Published in final edited form as:
J Am Chem Soc. 2006 May 3; 128(17): 5757–5763. doi:10.1021/ja0584222.

Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins

Dominique P. Frueh[†], **Zhen-Yu J. Sun**[†], **David A. Vosburg**[†,‡], **Christopher T. Walsh**[†], **Jeffrey C. Hoch**[§], and **Gerhard Wagner**[†,*]

†*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA; Fax:617-432-3283, Phone: 617-432-3213*

§*Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, Farmington, CT 06030, USA*

Abstract

The initial step of protein NMR resonance assignments typically identifies the sequence positions of ^{1}H-^{15}N HSQC cross peaks. This is usually achieved by tediously comparing strips of multiple triple-resonance experiments. More conveniently, this could be obtained directly with hNcaNH and hNcocaNH-type experiments. However, in large proteins and at very high fields, rapid transverse relaxation severely limits the sensitivity of these experiments, and the limited spectral resolution obtainable in conventionally recorded experiments leaves many assignments ambiguous. We have developed alternative hNcaNH experiments that overcome most of these limitations. The TROSY technique was implemented for semi-constant time evolutions in *both* indirect dimensions, which results in remarkable sensitivity and resolution enhancements. Non-uniform sampling in both indirect dimensions combined with Maximum Entropy (MaxEnt) reconstruction enables such dramatic resolution enhancement while maintaining short measuring times. Experiments are presented that provide either bidirectional or unidirectional connectivities. The experiments do not involve carbonyl coherences and thus do not suffer from fast chemical shift anisotropy-mediated relaxation otherwise encountered at very high fields. The method was applied to a 300 μM sample of a 37 kDa fragment of the *E. coli* enterobactin synthetase module EntF, for which high-resolution spectra with an excellent signal-to-noise ratio were obtained within four days each.

Keywords

protein backbone assignment; nuclear magnetic resonance; NRPS; non-uniform sampling; maximum entropy reconstruction; TROSY; enterobactin synthetase

Introduction

Assignment of the NMR signals of a protein is a prerequisite for any structural or functional studies. In large proteins, the complexity of the spectra requires the acquisition of many experiments to resolve ambiguities for backbone assignment.[1, 2] The hNcaNH [3] and hNcocaNH [4, 5] experiments provide direct connectivities in the HN-HSQC spectrum of the molecule, and are thus very useful additions to the set of back-bone experiments typically recorded.[6] The hNcaNH experiment provides correlations between coherences of the backbone

*E-mail: Gerhard_Wagner@hms.harvard.edu.
‡Current address: Department of Chemistry, Harvey Mudd College, Claremont, CA 91711, USA

Contribution from Harvard Medical School, Boston, MA 02115, USA

amide protons and nitrogens in two dimensions and provides cross-peaks of succeeding, preceding and intra-residue signals in a third dimension. The hNcocaNH experiment only features cross-peaks with succeeding residues. Unfortunately, these experiments feature several long transfer times and, for large proteins, the sensitivity is rapidly deteriorated by fast relaxation, even with perdeuteration. In addition, spectral crowding in the ^{15}N dimensions implies the need for high-resolution spectra, which is usually prevented by limitations to experimental time available, or due to sample instability. Thus, on large proteins, no satisfactory data can be obtained within reasonable time for these experiments.

We present modifications of the hNcaNH experiment [3, 5] that feature the TROSY (Transverse Relaxation Optimized Spectroscopy) technique to select the slowly relaxing component of the nitrogen and amide proton multiplet [7] and take advantage of non-uniform sampling [8-12] to optimize sensitivity and resolution. Although it is straightforward to apply the TROSY technique to the directly observed signals of amide protons H^N nuclei and their attached nitrogens, dramatic improvements are obtained when *both* ^{15}N evolution periods benefit from the TROSY technique as described here in our TROSY-hNcaNH-TROSY experiment. This is particularly important as the corresponding dimension provides the cross-peaks used to establish the sequential assignments. Our design makes use of the in-phase antiphase (IPAP) technique [13] to select the slowly relaxing doublet component during the evolution of the nitrogen from which the coherence transfer was originated. The chain elongations obtained with the TROSY-hNcaNH-TROSY experiment extend symmetrically towards both the N- and C-termini. Another experiment, a TROSY-s-hNcaNH-TROSY, has been designed to provide only connectivities with the succeeding residue of a given amino acid, which simplifies the spectrum and resolves ambiguities. Although the same information can be obtained with an hNcocaNH experiment, the TROSY-s-hNcaNH-TROSY experiment does not use transfer through the carbonyl carbon to avoid rapid relaxation at high magnetic fields and for large proteins.

All experiments, including a reference hNcaNH-TROSY, make use of non-uniform sampling (NUS) of the evolution times in both indirect dimensions. By tailoring the sampling schedule to the decay of the signals of ^{15}N coherences during a semi-constant time evolution, relaxation losses are minimized, while the optimal resolution can be attained in a minimum of measurement time. The time saved is then used to increase the number of transients accumulated for each FID (Free Induction Decay). The experiments have been tested on a challenging 300 μM sample of the 37 kDa T-TE didomain of the four-domain protein EntF. The EntF module is part of the *E.coli* enterobactin non ribosomal peptide synthetase (NRPS), a protein currently under structural investigation.[14] We describe several ways of using the data obtained to facilitate and accelerate the assignment of the H^N and N resonances.

Material and methods

A pET30a+ plasmid containing the gene coding for the N-terminal His_6-tagged protein was transformed into *E. coli* BL21(DE3) cells for protein expression. A (^{15}N,^{2}H,^{13}C) uniformly labelled sample was prepared by over-expression in M9 minimal media in D_2O containing ^{2}H-^{13}C glucose and $^{15}NH_4Cl$. The cells were allowed to grow for 24 h at 37 °C. After induction, the cell culture was cooled to 25 °C and allowed to grow for 3 h. The cells were pelleted, resuspended, and lysed with a French press. The protein was purified using Ni-NTA resin (Qiagen) followed by FPLC using a Sephadex gel-filtration column (S75). The sample was concentrated to a final concentration of 300 μM in 20 mM phosphate (pH = 6.7) 150 mM NaCl with 1 mM EDTA and DTT.

The data were recorded at 25 °C on a 750 MHz Bruker spectrometer equipped with a cryoprobe®. For all three hNcaNH experiments, the spectral widths were 16.033 ppm for

proton (ω_3, centred on water resonance at 4.690 ppm) and 35 ppm for nitrogen (ω_1 and ω_2 centred at 118 ppm). A total of 512 complex points randomly selected from a matrix of 100 × 100 complex points were sampled for the two indirect dimensions, a number substantially greater than the number of expected cross peaks for a single N-N plane, and thus more than sufficient to reconstruct the peaks. The random distribution was weighted by an exponential to account for the relaxation of the protein. The sampling schedule was generated with the samp2sched program included in the Rowland NMR toolkit package.[15] For Bruker instruments, this schedule is converted to a vclist, which the pulse program uses to control delays and phases incrementations. A recycling delay of 1 s was used and 128 scans were accumulated per increment. The total measuring time of each experiment was 3 days and 18 hours. The data were processed using maximum entropy (MaxEnt) reconstruction as implemented in the Rowland NMR Toolkit msa2d program.[15] The processing time (using the parameters DEF = 10 and LAMBDA = 0.5) was approximately 21 minutes for each spectrum on a Dell computer with dual Xenon 3.0 GHz processors running under the Fedora Core 1 Linux operating system. The final matrix size was 256 (ω_1) × 256 (ω_2) × 450 (ω_3, NH region only) points. The resulting NMR spectra were analyzed using the software CARA (available from http://www.nmr.ch/). [16]

Results and Discussion

Pulse Sequence

Figure 1a shows the hNcaNH-TROSY pulse sequence. The experiment is a straightforward modification of the well known hNcaNH,[3] with a gradient selection sensitivity enhanced HN-TROSY. It also features semi-constant time evolution periods in both ^{15}N dimensions to optimize resolution and sensitivity by using non-uniform sampling with subsequent Maximum Entropy reconstruction. Thus, the experiment will only be summarized to allow for comparisons with the modifications described bellow.

In the conventional hNcaNH-TROSY pulse sequence of Fig. 1a, a first INEPT between points a and b creates nitrogen single-quantum coherence (SQC), which evolves under both scalar couplings J(NH) and J(NC$^\alpha$) to become antiphase with respect to C$^\alpha$ and in-phase with respect to H^N:

$$\sigma_b = -2N_{y_i}H^N_{z_i} \rightarrow \sigma_c = 2N_{y_i}C^\alpha_{z_{i-1}} + 2N_{y_i}C^\alpha_{z_i} \quad (1)$$

The signal is encoded with ^{15}N chemical shifts in a semi-constant time fashion [21] extending the maximum length of the period b-c from $2T_{N'}$ = 24.8 ms to 33.2 ms. After the coherence transfer from nitrogen to carbon between c and d, the two types of C$^\alpha$ SQCs evolve under both one and two-bond scalar couplings J(NC$^\alpha$) during the period of length $2T_C$, to become antiphase with respect to different nitrogens:[5]

$$\sigma_d = 2C^\alpha_{y_i}N_{z_i} + 2C^\alpha_{y_{i-1}}N_{z_i} \rightarrow \sigma_e = 2C^\alpha_{y_i}N_{z_i} + 2C^\alpha_{y_{i-1}}N_{z_i} - 2C^\alpha_{y_{i-1}}N_{z_{i-1}} - 2C^\alpha_{y_i}N_{z_{i+1}} + \ldots \quad (2)$$

where "…" indicates non-detected terms of the density operator. After the coherences are converted to nitrogen SQC, they are refocused with respect to C$^\alpha$ and frequency labeled with ^{15}N chemical shifts during a semi-constant time evolution period (f-g). Finally, the high frequency component of the nitrogen SQC is converted to the low frequency component of the attached proton coherence by using a gradient-selection sensitivity enhanced TROSY scheme.[22, 23]

Non-uniform sampling was used during both semi-constant time evolution periods in order to be able to sample out to long maximum evolution times, with a limited number of sampling

points. This allows accumulation of more scans per increment and increases the signal-to-noise ratio of the data acquired at each increment. The sampling schedule generated by the program samp2sched used exponentially increasing intervals between sample times, with damping factors corresponding to 60 s^{-1} during b-c and to 30 s^{-1} during f-g where the TROSY technique is applied. This ensures that more points are sampled for short values of t_1 and t_2, where the signals are intense, than for long values of t_1 and t_2, where most of the signal has relaxed. Note that the use of 2D Maximum Entropy Reconstruction for the processing of the non-uniformly acquired FIDs in the indirect dimensions requires that the number of samples exceed by a small factor the number of expected cross-peaks in the N-N planes (see Material and methods, above). This condition is largely fulfilled with the 512 sampled complex points in our experiment.

Each finally detected $N_i(\omega_2)$-$H_i^N(\omega_3)$ cross peak is derived from amide proton coherences of residues i, i-1 and i+1 and frequency labelled during t_1 with the nitrogen frequencies of the three residues. Thus, by repeating the analysis described above three times and collect the relevant terms, we can summarize the coherence pathways as:

$$\left.\begin{array}{l} H^N_{i-1} \rightarrow N_{i-1}(t_1) \rightarrow C^{\alpha}_{i-1} \\ H^N_i \rightarrow N_i(t_1) \left\{ \begin{array}{l} \rightarrow C^{\alpha}_{i-1} \\ \rightarrow C^{\alpha}_i \end{array} \right. \\ H^N_{i+1} \rightarrow N_{i+1}(t_1) \rightarrow C^{\alpha}_i \end{array}\right\} \rightarrow N_i(t_2) \rightarrow H^N_i(t_3) \quad (3)$$

As shown in Figure 2, each intra-residue nitrogen-proton correlation in the ω_2 and ω_3 dimensions will be correlated in ω_1 to the nitrogens of both the preceding and the succeeding amino acids (cross-peaks), in addition to its own nitrogen (autocorrelation or diagonal peak). As can be seen from Eq. (2), the two sequential cross-peaks have an opposite phase to the intraresidue diagonal peak. The relative intensities of the cross-peaks depend on the values of the scalar couplings $J(NC^{\alpha})$ involved in the transfer periods b-c, d-e and f-g and on the relaxation properties of the various coherences that take place during these transfers. A C^{α} selective 180° pulse with a bandwidth of 22 ppm was used during the $2T_C$ period (d-e) to refocus evolution under $J(C^{\alpha}C^{\beta})$. The C^{β} of Ser and of a small number of Leu and Thr are not inverted during the period (d-e) and the corresponding signals may thus have opposite phases. Alternatively, evolution under the $C^{\alpha}C^{\beta}$ scalar coupling can be refocused by setting by setting the $2T_C$ transfer time to a multiple of $1/JC^{\alpha}C^{\beta}$ (ca 28 ms) at the cost of a non optimum transfer time. The C^{α} relaxation times can be roughly estimated by monitoring the decay of H^N-N cross-peaks in a modified 2D-HN(CO)CA experiment with a spin echo during a C^{α} SQC period. The average rate $R_2 = 1/T_2$ is 18 s^{-1}, which is higher than the predicted 10 s^{-1} for a protein of this size (calculated assuming a deuterated rigid protein with isotropic tumbling, τ_C = 17 ns, and accounting for the neighbouring amide protons). This is mainly due to a number of outlier residues, probably subject to conformational exchange, which have an average rate of 40 s^{-1}. The predicted rate for a protein of this size (c.a. 10 s^{-1}) would lead to an optimum transfer time of 48 ms, whereas a rate of 40 s^{-1} corresponds to an optimum value of 32 ms. However, at 32 ms, the small signals of fast relaxing residues are enhanced by 30 % compared to their intensities at 48 ms, whereas the more intense slow relaxing signals, with rates of 10 s^{-1}, suffer little from a 20% decrease in intensity (see supporting information, Figure S2). We thus chose to collect our data with a transfer time optimum for fast relaxing signals. Simulations of the transfer efficiencies for various relaxation rates are presented in the supporting information. For relaxation rates between 0 and 50 s^{-1}, the optimum transfer time, $2T_C$*, can be estimated from the transverse relaxation rate R_2 by the empirical relation (see supporting information Figure S3):

$$2T_C^* = 0.052 - 4.8 \cdot 10^{-4} R_2 \quad (4)$$

Where R_2 is in rad/s and T_C is in s.

While the experiment described so far uses a single TROSY module at the detection end of the pulse sequence, it would be desirable to benefit from TROSY effects during the two ^{15}N evolution periods. Since hNcaNH experiments involve out-and-stay type of coherence transfers, the TROSY selection employed during t_2 and t_3 evolutions cannot readily be applied to the t_1 evolution period to select the narrow component of the ^{15}N doublet as in conventional out-and-back type of experiments such as HNCA. Hence, this non-TROSY t_1 evolution period results in sensitivity losses and poor resolution in the ω_1 dimension, which is used for making the sequential connectivities.

Further improvement in both resolution and sensitivity can be achieved by designing an additional TROSY scheme for the ω_1 dimension. In brief, this is achieved by eliminating the two 180° pulses on the proton channel between points b and c (compare Fig. 1a and 1 b), which retains the ^1H-^{15}N coupling during t_1, and by selecting the sharp nitrogen component with phase alteration of the ^1H and ^{15}N pulses at point b. Figure 1b shows the scheme used to select the slowly relaxing component of the nitrogen multiplet in ω_1 with an in-phase anti-phase (IPAP) [13] strategy. Antiphase ^{15}N SQC is combined with in-phase SQC to obtain the desired low-field component. This is done by cycling the phases ϕ_1 and ϕ_7 simultaneously:

$$\begin{aligned}\phi_1=x:\sigma_b=-2N_{y_i}H^N_{z_i} &\xrightarrow{\phi_7=x,(2\varepsilon)J(NH)} [-N_{x_i}]\\ &\xrightarrow{\varpi_N t_1, J(NH)t_1, 2T_{N'}J(NC^\alpha)} \sigma_c=(2N_{y_i}C^\alpha_{z_{i-1}}+2N_{y_i}C^\alpha_{z_i})\cos(\omega_N t_1)\cos(\pi J t_1)\end{aligned} \quad (5)$$

and

$$\begin{aligned}\phi_1=y:\sigma_b=-2N_{x_i}H^N_{z_i} &\xrightarrow{\phi_7=-x,(\varepsilon-\varepsilon+T_{N'}-T_{N'})J(NH)} \left[-2N_{x_i}H^N_{z_i}\right]\\ &\xrightarrow{\varpi_N t_1, J(NH)t_1, 2T_{N'}J(NC^\alpha)} \sigma_c=(2N_{y_i}C^\alpha_{z_{i-1}}+2N_{y_i}C^\alpha_{z_i})\sin(\omega_N t_1)\sin(\pi J t_1)\end{aligned} \quad (6)$$

where $\omega_N t_1$ and $J(NH)t_1$ indicate evolution under the nitrogen chemical shift and under the J (NH) scalar coupling during t_1, $2T_{N'}J(NC^\alpha)$ denotes evolution under $J(NC^\alpha)$ during $2T_{N'}$, $2\varepsilon J$ (NH) represents the evolution under J(NH) during 2ε and $(\varepsilon\text{-}\varepsilon+T_{N'}\text{-}T_{N'})$ depicts refocusing of the evolution under J(NH). The term in squared-brackets is an intermediate state. The phases are those used for our Bruker spectrometer, where y actually corresponds to a -y rotation for nuclei with negative gyromagnetic ratios.[24] Subtraction of Eq. 6 from Eq. 5, gives:

$$\begin{aligned}\sigma_c&=\left(2N_{y_i}C^\alpha_{z_{i-1}}+2N_{y_i}C^\alpha_{z_i}\right)(\cos(\omega_N t_1)\cos(\pi J t_1)-\sin(\omega_N t_1)\sin(\pi J t_1))\\ &=\left(2N_{y_i}C^\alpha_{z_{i-1}}+2N_{y_i}C^\alpha_{z_i}\right)\cos((\omega_N+\pi J)t_1)\end{aligned} \quad (7)$$

This is the high frequency component of the doublet in the nitrogen dimension. The data subtraction is achieved by negating the receiver phase concomitantly with the phase cycling of ϕ_1 and ϕ_7 in the pulse program. The imaginary part is simply obtained by applying the States-TPPI method [18] on phase ϕ_1. Alternatively, the selection could be achieved by using the S^3 approach,[25] similarly to what has been used for the methylene TROSY.[26]

As mentioned above, the two hNcaNH spectra (with or without the double TROSY) contain correlations with both the following and the preceding residues. Thus, the directionality of the chain extension cannot be achieved with this experiment alone. The NUS-hNcocaNH experiment [6] can be modified to incorporate the double TROSY technique. However, the experiment is then subject to carbonyl relaxation, which leads to losses in sensitivity, in particular for large proteins and when recorded at high magnetic fields.

To overcome this problem, we have developed an experiment where selection of one of the two sequential pathways in Eq. 3 is achieved in a similar way as in some modifications of the HNCA experiment [27-30] involving, however, different refocusing schemes. This is illustrated in Figure 2, in which the grey arrows indicate pathways that have to be suppressed. The method relies on the fact that during the periods $T_{N'}$ and $2T_C$, pairs of nitrogens and C^α that are involved in a given pathway are scalar coupled with the same carbonyl carbons, whereas those involved in the other pathway are not. We call this a TROSY-s-hNcaNH-TROSY experiment (where the "s", indicates selection of the pathway originating from the *succeeding* residue) and the modifications of the parent pulse sequence (Figure 1a) for the periods b-c and d-e are shown in Figure 1c. No carbonyl coherences are involved in the pathway selection of this experiment. During the first semi-constant time evolution, ^{15}N SQC is allowed to evolve under J(NC′) to become antiphase with respect to the carbonyl carbon:

$$\sigma_b = -2N_{y_i}H^N_{z_i} \rightarrow \sigma_c = 4N_{x_i}C^\alpha_{z_{i-1}}C'_{z_{i-1}} + 4N_{x_i}C^\alpha_{z_i}C'_{z_{i-1}} \quad (8)$$

During the period (d-e), ^{13}C SQC evolution under J(C^αC′) brings the coherence of the preceding C^α in-phase with respect to carbonyl and further defocuses the coherence of the intra residue C^α:

$$\begin{aligned}\sigma_d &= -4C^\alpha_{y_i}N_{z_i}C'_{z_{i-1}} - 4C^\alpha_{y_{i-1}}N_{z_i}C'_{z_{i-1}} \\ \rightarrow \sigma_e &= -8C^\alpha_{x_i}N_{z_i}C'_{z_{i-1}}C'_{z_i} - 2C^\alpha_{x_{i-1}}N_{z_i} + 2C^\alpha_{x_{i-1}}N_{z_{i-1}} + 8C^\alpha_{x_i}N_{z_{i+1}}C'_{z_{i-1}}C'_{z_i} + \ldots\end{aligned} \quad (9)$$

Only the second and third terms of Eq. 9 are detected after the periods f-g and g-h. Thus, only a signal originating from the succeeding residue will appear together with a weaker intra-residue signal in the ω_1 dimension of the spectrum (see again Figure 2 for a schematic representation):

$$\left.\begin{aligned}H^N_i &\rightarrow N_i(t_1) \rightarrow C^\alpha_i \\ H^N_{i+1} &\rightarrow N_{i+1}(t_1) \rightarrow C^\alpha_i\end{aligned}\right\} \rightarrow N_i(t_2) \rightarrow H^N_i(t_3) \quad (10)$$

Similarly, in a TROSY-p-hNcaNH-TROSY experiment, the connection to the *preceding* residues alone can also be obtained by first defocusing carbon coherences between points d and e, and refocusing nitrogen with respect to both C^α and C′ during the period f-g. Such an experiment has been designed, but no data set was actually recorded since only one of the selective TROSY-hNcaHN-TROSY experiments is sufficient to provide the directionality information during chain elongation.

Data Analysis

Non-uniform sampling allowed us to acquire data to very long evolution times in both indirect dimensions and to fully benefit from the slow transverse relaxation of the coherences selected with the pulse sequences described. The data cannot be transformed with the discrete Fourier transformation and were thus processed by using Maximum Entropy (MaxEnt) reconstruction[15], which causes some non-linearity of peak intensities. Without constructing a calibration curve, or computing the reconstruction using parameters that minimize the nonlinearity at the expense of sensitivity,[31] only a qualitative comparison can be estimated from the traces, as shown in Figure 3. In the judgement of the overall quality of the experiments presented here, we will rely on the number of connectivities that could be confirmed amongst previously assigned residues. These reflect the threshold for which a signal can be identified in a spectrum. In the analysis following below, connectivities to prolines and non-exchangeable amide protons, which are obviously both absent from the spectra, were not counted.

Figure 3 shows an example of strips obtained with the three variations of the hNcaNH experiment. Even without the TROSY in ω_1, a fair number of signals can be detected. However, as can be seen in Figure 3a, the lower intensity signals are often absent or hard to distinguish from the noise (as seen for L102, G105, and N106). Consequently, only 36 % of the possible connectivities could be established in the single-TROSY version (Fig. 1a). In contrast, implementation of the double-TROSY technique greatly improves the quality of the spectrum, as can be seen in Figure 3b where all connections are observed in the strips displayed. Thus, in the 37 kDa fragment of the EntF protein studied here, 85% of the assignable residues could be linked, of which 51% contained connections to both successors and predecessors. The redundant information contained in the spectrum (the presence of cross-peaks originating both from succeeding and preceding residues) provides an additional robustness since only one of the signals is sufficient to establish a new connectivity. As mentioned earlier, an inversion of the phase of one or more cross-peaks can occur depending on the composition of a triplet of connected residues, which can be exploited to facilitate the assignment. The high resolution provided by the experiment, which leads to precise values of the nitrogen chemical shifts, often enables one to directly identify the signal of a neighbouring residue in an HSQC during the initial stage of sequential assignment. However, it remains uncertain whether a cross-peak belongs to a successor or a predecessor with only this experiment. This information is provided by the TROSY-s-hNcaNH-TROSY experiment where only cross-peaks of the succeeding residues are selectively detected. The spectrum of the TROSY-s-hNcaNH-TROSY experiment is much simplified since only one of the sequential cross peaks is present and the uninformative intra-residual diagonal peaks are of lower intensities. At the outset it was expected that the sensitivity of the cross peaks would be the same as for the non-selective experiment, since the length of the pulse sequence remained the same and since no additional coherences are created, in contrast to inter- or intra-HNCA experiments. However, some losses of signal intensities have been observed (Figure 3b and c). These are probably due to the fact that additional pulses are required for the selection and because there is a slight mismatch between transfers required for evolutions under J(NC′) and J(NC^{α}) scalar couplings. The latter effect could be recovered by shifting the position of the first C′ inversion pulse with respect to the C^{α} pulse. Strips of the resulting spectrum are shown in Figure 3c. A total of 66 % of the known connections were obtained with this experiment alone. Compared to the non-selective experiment, the losses are due to the above-mentioned reduction in sensitivity and the absence of cross-peaks to the preceding residue, which are sometimes the only link between two residues. Nonetheless, the directional information provided by the TROSY-s-hNcaNH-TROSY experiment is valuable for resolving some ambiguities in sequential assignments. In particular, note that information from a single residue is sufficient to determine the directionality of a given fragment.

The validation of the connections established with the three hNcaNH experiments is based on a comparison with assignments obtained previously with more conventional methods. In order to obtain assignments for a previously unassigned protein, a strip comparison is efficient when both cross-peaks are present in each strip of a pair of residues in the non selective TROSY-hNcaNH-TROSY. In other situations, the spectra are best used in conjunction with a pair of HNCA and HNCOCA experiments (or their counterpart for C^{β} and C′ nuclei). The TROSY-hNcaNH-TROSY provides the nitrogen chemical shift of the preceding residue (identified with help of the TROSY-s-hNcaNH-TROSY) while an HNCOCA provides its C^{α} chemical shift. The procedure can also be adapted to the determination of the successor by first looking at the intra-residue signal of an HNCA. A cross-peak with the same C^{α} chemical shift will then be present in the corresponding H-N plane of the HNCOCA, at the ^{15}N chemical shift predicted by the hNcaNH. The combined information is equivalent to the one provided by a 4D hNCANH or hNcoCANH and greatly reduces ambiguities in sequential assignment. In an alternative but more practical approach, the hNcaNH data enables one to identify the correct candidate in a C^{α} (or C^{β} or C′) strip comparison by comparing their ^{15}N chemical shifts with the one provided by the hNcaNH experiment. Instead of comparing candidates from various independent C^{α},

C^{β} and C′ strips matching, the best candidate can be quickly identified *within* any of these strips comparison by combining with the hNcaNH correlations. This results from the fact that the sequential information supplied by the hNcaNH is provided by ^{15}N nuclei, which are correlated in each of the HNCA, HNCACB and HNCACO spectra (and their inter-residue analogues). Any of these strategies greatly accelerates the sequential assignment when compared to a convolution of strip comparisons resulting from different nuclei.

Conclusion

The variants of the hNcaNH experiments developed enabled us to record spectra on a 37 kDa protein at a fairly low concentration. Significant improvements were obtained by designing a TROSY sequence applied to the first indirect dimension in addition to the conventional TROSY implementation. This improved the number of achievable connections from a mere 36% of the detected signals to 85%. Additional information can be obtained by recording double TROSY versions of the hNcocaNH experiment or the newly designed s-hNcaNH where only signals with successors are present. Thus, all the benefits from the original experiments can now be applied to large proteins. Together with the conventional back-bone experiments, these provide a powerful tool for sequential assignment of large proteins at high magnetic fields. In particular, at the early stage of sequential assignment, the ^{15}N correlation information provided by hNcaNH can be directly combined with any of the backbone experiments so that the spin-system chains can be rapidly extended. In the later stage of connecting the chain fragments, the additional ^{15}N correlation is welcome information for resolving ambiguities due to spectral overlap. The pulse sequences, together with the rnmrtk script to process the data, can be downloaded at http://gwagner.med.harvard.edu.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the National Institutes of Health (grants GM47467, AI37581 and RR00995) and a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund (D.A.V.).

References

1. Ferentz AE, Wagner G. Q Rev Biophys 2000;33(1):29–65. [PubMed: 11075388]
2. Tugarinov V, Hwang PM, Kay LE. Annu Rev Biochem 2004;73:107–46. [PubMed: 15189138]
3. Weisemann R, Ruterjans H, Bermel W. J Biomol NMR 1993;3(1):113–20. [PubMed: 8448431]
4. Grzesiek S, Bax A. J Am Chem Soc 1993;115:4369.
5. Panchal SC, Bhavesh NS, Hosur RV. J Biomol NMR 2001;20(2):135–47. [PubMed: 11495245]
6. Sun ZY, Frueh DP, Selenko P, Hoch JC, Wagner G. J Biomol NMR 2005;33(1):43–50. [PubMed: 16222556]
7. Pervushin K, Riek R, Wider G, Wuthrich K. Proc Natl Acad Sci U S A 1997;94(23):12366–71. [PubMed: 9356455]
8. Marion D. J Biomol NMR 2005;32(2):141–50. [PubMed: 16034665]
9. Orekhov VY, Ibraghimov IV, Billeter M. J Biomol NMR 2001;20(1):49–60. [PubMed: 11430755]
10. Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. J Am Chem Soc 2005;127(8):2767–75. [PubMed: 15725035]
11. Rovnyak D, Frueh DP, Sastry M, Sun ZY, Stern AS, Hoch JC, Wagner G. J Magn Reson 2004;170 (1):15–21. [PubMed: 15324754]
12. Schmieder P, Stern AS, Wagner G, Hoch JC. J Biomol NMR 1993;3(5):569–76. [PubMed: 8219741]
13. Ottiger M, Delaglio F, Bax A. J Magn Reson 1998;131(2):373–8. [PubMed: 9571116]

14. Roche ED, Walsh CT. Biochemistry 2003;42(5):1334–44. [PubMed: 12564937]
15. Hoch, JC.; Stern, AS. NMR data processing. Wiley-Liss; New York: 1996.
16. Keller, RLJ. The Computer Aided Resonance Assignment Tutorial. Cantina Verlag; Goldau: 2004.
17. Emsley L, Bodenhausen G. J Magn Reson 1992;97:135–148.
18. Marion D, Ikura M, Tschudin R, Bax A. J Magn Reson 1989;85:393–399.
19. Kay LE, Keifer P, Saarinen T. J Am Chem Soc 1992;114:10663–10665.
20. Shaka AJ, Keeler J, Frenkiel T, Freeman R. J Magn Reson 1983;52:335–338.
21. Grzesiek S, Bax A. J Biomol NMR 1993;3:185–204. [PubMed: 8477186]
22. Loria JP, Rance M, Palmer AG 3rd. J Magn Reson 1999;141(1):180–4. [PubMed: 10527755]
23. Salzmann M, Wider G, Pervushin K, Wuthrich K. J Biomol NMR 1999;15(2):181–4. [PubMed: 10605091]
24. Roehrl MH, Heffron GJ, Wagner G. J Magn Reson 2005;174(2):325–30. [PubMed: 15862251]
25. Meissner A, Duus JO, Sørensen OW. J Biomol NMR 1997;10:89–94. [PubMed: 9453800]
26. Miclet E, Williams DC Jr, Clore GM, Bryce DL, Boisbouvier J, Bax A. J Am Chem Soc 2004;126 (34):10560–70. [PubMed: 15327312]
27. Meissner A, Sorensen OW. J Magn Reson 2001;150(1):100–4. [PubMed: 11330987]
28. Brutscher B. J Magn Reson 2002;156(1):155–9. [PubMed: 12081454]
29. Nietlispach D, Ito Y, Laue ED. J Am Chem Soc 2002;124(37):11199–207. [PubMed: 12224968]
30. Permi P, Annila A. J Biomol NMR 2001;20(2):127–33. [PubMed: 11495244]
31. Schmieder P, Stern AS, Wagner G, Hoch JC. J Magn Reson 1997;125:332–339. [PubMed: 9144266]

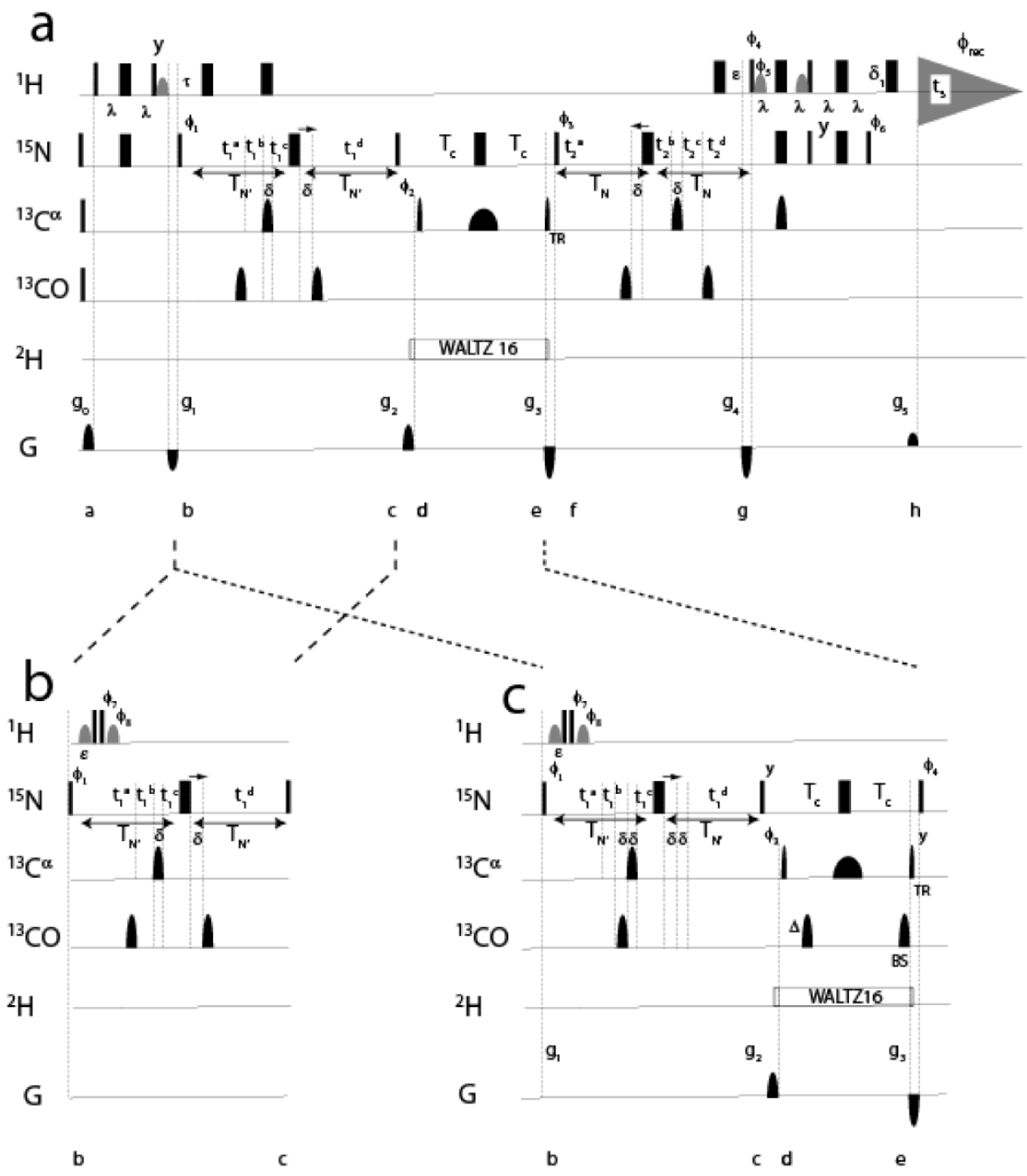

Figure 1.

(a, top) Pulse sequence of the NUS-hNcaNH-TROSY experiment. Narrow and wide solid rectangles indicate 90° and 180° pulses respectively. The pulses are applied along the x axis unless specified otherwise. Grey semi-ellipses indicate 800 μs 90° Sinc1 pulses (2000 Hz bandwidth, at 4.696 ppm) used to selectively affect the water. Narrow and wide hemi-ellipses indicate 90° (256 μs GaussCascadeQ5, 24.1 kHz) and 180° (205 μs GaussCascadeQ3, 16.5 kHz) shaped pulses respectively.[17] The 180° shaped pulse between points d and e is a 1039 μs Q3 refocusing pulse with a bandwidth of 4.1 kHz (22 ppm). The label TR at point e denotes a time-reversal of the de-excitation pulse. The delays are: $\varepsilon = 2.75$ ms $\approx 1/(4J(NH))$; $\lambda = 2.3$ ms $\approx 1/(4J(NH))$; $T_{N'} = 12.4$ ms; $T_N = 8.9$ ms = 13.5 ms - 2λ; $\tau = 9.65$ ms = $T_{N'} - \varepsilon/2 \approx T_{N'} -$

1/(4J(NH)). λ has been shortened compared to ε in order to account for proton relaxation. $T_{N'}$ differs from T_N since the latter benefits from TROSY implementation and incorporates the first half of the S^3 transfer. T_C is set to 16 ms as a compromise between relaxation and evolution under J(NC^{α}) scalar couplings (see text and supporting information). The short delay δ = 210μs compensates for a ^{13}C 180° pulse. The delays and increments for the ^{15}N t_1 semi-CT periods are: $t_1^a = t_1^d = T_{N'}$; $t_1^b = t_1^c = 0$ ms; $\Delta t_1^a = 1/(2SW_N) - T_{N'}/n_{1max}$; $\Delta t_1^b = 16.6ms/n_{1max}$; $\Delta t_1^c = 1/(2SW_N) - 16.6ms/n_{1max}$; $\Delta t_1^d = -T_N/n_{1max}$. The parameters for the ^{15}N t_2 semi-CT periods are: $t_2^a = t_2^d = 8.9$ ms $= T_N$; $t_2^b = t_2^c = 0$ms; $\Delta t_2^a = -8.9$ ms$/n_{2max}$; $\Delta t_2^b = 1/2SW_N - 15.7$ ms$/n_{2max}$; $\Delta t_2^c = 15.7$ ms$/n_{2max}$; $\Delta t_2^d = 1/2SW_N - 13.5$ ms$/n_{2max}$. The values n_{1max} and n_{2max} are the maximum incremented points for the respective dimensions. In our case, $n_{1max} = n_{2max} = 100$. The phase cycling is: ϕ_1 = x, -x; ϕ_2 = 2(x), 2(-x); ϕ_3 = 4(x), 4(-x); ϕ_4 = y; ϕ_5 = y; ϕ_6 = x; ϕ_{rec} = x, -x, -x, x, -x, x, x, -x. Pulse field gradients, 1 ms, with a sinusoidal shape are used with amplitudes of 25 G/cm for g_0, -18.5 G/cm for g_1, 22 G/cm for g_2, -38.5 G/cm for g_3, -40 G/cm for g_4 and 4 G/cm for g_5. All gradients were followed by a recovery delay of at least 200 μs. Quadrature detection in t_1 is achieved by the States-TPPI technique [18] applied to the phase ϕ_1. The gradient selection echo-antiecho sensitivity enhanced method [19] is used for evolution in t_2 where two fids are acquired with (g_4, g_5, ϕ_4, ϕ_5, ϕ_6) and with (-g_4, -g_5, ϕ_4 + 180°, ϕ_5 + 180°, ϕ_6 + 180°). Deuterium decoupling is achieved by using the WALTZ16 sequence.[20] **(b)** Modification to enable TROSY evolution in t_1 for the NUS-TROSY-hNcaNH-TROSY experiment. The period (b-c) replaces the period (b-c) in (**a**). The remainder of the sequences is described above. The phases are the same as above, except for ϕ_1 = x, y; ϕ_7 = x, -x and ϕ_8 = -x, x. **(c)** Modification to select correlations only with the following residue in the NUS-TROSY-s-hNcaNH-TROSY. The period (b-e) replaces the period (b-e) in (**a**). Δ = 4.5 ms ≈ 1/(4J($C^{\alpha}C'$)). The phases ϕ_1, ϕ_7 and ϕ_8 are as in (**b**).

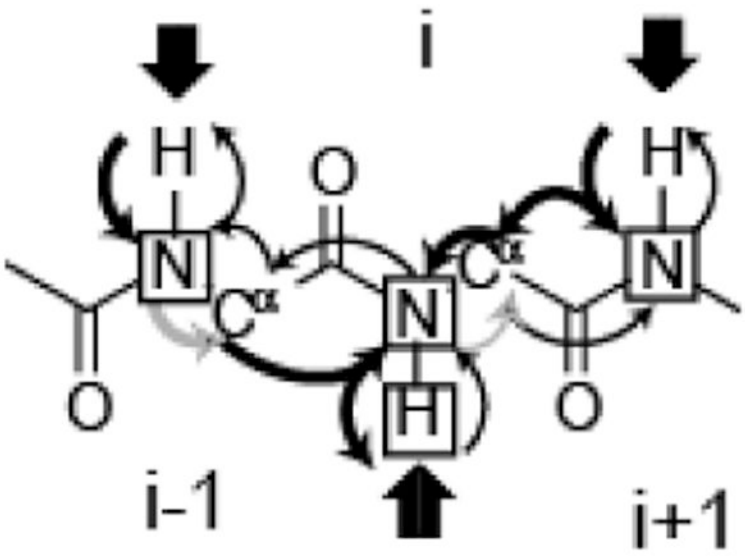

Figure 2.

Schematic representation of the transfers involved in the hNcaNH experiments. Only pathways leading to detection of the amide proton of the i-th residue are considered. A total of four transfers (represented by the curved arrows) are effected, starting from amide protons. The bold arrows highlight the inter-residue pathways that converge to the i-th residue. The squares indicate the coherences that are frequency labelled in the experiments. The grey arrows denote transfers that are eliminated in the NUS-TROSY-s-hNcaNH-TROSY experiment.

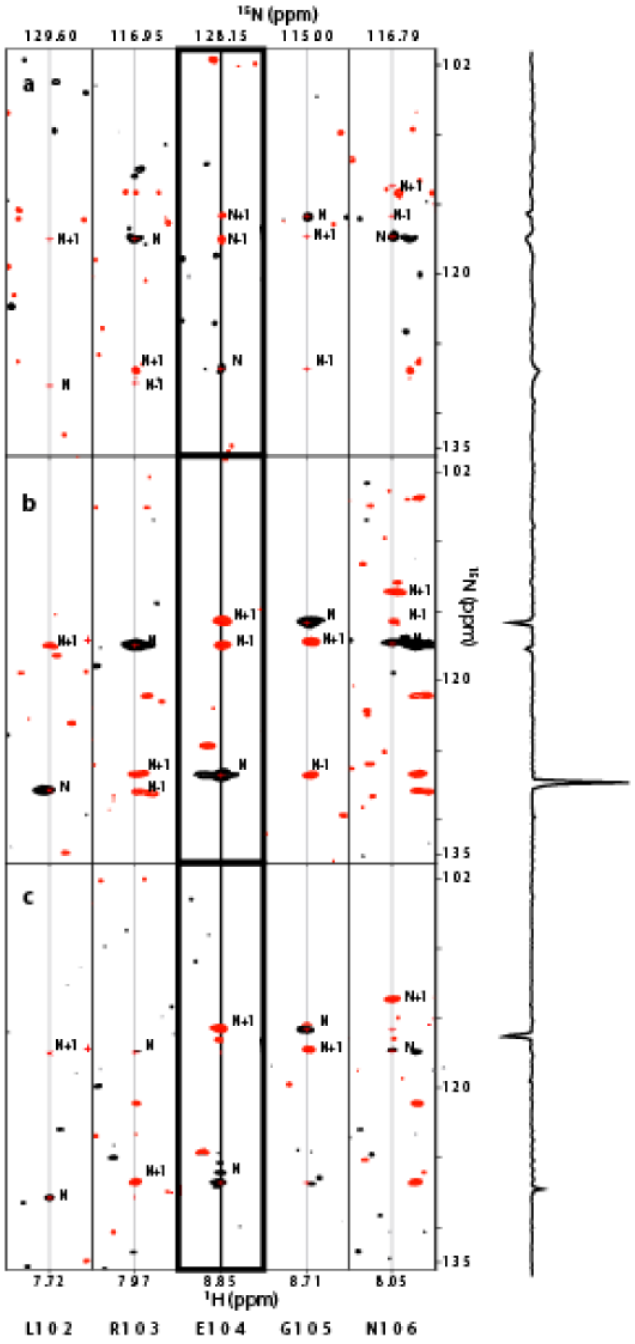

Figure 3.

Strips comparison obtained with the experiments described in Figure 1. The labels indicate where a cross-peak is expected. The same contour level was used for the three spectra. At the right hand side, cross-sections along ω_1 are shown for E104. The cross-sections have been rescaled in order to observe all signals for each experiment. **(a)** NUS-hNcaNH-TROSY (single TROSY experiment). Only two connectivities can be found, between R103 and E104 and between E104 and G105. **(b)** Double-TROSY NUS-TROSY-hNcaNH-TROSY. All connections can now be established in these strips. **(c)** Double-TROSY NUS-TROSY-s-hNcaNH-TROSY. Only cross-peaks to successors are present. The residue 101 is a proline and thus no N-1 signal is present in the strips of L102.