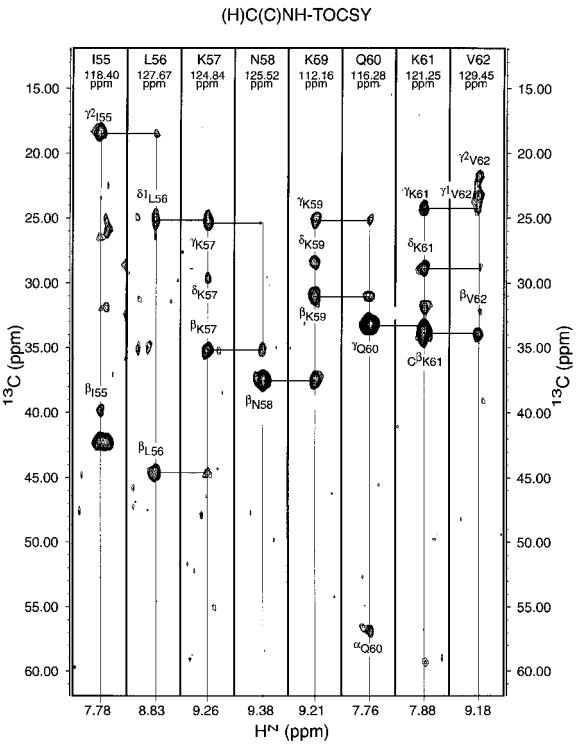


Figure 6. Strips from the (H)C(C)NH-TOCSY spectrum of a 65% deuterated tGCN5 sample. Sequential connectivities are indicated by horizontal lines.

65% deuterated sample of tGCN5 showed no improvement compared to the fully protonated sample (data not shown).

Advantages of the optimized H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments for side-chain assignments

The primary motivation for optimizing the H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments was to obtain side-chain assignments in a protein with poor relaxation properties. The optimized H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments recorded on a partially deuterated sample resolve the side-chain assignment problem for large proteins. Because of the fast relaxation of protonated carbons, the pulse sequences favor transfer in side-chains that do not contain protonated carbons between the excited CH group and the detected NH. Clearly, in a protein that has a fraction ϵ of protons remaining (ϵ is typically around 0.3), the intensity of the cross peaks is scaled down by only ϵ, compared to ϵ^2 in an HCCH-TOCSY, since the NH group is fully protonated. Furthermore, the NH group has more favorable relaxation properties than a CH group (which must be used for detection in an HCCH-TOCSY) because the dipolar interaction between the amide nitrogen and proton is weaker due to the smaller gyromagnetic ratio of the nitrogen. In addition, the N-H detection makes it possible to use the TROSY line-narrowing scheme (Pervushin et al., 1997). While implementing TROSY in the experiments discussed here is straightforward, it would be most favorably achieved at very high field,

where the TROSY effect is more prominent. However, this would require hardware that can provide a carbon spin-lock field covering the whole aliphatic range (i.e. ~40 ppm). This is possible but not trivial with current standard equipment.

Recently, Kessler and co-workers have proposed techniques that use multiple quantum coherence and phase labeling of different spin system topologies in experiments related to H(CC-CO)NH-TOCSY. These schemes produce side-chain signals with different signs for different spin systems (Gschwind et al., 1998). In the experiments presented here, almost complete carbon and proton side-chain spectra are observed for most residues, and carbon chemical shifts provided enough information for identification of residue types. More elaborate experiments for residue type identification may not be necessary.

An alternative version of the experiments recording side-chain carbon chemical shifts, suggested by Farmer and Venters (1995), starts from ^{13}C magnetization. However, we found that the experiments starting from proton magnetization had better sensitivity (data not shown).

Non-uniform deuteration in E. coli

Partially deuterated protein expressed in *E. coli* on minimal medium using protonated glucose has a non-uniform deuterium distribution that is particularly suitable for the experiments presented here. The *E. coli* glucose biosynthesis pathway results in deuteration of C^{α} carbons with essentially the same deuteration level as the solvent. Side-chains show a lower degree of deuteration, however, since hydrogen isotopes are incorporated into side-chains from both the solvent and the glucose metabolic products (Stryer, 1988). In particular, the protonation level of methyl groups remains significantly higher than that of α-carbons. This is clearly visible in the 1D ^{1}H spectrum of a protein grown in 100% D_2O with protonated glucose (data not shown). Such a distribution of protons favors coherence transfer from the protonated side-chains to NH groups via deuterated α-carbons. The incomplete deuteration of side-chain groups makes it feasible to apply the experiments discussed here to proteins that are produced from *E. coli* in 100% D_2O with protonated glucose. Using these experiments we have obtained assignment for approximately 80% of the side-chains in a 27.5 kDa heterodimeric protein expressed in 100% D_2O with protonated glucose (data not shown).

Gardner and co-workers have used related experiments to assign protonated methyl groups in otherwise deuterated proteins (Gardner et al., 1996). While this is an elegant approach for large proteins, preparation of these samples is complicated and expensive, and the assignment is limited to methyl groups. Straightforward expression in *E. coli* with protonated glucose results in more deuteration at C^{α} protons than in the side-chains, allowing all side-chain protons, including methyl groups, to be efficiently correlated to backbone amide signals via deuterated α-carbons. Thus, the simple labeling methods and the NMR experiments described here have some of the benefits of Gardner's experiment (Gardner et al., 1996), but also provide additional information.

In partially and fully deuterated samples, the ^{13}C isotope shift induced by deuteration can create potential problems for spectral resolution. However, we did not observe adverse effects of the isotope shift, probably because side-chain carbon and proton signals are recorded in the indirect dimensions, making digitization the limiting factor for resolution.

An efficient way to achieve backbone and side-chain assignments simultaneously

The high cost of isotope labeling has been a limitation for structural studies using NMR, making it desirable to solve a structure using a smaller number of expensively labeled samples. With the approach described here, only two ^{15}N and ^{13}C labeled samples, one partially (65% to 90%) deuterated and the other protonated, will be needed. Complete backbone and side-chain assignments can be obtained from H(CC-CO)NH-TOCSY, (H)C(C-CO)NH-TOCSY, and ^{13}C-NOESY-HSQC spectra. The ^{13}C-NOESY-HSQC simultaneously provides some of the distance restraints needed for complete structure determination.

This side-chain assignment strategy has the additional advantage that it can easily be incorporated into automated assignment programs, which typically make use of the well-dispersed backbone NH resonances for backbone assignment.

Furthermore, the dramatic gain in sensitivity resulting from partial deuteration is also seen in the related HCCNH-TOCSY experiments (Figures 1c, 1d and 6), which correlate the side-chain resonances with the backbone amide group of the same residue as well as the following one. A combination of these two pairs of experiments may then be sufficient for both backbone and side-chain assignment. Residue type information, which is essential for sequential backbone

assignment, is provided by characteristic side-chain carbon chemical shifts, so there is no need for specifically labeled samples. We actually detected a DNA sequencing error in the original tGCN5 sequence from these experiments (Lin et al., 1999).

Thus, it seems possible to determine a protein structure using only the following experiments: HNCA, H(CC-CO)NH-TOCSY, (H)C(C-CO)NH-TOCSY, (H)C(C)NH-TOCSY, ^{13}C-NOESY-HSQC, ^{15}N-NOESY-HSQC and 2D NOESY. HNCA is needed to identify the C^{α} signals which are weak in the experiments presented here. The first six experiments are used for assignment, while the last three are used for obtaining NOE distance restraints. This set of seven experiments may well be sufficient for characterizing the overall fold of a protein and may be of interest for applications in structural genomics.

Conclusions

We have presented a strategy for complete side-chain assignment using optimized H(CC-CO)NH-TOCSY and (H)C(C-CO)NH-TOCSY experiments on a partially deuterated sample, in combination with a 3D ^{13}C-dispersed NOESY-HSQC. This strategy benefits from the nonuniform deuteration pattern generated by the glucose metabolic pathway in *E. coli*. There is also a dramatic gain in sensitivity from partial deuteration. These experiments alleviate difficulties encountered during assignment using conventional HCCH-TOCSY experiments. Because of the well-dispersed backbone NH resonances and well-characterized side-chain carbon chemical shifts that depend upon the residue type, the assignment procedure is efficient and reliable. Extending this strategy can make it possible to determine a structure with a limited number of experiments.

Acknowledgements

We thank Mr. Greg Heffron for help with the spectrometers and Dr. C.M. Fletcher and Dr. A.E. Ferentz for proofreading the manuscript. This work was supported by grants from NSF (MCB-9816072) and NIH (GM47467). Acquisition and maintenance of spectrometers and computers used for this work was supported by NSF (MCB 9527181) and the Harvard Center for Structural Biology and the Giovanni Armenise-Harvard Foundation for Advanced Scientific Research.

References

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.

Brünger, A.T. (1992) *X-PLOR Version 3.1*, Yale University Press, New Haven, CT and London.

Farmer II, B.T. and Venters, R.A. (1995) *J. Am. Chem. Soc.*, **117**, 4187–4188.

Fesik, S.W. and Zuiderweg, E.R.P. (1990) *Quart. Rev. Biophys.*, **23**, 97–131.

Fischer, M.W.F., Zeng, L. and Zuiderweg, E.R.P. (1996) *J. Am. Chem. Soc.*, **118**, 12457–12458.

Gardner, K.H., Konrat, R., Rosen, M.K. and Kay, L.E. (1996) *J. Biomol. NMR*, **8**, 351–356.

Grzesiek, S., Anglister, J. and Bax, A. (1993a) *J. Magn. Reson.*, **B101**, 114–119.

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993b) *J. Am. Chem. Soc.*, **115**, 4369–4370.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Gschwind, R.M., Gemmecker, G. and Kessler, H. (1998) *J. Biomol. NMR*, **11**, 191–198.

Kay, L.E., Ikura, M. and Bax, A. (1990) *J. Am. Chem. Soc.*, **112**, 888–889.

Kay, L.E., Keiffer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.

Kay, L.E., Xu, G., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) *J. Magn. Reson.*, **B101**, 333–337.

Kupce, E. and Wagner, G. (1995) *J. Magn. Reson.*, **B109**, 329–333.

Kupce, E. and Wagner, G. (1996) *J. Magn. Reson.*, **B110**, 309–312.

Kupce, E. and Freeman, R. (1996a) *J. Magn. Reson.*, **A118**, 299–303.

Kupce, E. and Freeman, R. (1996b) *Chem. Phys. Lett.*, **250**, 523–527.

Kupce, E., Matsuo, H. and Wagner, G. (1999) In *Biological Magnetic Resonance: Modern Techniques in Protein NMR*, Vol. 16 (Eds, Krishna and Berliner), Kluwer Academic/Plenum Publishers, Dordrecht, pp. 149–193.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry*, **27**, 142–150.

LeMaster, D.M. (1994) *Prog. NMR Spectrosc.*, **26**, 317–419.

Lin, Y., Fletcher, C.M., Zhou, J., Allis, C.D. and Wagner, G. (1999) *Nature*, **400**, 86–89.

Logan, T.M., Olejniczak, E.T., Xu, R.X. and Fesik, S.W. (1993) *J. Biomol. NMR*, **3**, 225–231.

Lyons, B.A. and Montelione, G.T. (1993) *J. Magn. Reson.*, **B101**, 206–209.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1998a) *J. Am. Chem. Soc.*, **111**, 1515–1517.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989b) *J. Magn. Reson.*, **85**, 393–399.

Matsuo, H., Li, H. and Wagner, G. (1996a) *J. Magn. Reson.*, **B110**, 112–115.

Matsuo, H., Kupce, E., Li, H. and Wagner, G. (1996b) *J. Magn. Reson.*, **B111**, 194–198.

Matsuo, H., Kupce, E., Li, H. and Wagner, G. (1996c) *J. Magn. Reson.*, **B113**, 91–96.

McCoy, M.A. and Mueller, L. (1992) *J. Am. Chem. Soc.*, **114**, 2108–2112.

Morris, G.A. and Freeman, F. (1979) *J. Am. Chem. Soc.*, **101**, 760–762.

Muhandiram, D.R. and Kay, L.E. (1994) *J. Magn. Reson.*, **103**, 203–216.

Nietlispach, D., Clowes, R.T., Broadhurst, W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P.J. and Laue, E.D. (1996) *J. Am. Chem. Soc.*, **118**, 407–415.

Pervushin, K., Riek, R., Wider, G. and Wüthrich, K. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.

Shaka, A. (1985) *Chem. Phys. Lett.*, **120**, 201–205.

Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H. and Kay, L.E. (1996) *J. Am. Chem. Soc.*, **118**, 6570–6579.

Stryer, L. (1988) *Biochemistry*, W.H. Freeman and Co., New York, NY.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.