Journal of Biomolecular NMR, **20:** 135–147, 2001.
KLUWER/ESCOM

Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: Application to unfolded proteins

Sanjay C. Panchal, Neel S. Bhavesh & Ramakrishna V. Hosur*
Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India

Received 20 March 2001; Accepted 6 April 2001

Key words: HN(C)N, HNN, sequential ^{15}N correlations, triple resonance experiments, triplets, unfolded proteins

Abstract

Two triple resonance experiments, HNN and HN(C)N, are presented which correlate H^N and ^{15}N resonances sequentially along the polypeptide chain of a doubly (^{13}C, ^{15}N) labeled protein. These incorporate several improvements over the previously published sequences for a similar purpose and have several novel features. The spectral characteristics enable direct identification of certain triplets of residues, which provide many starting points for the sequential assignment procedure. The experiments are sensitive and their utility has been demonstrated with a 22 kDa protein under unfolding conditions where most of the standard triple resonance experiments such as HNCA, CBCANH etc. have limited success because of poor amide, C^α and C^β chemical shift dispersions.

Abbreviations: HIV-1 TD, human immunodeficiency virus-1 protease tethered dimer; INEPT, insensitive nuclei enhanced by polarization transfer; PEP, preservation of equivalent pathway; PFG, pulse field gradient; TROSY, transverse relaxation-optimized spectroscopy.

Introduction

Multidimensional triple resonance NMR experiments have provided a major break-through in the structure determination of large protein molecules in aqueous solutions. Assignments of ^1H, ^{13}C and ^{15}N nuclei along the backbone of a polypeptide chain can now be obtained fairly routinely using a set of standard experiments, such as HNCA (Kay et al., 1990), HN(CO)CA (Bax and Ikura, 1991), HNCO (Kay et al., 1990), CBCANH (Grzesiek and Bax, 1992a) and CBCA(CO)NH (Grzesiek and Bax, 1992b). Nonetheless, methodological improvements are continuing in order to enhance the speed of data analysis, reduce the ambiguity in the assignments, obtain connectivities across prolines (Lohr et al., 2000), and minimize the amount of experimental data required (Kong et al., 1999; Lin and Wagner, 1999; Liu et al., 2000). In all the above spectral analysis methods, the time consuming step is the repeated scanning through the ^{15}N planes of the 3D spectrum to locate the peaks at the desired chemical shifts. Equivalence of carbon chemical shifts complicates the analysis further. Thus, any help in such a search is highly desirable. Rapid assignment of backbone resonances is also of considerable importance at present in the context of high throughput structure determinations demanded by research in the area of 'structural genomics'. Besides, in flexible and unfolded proteins, ^1H and ^{13}C chemical shift dispersions are very poor and this poses a serious problem in obtaining sequence specific assignments (Logan et al., 1994; Schwalbe et al., 1997; Penkett et al., 1998; Henning et al., 1999) by the standard procedures mentioned above. At the same time, however, study of unfolded proteins is of considerable interest from the point of view of understanding protein folding mechanisms, protein misfolding and aggrega-

*To whom correspondence should be addressed. E-mail: hosur@tifr.res.in

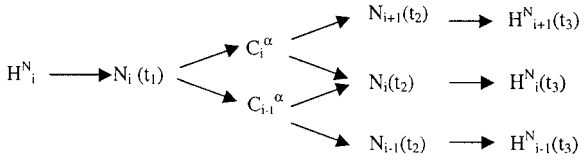
tion properties which lie at the heart of many diseases (Dill and Shortle, 1991; Thomas et al., 1995; Booth et al., 1997). Thus it has become highly necessary to develop new strategies and techniques to enable unequivocal assignment of resonances and consequently enable detailed characterization of unfolded proteins. In this context it is useful to note that ^{15}N chemical shift dispersions are very good even in unfolded proteins and one can base the strategies and experiments on these dispersions.

Some experiments have been described in the literature, which generate correlations between the sequential backbone amide and ^{15}N atoms (Grzesiek et al., 1993; Weisemann et al., 1993; Matsuo et al., 1996; Bracken et al., 1997; Ikegami et al., 1997; Liu et al., 2000; Lohr et al., 2000), and these employed two nitrogen evolutions to achieve maximum dispersion of peaks in the spectra. The experiments were based on HNCA, HN(CO)CA or HNCO types of transfers and employed a relay step to obtain i, $i-1$ and $i+1$ correlations along the polypeptide chains. However, the original implementation of sequences derived from HNCA and HN(CO)CA had very low sensitivities, since long delays were used and this caused a significant loss of the signal due to relaxation, especially of C^{α} carbons in the folded proteins to which they were applied. The situation would of course be much better in deuterium labeled proteins (Kay and Gardner, 1997), which behave significantly better from the point of view of relaxation losses. In fact, Ikegami et al. (1997) did observe good signals with their deuterium labeled proteins. Even otherwise, in unfolded proteins, the relaxation losses are not too serious because of the flexible nature of the polypeptide chain, and one may expect the experiments to be more successful.

In this paper we describe some improvements to the two pulse sequences based on HNCA and HN(CO)CA and with these we have been able to record excellent spectra with high resolution and sensitivity on a 22 kDa protein, HIV-1 protease tethered dimer under unfolding conditions. These have enabled rapid assignment of all the backbone resonances and some of these are indicated in the illustrative spectra shown for the two experiments. The modified experiments will be referred to as HNN and HN(C)N, respectively in the following description.

Methods

The pulse sequence for the 3D HNN is shown in Figure 1A. It begins with an INEPT transfer of magnetization from the H^N proton of residue i to the directly attached ^{15}N spin via the one-bond ^1H-^{15}N coupling (90–95 Hz). A constant time evolution period including t_1 follows. During this period the $^{15}N_i$ magnetization also evolves under one- (7–11 Hz) and two-bond (4–9 Hz) ^{15}N-$^{13}C^{\alpha}$ couplings. At the end of this period $^{15}N_i$ magnetization is antiphase with respect to $^{13}C_i^{\alpha}$ and $^{13}C_{i-1}^{\alpha}$. It is transferred to $^{13}C^{\alpha}$ through a pair of 90° pulses on ^{15}N and $^{13}C^{\alpha}$. Subsequently, during the time τ_{CN} the $^{13}C_i^{\alpha}$ and $^{13}C_{i-1}^{\alpha}$ magnetizations are transferred to their own ^{15}N spins via one-bond and to neighbouring ^{15}N spins via two-bond couplings, respectively. At the end of this period the $^{13}C^{\alpha}$ magnetization is antiphase with respect to ^{15}N, and is transferred to ^{15}N by a pair of 90° pulses. This is followed by a second ^{15}N constant time evolution period including t_2. During this period the ^{15}N magnetization becomes in-phase with respect to $^{13}C^{\alpha}$ and antiphase with respect to its amide proton. Finally, the ^{15}N magnetization is transferred back to the amide proton by a PFG coherence selection scheme (Kay et al., 1992; Muhandiram and Kay, 1994), and the proton magnetization is detected during the acquisition period t_3. The ^{13}CO spins are decoupled during both t_1 and t_2 periods. The magnetization transfer through the pulse sequence can be described in short as:



For the most general case of a sequence of amino acids ZXY, where Z, X, Y can be any residue other than glycine (which does not have a C^{β} carbon) and proline (which does not have an amide proton), the relevant density operator at point f in the pulse sequence is:

$$\sigma_f^A = \left\{ \begin{array}{l} -2H_{iz}N_{iy}\Gamma_{3A}\cos\left(\Omega_{N_i}t_2\right)\left[\Gamma_{1A}^2+\Gamma_{2A}^2\right]+ \\ \left[\begin{array}{l} 2H_{(i-1)z}N_{(i-1)y}\cos\left(\Omega_{N_{(i-1)}}t_2\right) \\ +2H_{(i+1)z}N_{(i+1)y}\cos\left(\Omega_{N_{(i+1)}}t_2\right) \end{array}\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A} \end{array} \right\} \Gamma_{5A}\cos(\Omega_{N_i}t_1) \quad (1)$$

where

$$\Gamma_{1A} = \left(2\pi^2 J_{C^{\alpha}N}T_N\right)\cos\left(2\pi^1 J_{C^{\alpha}N}T_N\right) \quad (2)$$

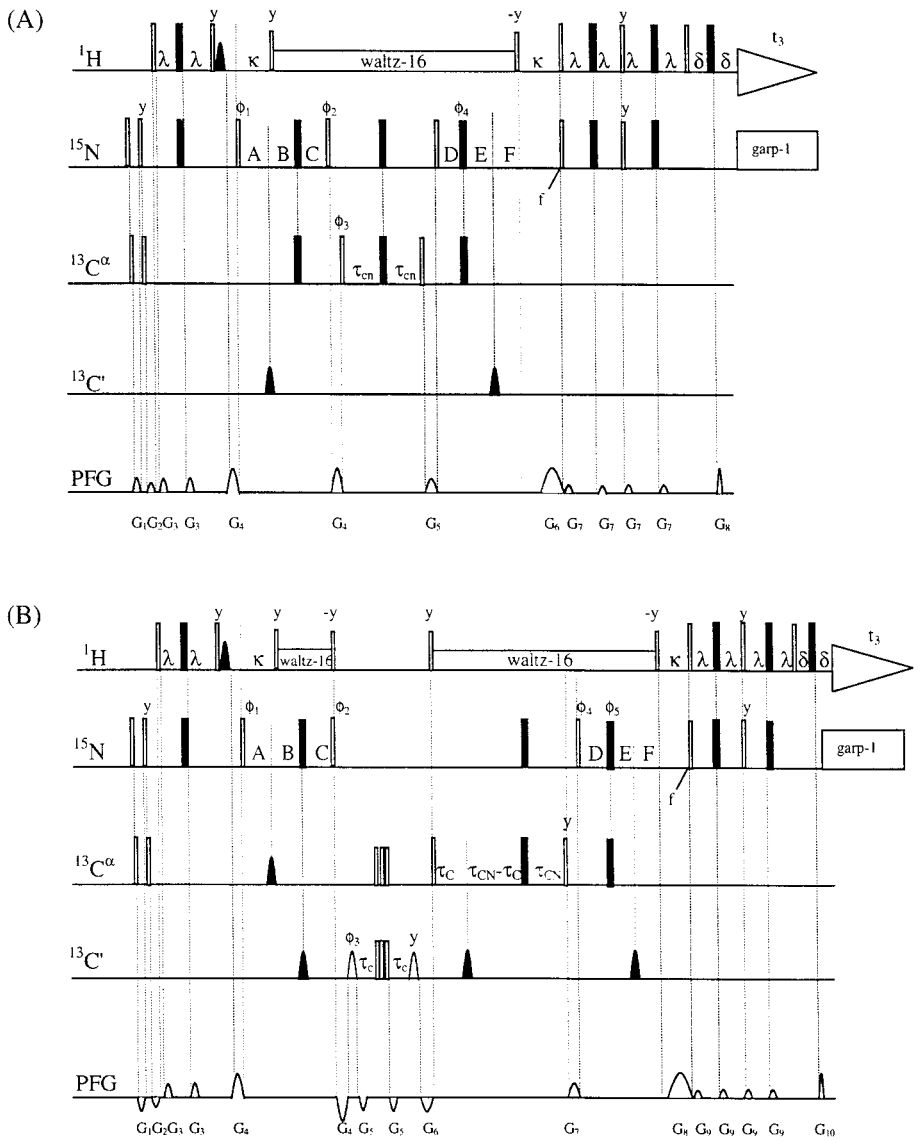


Figure 1. Pulse sequence for the (A) HNN and (B) HN(C)N experiment. Narrow (hollow) and wide (filled black) rectangular bars indicate non-selective 90° and 180° pulses, respectively. Unless indicated otherwise, the pulses are applied with phase x. The ^{1}H and ^{15}N carrier frequencies are set at 4.71 ppm (water) and 119.0 ppm, respectively. The ^{13}C carrier frequency is set at 56.0 ppm. The selective ^{1}H 90° pulse used for water flip-back during the first INEPT transfer is applied for a duration of 1.4 ms with one-lobe sinc profile. Proton decoupling using the Waltz-16 decoupling sequence with a field strength of 7.5 kHz is applied during most of the t_1 and t_2 evolution periods, and ^{15}N decoupling using the Garp-1 sequence with 0.5 kHz is applied during acquisition. The strength of the $^{13}C^\alpha$ pulses is adjusted so that they cause minimal excitation of carbonyl carbons. The 180° ^{13}CO shaped pulse had a one-lobe sinc profile with minimal excitation of $^{13}C^\alpha$. The delays are $\lambda = 2.7$ ms, $\kappa = 5.4$ ms and $\delta = 0.35$ ms. τ_{CN} must be optimized and is around 12–16 ms. The values of the individual periods containing t_1 are: $A = t_1/2$, $B = T_N$ and $C = T_N - t_1/2$. The values of the individual periods containing t_2 are: $D = T_N - t_2/2$, $E = T_N$ and $F = t_2/2$. Phase cycling for the experiment is $\phi_1 = 2(x), 2(-x)$; $\phi_2 = x, -x, x, -x$; $\phi_3 = \phi_2$; $\phi_4 = x$; $\phi_5 = 4(x), 4(-x)$ and receiver = 2(x), 2(−x). Frequency discrimination in t_1 is achieved using States-TPPI phase cycling of ϕ_1 along with the receiver phase; frequency discrimination in t_2 is achieved using the PEP sensitivity-enhanced gradient method. The N- and P-type signals are collected separately by inverting the sign of the G_6 gradient pulse (G_7 in case of HN(C)N). Signals recorded are manipulated post acquisition to generate pure absorption data. The gradient duration and levels are as follows: $G_1 = 0.5$ ms, 8 G/cm; $G_2 = 0.5$ ms, 5.6 G/cm; $G_3 = 0.5$ ms, 8 G/cm; $G_4 = 1$ ms, 11 G/cm; $G_5 = 1.0$ ms, 8 G/cm; $G_6 = 2.5$ ms, 26 G/cm; $G_7 = 0.5$ ms, 2 G/cm; $G_8 = 0.25$ ms, 25.8 G/cm. (B) Unless indicated otherwise, notations are the same as in (A). Composite pulses (indicated by three continuous 90° pulses) (Shaka, 1985) can be used to evolve $^{13}C^\alpha$ and ^{13}CO under one-bond $^{13}C^\alpha$-^{13}CO coupling. The delay τ_C can be set to 4.5 ms. The values of the individual periods containing t_1 and t_2 periods are the same as in (A). Phase cycling for the experiment is $\phi_1 = 2(-x), 2(x)$; $\phi_2 = x, -x, -x, x$; $\phi_3 = \phi_2$; $\phi_4 = x$; $\phi_5 = 4(x), 4(-x)$; and receiver = 2(x), 2(−x). The gradient duration and levels are as follows: $G_1 = 0.5$ ms, 8 G/cm; $G_2 = 0.5$ ms, 5.6 G/cm; $G_3 = 0.5$ ms, 8 G/cm; $G_4 = 1$ ms, 11 G/cm; $G_5 = 0.5$ ms, 7 G/cm; $G_6 = 1.0$ ms, 9 G/cm; $G_7 = 1.0$ ms, 8 G/cm; $G_8 = 2.5$ ms, 26 G/cm; $G_9 = 0.5$ ms, 2 G/cm; $G_{10} = 0.25$ ms, 25.8 G/cm.

$$\Gamma_{2A} = \cos\left(2\pi^2 J_{C^\alpha N} T_N\right) \sin\left(2\pi^1 J_{C^\alpha N} T_N\right) \quad (3)$$

$$\Gamma_{3A} = \cos\left(2\pi^1 J_{C^\alpha N} \tau_{CN}\right) \cos\left(2\pi^2 J_{C^\alpha N} \tau_{CN}\right) \quad (4)$$

$$\Gamma_{4A} = \sin\left(2\pi^1 J_{C^\alpha N} \tau_{CN}\right) \sin\left(2\pi^2 J_{C^\alpha N} \tau_{CN}\right) \quad (5)$$

$$\Gamma_{5A} = \cos\left(2\pi J_{C^\alpha C^\beta} \tau_{CN}\right) \quad (6)$$

and τ_{CN}, $2T_N = A + B + C = D + E + F$ are the delays as shown in Figure 1A. The ^1J and ^2J in Equations 2–6 are one-bond and two-bond coupling constants corresponding to their respective subscripts.

The first term in Equation 1 gives rise to the self or the diagonal (F_1=F_2) peak and the second term gives the sequential peaks to $i - 1$ and $i + 1$ residues, respectively. It follows that each residue i generates three peaks with the following coordinates: $F_1 = N_i$, (F_2, F_3) = (H^N_i, N_i), (H^N_{i-1}, N_{i-1}), (H^N_{i+1}, N_{i+1}). As a consequence, for $F_2 = N_i$, the (F_1, F_3) plane will have three peaks at the coordinates: (H^N_i, N_i), (H^N_i, N_{i-1}), (H^N_i, N_{i+1}).

The transfer efficiency, which dictates the intensity of the sequential peaks, is given by.

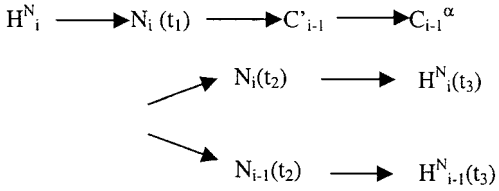
$$\Gamma_c^A = \Gamma_{1A}\Gamma_{2A}\Gamma_{4A}\Gamma_{5A} \quad (7)$$

Similarly the intensity of the diagonal peak is given by

$$\Gamma_d^A = -\Gamma_{3A}\Gamma_{5A}\left(\Gamma_{1A}^2 + \Gamma_{2A}^2\right) \quad (8)$$

For the standard values of the one- and two-bond C^α-N coupling constants (7–11 Hz), the one-bond C^α-C^β coupling constant (35–40 Hz), and the commonly used delays (12–16 ms), the functions Γ_{1A}–Γ_{4A} are positive and Γ_{5A} is negative. Thus from Equations 7 and 8 it follows that the diagonal and the sequential peaks will have positive and negative signs, respectively.

The pulse sequence for HN(C)N is shown in Figure 1B. The pathway followed by the magnetization through the pulse sequence can be traced as before and can be briefly described as:



Again, for the most general case of the sequence of amino acids, as considered earlier, the relevant density operator at point f in Figure 1B is:

$$\sigma_f^B = \begin{Bmatrix} 2H_{iz}N_{iy}\cos\left(\Omega_{N_i}t_2\right)\Gamma_{2B}\Gamma_{4B} \\ -2H_{(i-1)z}N_{(i-1)y}\cos\left(\Omega_{N_{(i-1)}}t_2\right)\Gamma_{3B}\Gamma_{5B} \end{Bmatrix} \sin(2\pi J_{C^\alpha C O}\tau_C)\Gamma_{6B}\Gamma_{1B}\cos(\Omega_{N_i}t_1) \quad (9)$$

where

$$\Gamma_{1B} = \sin\left(2\pi J_{C^\alpha C O}\tau_C\right)\sin\left(2\pi J_{NCO}T_N\right) \quad (10)$$

$$\Gamma_{2B} = \cos\left(2\pi^2 J_{C^\alpha N}\tau_{CN}\right)\cos\left(2\pi^1 J_{C^\alpha N}\tau_{CN}\right) \quad (11)$$

$$\Gamma_{3B} = \sin\left(2\pi^1 J_{C^\alpha N}\tau_{CN}\right)\sin\left(2\pi^2 J_{C^\alpha N}\tau_{CN}\right) \quad (12)$$

$$\Gamma_{4B} = \cos\left(2\pi^1 J_{C^\alpha N}T_N\right)\sin\left(2\pi^2 J_{C^\alpha N}T_N\right) \quad (13)$$

$$\Gamma_{5B} = \sin\left(2\pi^1 J_{C^\alpha N}T_N\right)\cos\left(2\pi^2 J_{C^\alpha N}T_N\right) \quad (14)$$

$$\Gamma_{6B} = \cos\left(2\pi J_{C^\alpha C^\beta}\tau_{CN}\right) \quad (15)$$

and τ_C, τ_{CN}, $2T_N = A + B + C = D + E + F$ are the delays as shown in Figure 1B. The ^1J and ^2J in Equations 9–15 are one-bond and two-bond coupling constants corresponding to their respective subscripts.

The first term in Equation 9 gives rise to the self (diagonal) peak and the second term gives the sequential peak to residue $i - 1$ only. It follows that each residue i generates two peaks with the following coordinates: $F_1 = N_i$, (F_2, F_3) = (H^N_i, N_i), (H^N_{i-1}, N_{i-1}). As a consequence, for $F_2 = N_i$, the (F_1, F_3) plane will have two peaks at the coordinates: (H^N_i, N_i), (H^N_i, N_{i+1}). Thus HN(C)N provides directionality to the sequential correlations.

The transfer efficiency which dictates the intensity of the sequential peak is given by

$$\Gamma_c^B = -\Gamma_{1B}\Gamma_{3B}\Gamma_{5B}\Gamma_{6B}\sin\left(2\pi J_{C^\alpha C O}\tau_c\right) \quad (16)$$

Similarly, the intensity of the diagonal peak is given by

$$\Gamma_d^B = \Gamma_{1B}\Gamma_{2B}\Gamma_{4B}\Gamma_{6B}\sin\left(2\pi J_{C^\alpha C O}\tau_c\right) \quad (17)$$

Here again it follows that for the standard values of coupling constants and delays, as mentioned earlier, the sequential and diagonal peaks will have negative and positive signs, respectively.

The HNN pulse sequence incorporates the following improvements over the HN(CA)NNH sequence originally proposed (Weisemann et al., 1993) to obtain sequential ^{15}N correlations: (i) HNN employs pulse field gradients as opposed to the phase cycling strategy used in HN(CA)NNH, for coherence selection and artifact suppression. This reduces the minimum phase cycle substantially, 2 versus 16 steps. (ii) Part of the phase cycle (ϕ_2, ϕ_3) used in HNN is to eliminate the ^{15}N z-magnetization that will be created after the ^{15}N to C$^{\alpha}$ transfer at the end of the first constant time period. This magnetization, if not eliminated, eventually appears on the diagonal peaks with different phases and causes serious distortion of the peaks and also produces undesirable tails. Besides, it will also produce peaks for the side chains of tryptophans, glutamines and asparagines, which are not required for backbone assignments and may lead to unnecessary confusion. We indeed observed such peak distortions and side chain peaks when the spectra were recorded without the phase cycle mentioned above. Surprisingly, however, this problem has not been addressed either in the original or in subsequent modified implementations, although PFGs were used in some cases for artifact suppression. (iii) The original pulse sequence (Weisemann et al., 1993) does not use a constant time (CT) approach for the nitrogen evolutions. Couplings to carbonyl and alpha carbons are active during the t_1 evolution period. In the subsequent 4D implementation (Ikegami et al., 1997), constant time has been used, but the carbonyl coupling is still active during the evolution period. Active carbonyl coupling during t_1 evolution leads to loss of resolution, if not splitting, and the C$^{\alpha}$-N coupling being active for the period ($t_1 + 2\Delta_1$) in the non-CT method leads to mixed phases. We may mention that while working with unfolded proteins or even with folded proteins with sharp lines, the highest possible resolution is desirable along the nitrogen axis. Since the spectral width along the nitrogen axis is usually quite small, it is practical to obtain good resolution even with a small number of t_1 and t_2 increments. (iv) The gradient sensitivity enhancement scheme (Kay et al., 1992) has been incorporated in the HNN scheme. Further modifications can be envisaged by incorporating TROSY (Pervushin et al., 1997) optimization in the pulse sequence.

Weisemann et al. ignored evolution under C$^{\alpha}$-C$^{\beta}$ coupling during the $2\Delta_2$ period of C$^{\alpha}$ evolution and the resultant transfer functions motivated the authors to use long delays so as to reduce the diagonal peaks and enhance the cross peaks. These protocols were also used in other subsequent publications (Ikegami et al., 1997). The long delays used resulted in a substantial loss of sensitivity. Our calculations (see below), which explicitly included evolution under C$^{\alpha}$-C$^{\beta}$ coupling, however, indicated that there is no need to use long delays to suppress diagonal peaks and the cross peaks can also have reasonable intensities with shorter delays. This optimization is really a compromise between relaxation losses and transfer efficiencies. We may also mention at this stage that we do not wish to suppress the diagonal peaks, since as we have shown below, the possibility of different signs of diagonal and cross peaks under short delay conditions has useful applications.

Likewise, the HN(C)N pulse sequence also has the same advantages of phase cycling, field gradients, sensitivity enhancement, etc., over the original implementation (Grzesiek et al., 1993; Matsuo et al., 1996; Bracken et al., 1997). In addition, a major difference lies in the magnetization transfer pathway: the ^{13}CO magnetization which is antiphase with respect to ^{15}N following the transfer from the latter is not immediately refocused. This is achieved indirectly later during the τ_{CN} period via the C$^{\alpha}$-CO coupling evolution. This requires a shorter time, since the C$^{\alpha}$-CO coupling constant is much larger and is readily absorbed into the τ_{CN} period. Thus there is a substantial reduction in the length of the pulse sequence and a consequent enhancement in the sensitivity. Further, during the CO evolution period, $2\tau_C$, the CO frequencies are refocused and the C$^{\alpha}$ are inverted to keep the C$^{\alpha}$-CO coupling active, by the application of a composite pulse (Shaka, 1985).

Experimental

NMR samples

The pulse sequences have been tested first on an eight-residue peptide (YGGFLRRI) which is ^{13}C/^{15}N labeled at G3, F4 and L5 and then on a ^{13}C/^{15}N labeled HIV-1 protease tethered dimer (HIV-TD) (MW = 22 kDa) protein. The peptide was purchased from Cambridge Isotope Laboratories, U.S.A. The concentration of the peptide was 1 mM in 100% DMSO. ^{13}C and ^{15}N labeled HIV-1 protease tethered

dimer was produced by over-expression, the details of which have been described elsewhere (Panchal et al., 2000). For NMR sample preparation, the protein solution was concentrated to $\sim$ 1 mM and exchanged with pH 5.2 NMR buffer consisting of 50 mM Na-acetate, 5 mM EDTA, 150 mM DTT, 6 M guanidine hydrochloride by ultrafiltration.

Data acquisition and processing

Spectra were recorded at ^{1}H Larmor frequencies of 600.052 MHz using a three-channel Varian Unity$^+$ spectrometer equipped with pulsed field gradients. Temperature settings were 25 °C and 32 °C for experiments on peptide and protein samples, respectively.

In the HNN and HN(C)N experiments on peptide, spectral widths comprised 26.3 ppm (^{15}N, F_1 and ^{15}N, F_2) and 14.2 ppm (^{1}H, F_3). Time domain data consisted of 24 $\times$ 24 $\times$ 512 complex points along t_1, t_2 and t_3, respectively in both the experiments. In each experiment, two scans were accumulated per transient, resulting in a measuring time of 2 h for each 3D spectrum (for $\tau_{CN} = 10$ ms). For quantitative calculation of transfer functions, a series of 3D spectra (for both experiments) were collected by varying the τ_{CN} period from 5 ms to 50 ms in steps of 5 ms.

For the protein, the spectral widths comprised 37.8 ppm (^{15}N, F_1 and ^{15}N, F_2) and 14.2 ppm (^{1}H, F_3) in both the HNN and HN(C)N experiments. Time domain data consisted of 68 (t_1) $\times$ 68 (t_2) $\times$ 512 (t_3) and 72 (t_1) $\times$ 72 (t_2) $\times$ 512 (t_3) complex points for the HNN and HN(C)N experiments, respectively. Four scans were used for each FID in both the experiments and the total acquisition time was 29.5 h for the HNN and 33.5 h for the HN(C)N experiment.

Processing of spectra was carried out using Felix 97.0 software. Prior to Fourier transformation and zero-filling, data were apodized with a sine-squared weighting function shifted by 60° in all dimensions. The matrices consisted of 64 $\times$ 64 $\times$ 1024 and 256 $\times$ 256 $\times$ 1024 real data points, respectively for peptide and protein.

Results and discussion

Transfer efficiency calculations

The intensities of the different diagonal and sequential peaks in the two types of spectra are dictated, as mentioned earlier, by the respective transfer functions (Equations 7, 8, 15 and 16). These in turn depend on the delays used. For the standard values of the one- and two-bond coupling constants (7 and 10 Hz respectively) and the T_N delays, 14 ms for HNN and 16 ms for HN(C)N, the transfer functions for the diagonal and sequential peaks as a function of τ_{CN}, are plotted in Figure 2. The transfer functions show a highly oscillatory behaviour in comparison with those reported earlier (Weisemann et al., 1993). This is a consequence of the contribution from C^α-C^β coupling evolution, which was not explicitly included by Weisemann et al. (1993). We see several nodes at which points the intensities of both the diagonal and the sequential peaks will be zero. It is interesting to note that in between two nodes, the patterns of oscillations of the peak intensities are different for different pairs of nodes and the signs of the peaks can readily change. These will be helpful in optimizing the experimental parameters for obtaining good intensities of the peaks and they will also be useful in proper interpretation of the spectra taking into consideration the relative signs of the sequential and diagonal peaks under chosen τ_{CN} delay conditions. We may hasten to add that the above curves will look differently whenever a glycine is present at the i or $i-1$ positions. An example of this is included in Figure 3.

In the above calculations, relaxation effects have, however, not been explicitly included. Although inclusion of the relaxation effects is not expected to change the relative profiles of the curves, transverse relaxation of C^α is very efficient and if the magnetization stays on C^α for a long time, as it does in these pulse sequences, there will be substantial loss of the signal. Nevertheless, in unfolded and flexible proteins, because of their longer relaxation times compared to folded proteins, the losses may be expected to be tolerable. In order to get an experimental assessment of these losses in unfolded proteins and arrive at suitable protocols for the delay times, we have carried out series of 3D experiments on a short eight-residue peptide (YGGFLRRI), labeled (^{13}C and ^{15}N) at G3, F4 and L5, by varying the τ_{CN} period in the range 5–50 ms. Figure 3 shows illustrative experimental spectra (top panels in B and C) and plots of the intensities of the diagonal and sequential peaks as a function of τ_{CN} (bottom panels in B and C), for both pulse sequences. Calculated transfer functions, ignoring relaxation, for the special situation here due to partial labeling of the peptide and due to the presence of glycine are also shown for ready comparison. It is seen that the sequential peaks are more affected by relaxation than diagonal peaks and this is due to the fact that the sequential peaks build up from zero, while the diagonal peaks die down

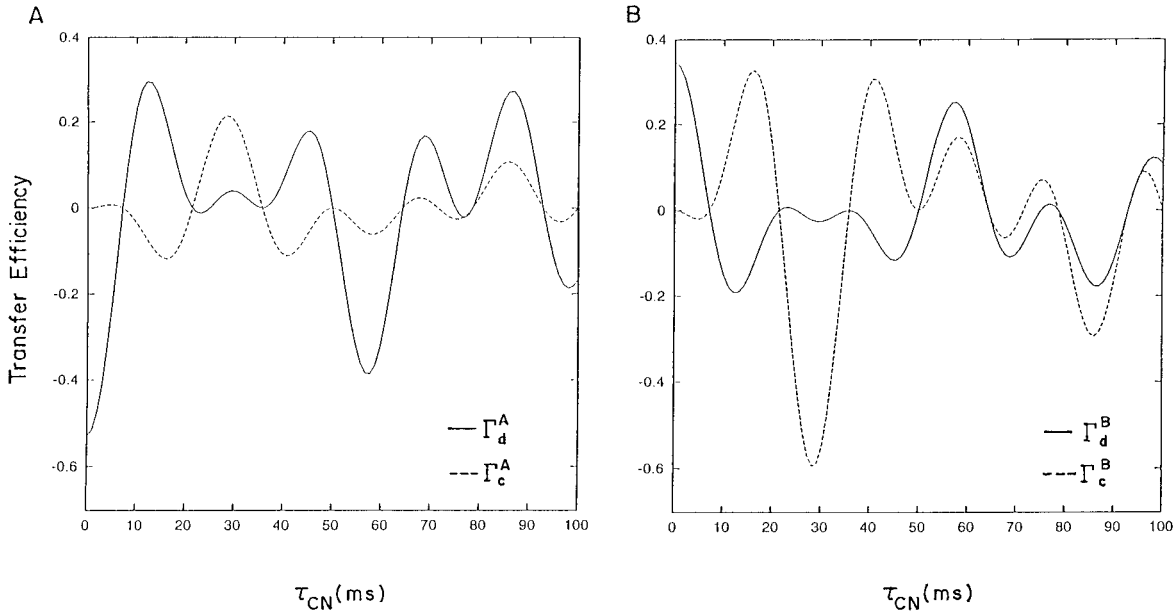


Figure 2. Plots of the (A) HNN and (B) HN(C)N coherence transfer efficiencies. The transfer functions Γ_c^A (Equation 7), Γ_d^A (Equation 8), Γ_c^B (Equation 15) and Γ_d^B (Equation 16) are described in the text. In each panel the diagonal and cross peak transfer efficiencies are shown by (—) and (- - - - -) lines, respectively. The transfer efficiency is plotted as a function of τ_{CN}. The plots were calculated by using $^1J_{C^\alpha N}$, $^2J_{C^\alpha N}$, $J_{C^\alpha C^\beta}$, $J_{C^\alpha CO}$, J_{NCO} values of 10, 7, 35, 55 and 15 Hz, respectively. The values of T_N used in transfer functions for HNN and HN(C)N were 14.0 ms and 16.0 ms, respectively and $\tau_C = 4.5$ ms.

as τ_{CN} increases, and the relaxation losses are obviously greater at higher τ_{CN} values. Nonetheless, for the peptide sample (concentration 1 mM), though each experiment used only 2 transients (acquisition time 2 h), the sequential peaks have good intensities even for fairly short τ_{CN} periods.

Amino acid sequence-dependent patterns in HNN and HN(C)N spectra

The facts that HNN and HN(C)N spectra have positive and negative peaks for the most commonly affordable delays, and that sequential correlations to both the neighbouring residues in a chain are seen, generate interesting features. These are described in the following paragraphs.

As mentioned before, the evolution of the product operators and hence the magnetization components are slightly different for glycine and proline neighbours. Four different cases of triplet amino acid sequences can be considered: (i) ZGY; (ii) GXY; (iii) GGY; and (iv) ZXY, where X, Y, Z can be any amino acid residue other than glycine and proline and G is glycine. The cases (i)–(iii) are special cases containing glycine in the triplet sequence and case (iv) is a general case. The final expressions for the observable magnetization for the three special cases of triplet sequences at time point f in the HNN pulse sequence are as follows:

(i)

$$\sigma_f^A = \left\{ \begin{array}{l} -2H_{iz}N_{iy}\Gamma_{3A}\cos\left(\Omega_{N_i}t_2\right)\left[\Gamma_{1A}^2\Gamma_{5A} + \Gamma_{2A}^2\right] + \\ \left[2H_{(i-1)z}N_{(i-1)y}\cos\left(\Omega_{N_{(i-1)}}t_2\right)\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A}\Gamma_{5A} + \\ \left[2H_{(i+1)z}N_{(i+1)y}\cos\left(\Omega_{N_{(i+1)}}t_2\right)\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A} \end{array} \right\} \cos(\Omega_{N_i}t_1) \quad (18)$$

(ii)

$$\sigma_f^A = \left\{ \begin{array}{l} -2H_{iz}N_{iy}\Gamma_{3A}\cos\left(\Omega_{N_i}t_2\right)\left[\Gamma_{1A}^2 + \Gamma_{2A}^2\Gamma_{5A}\right] + \\ \left[2H_{(i-1)z}N_{(i-1)y}\cos\left(\Omega_{N_{(i-1)}}t_2\right)\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A} + \\ \left[2H_{(i+1)z}N_{(i+1)y}\cos\left(\Omega_{N_{(i+1)}}t_2\right)\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A}\Gamma_{5A} \end{array} \right\} \cos(\Omega_{N_i}t_1) \quad (19)$$

(iii)

$$\sigma_f^A = \left\{ \begin{array}{l} -2H_{iz}N_{iy}\Gamma_{3A}\cos\left(\Omega_{N_i}t_2\right)\left[\Gamma_{1A}^2 + \Gamma_{2A}^2\right] + \\ \left[\begin{array}{l} 2H_{(i-1)z}N_{(i-1)y}\cos\left(\Omega_{N_{(i-1)}}t_2\right) \\ +2H_{(i+1)z}N_{(i+1)y}\cos\left(\Omega_{N_{(i+1)}}t_2\right) \end{array}\right]\Gamma_{1A}\Gamma_{2A}\Gamma_{4A} \end{array} \right\} \cos(\Omega_{N_i}t_1) \quad (20)$$

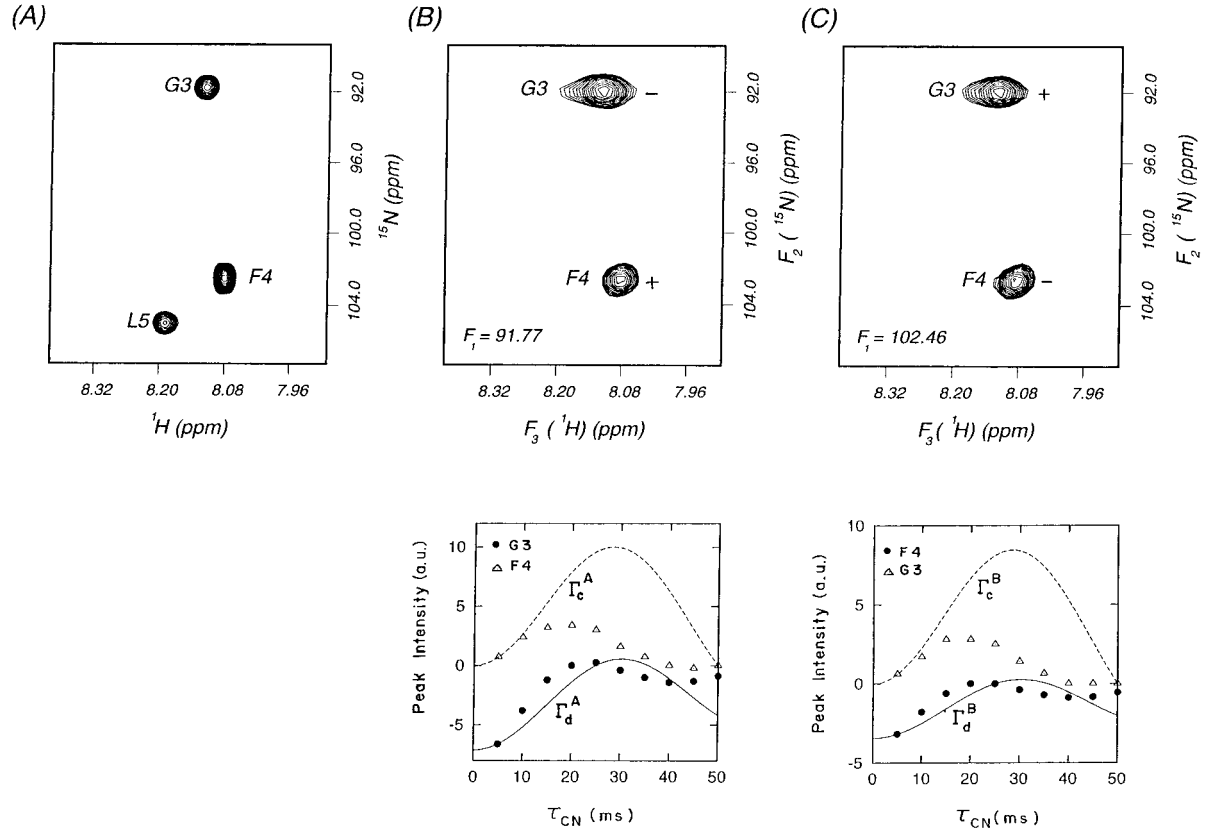


Figure 3. (A) HSQC spectrum of the peptide YGGFLRRI labeled (^{13}C, ^{15}N) at the G3, F4 and L5 positions. (B and C) Top panels: F_1-F_3 slices through the HNN (B) and HN(C)N (C) spectra at the ^{15}N (F_2) chemical shifts of G3 and F4, respectively. The signs of the peaks are indicated by + and − symbols. τ_{CN} = 10 ms in both the experiments. Bottom panels: Intensities of the diagonal (filled circle) and sequential (open triangles) peaks in the respective spectra as a function of τ_{CN}. In B, G3 is the diagonal peak and F4 the cross peak and in C, F4 is the diagonal peak and G3 the cross peak. The calculated transfer functions (excluding relaxation) for the diagonal and sequential peaks are shown by continuous and dashed curves, respectively in each panel. The calculated values have been normalized to match the experimental intensities at the shortest time value, 5 ms, assuming the relaxation losses to be minimal at this time point. The transfer functions are as follows: $\Gamma_d^A = \sin^2\left(2\pi^1 J_{C^\alpha N} T_N\right)\Gamma_{3A}$, $\Gamma_c^A = -[\Gamma_{1A} + \Gamma_{2A}]\Gamma_{4A}\sin\left(2\pi^1 J_{C^\alpha N} T_N\right)$, $\Gamma_d^B = \sin\left(2\pi J_{C^\alpha C^O}\tau_C\right)\Gamma_{1B}\Gamma_{2B}\Gamma_{4B}$, $\Gamma_c^B = \sin\left(2\pi J_{C^\alpha C^O}\tau_C\right)\Gamma_{1B}\Gamma_{3B}\Gamma_{5B}$.

The final expression for the observable magnetization for case (iv) is the same as that in Equation 8 and all Γ values given in Equations 2–6.

It is clear that glycines make an important difference in the expected patterns of peaks, as do also the prolines, which do not have an amide proton; the prolines show their presence by the absence of a peak.

Figure 4 shows the expected peak patterns in the F_1-F_3 planes of the HNN spectrum for several triplets of residues. The filled circle/square peak is taken to be positive, and the empty circle/square is taken to represent a negative peak. Figures 4A, B and C are F_1-F_3 planes at the ^{15}N chemical shift (F_2) of the central glycine in the triplet and Figures 4D, E, F and G are F_1-F_3 planes at the ^{15}N chemical shift (F_2) of the central X amino acid residue in the triplet. Note that in all these peak patterns the glycine ^{15}N chemical shift is taken to be distinctively upfield compared to that of the other residues. This is largely true and facilitates the assignment process further. The choice of the relative ^{15}N chemical shifts of i, $i-1$ and $i+1$ residues is quite arbitrary in each of the planes. In reality, the distribution of the positive and negative peaks can get altered as per the relative chemical shifts in each of the planes. The important points to consider are (i) the sign of the self or the diagonal peak; and (ii) the signs of the sequential peaks relative to the diagonal peak. The sign of the self peak for glycine is always opposite to that of any other residue. In order to em-

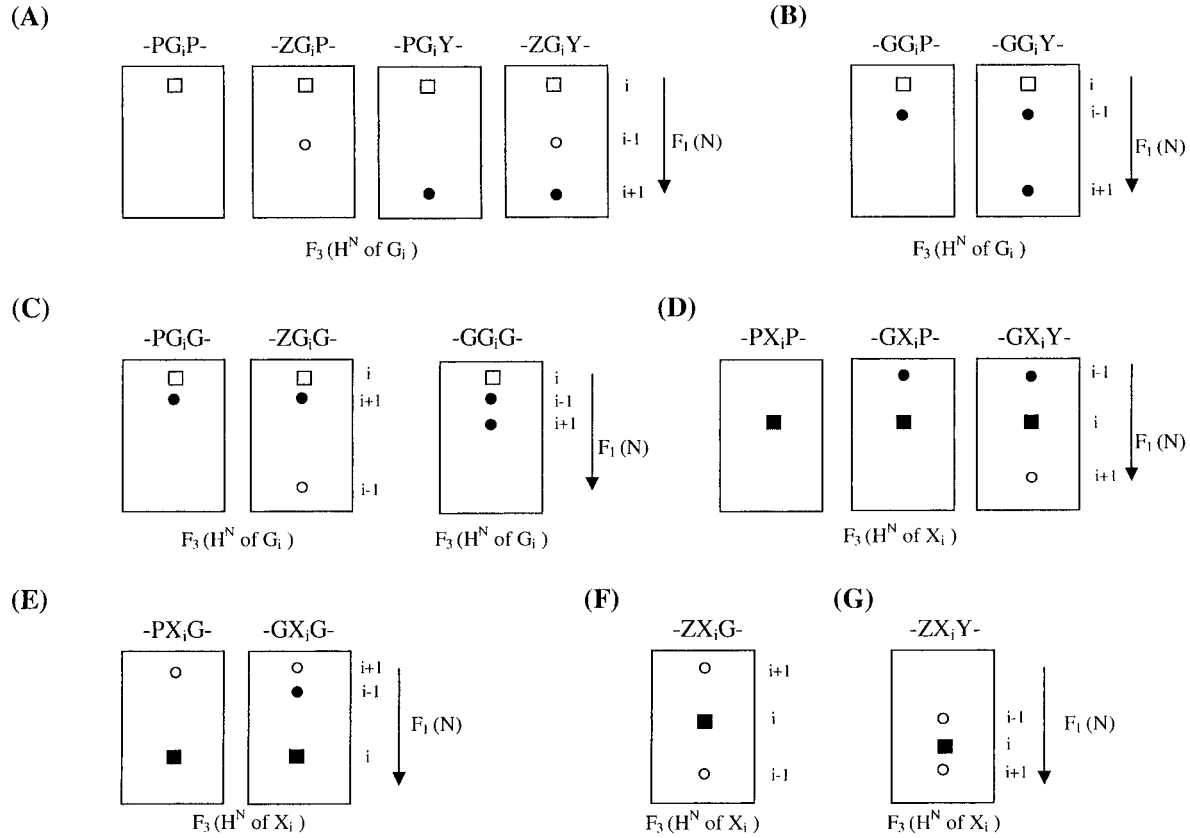


Figure 4. The expected peak patterns in the F_1-F_3 planes of the HNN spectrum for several triplets of residues. A, B, C are F_1-F_3 planes at the F_2 (^{15}N) chemical shift of the central glycine (*i*th residue). D, E, F and G are F_1-F_3 planes at the F_2 (^{15}N) chemical shift of the central X amino acid (*i*th residue). The diagonal ($F_1 = F_2$) peaks are indicated in squares and the sequential peaks in circles. The filled and empty squares/circles indicate positive and negative signs, respectively. The arrow indicates the direction of increase in the ^{15}N chemical shift values along that axis. It is assumed that the ^{15}N chemical shift of glycine is upfield compared to that of the other amino acid residues. For the other residues the ^{15}N chemical shifts are arbitrary.

phasize this feature, the diagonal peaks are shown with a different symbol (square) in the figure. This enables unambiguous discrimination between triplets having G and triplets having X as the central residues. For example, the patterns for PGY and PXG seem similar, but in the former the diagonal peak is negative and the sequential peak is positive, whereas the reverse is true in the latter case. The same observations hold for GGY and GXG patterns. The patterns for PGG and GGP are similar, but in a given protein, if only one of them is present then that can be readily identified. Out of the 16 triplet sequences shown in Figure 4, 14 (Figures 4A, B, C, D and E) can be easily identified. Out of the remaining two triplet sequences the triplet in Figure 4F can be easily identified if the glycine peak appears at its expected chemical shift, which is distinct from the rest of the amino acid residues. The pattern in Figure 4G is the most common one and has been included only as a reference.

A similar analysis has been carried out for the HN(C)N sequence, which also shows many distinctive features. Here the signs of the peaks for i and $i-1$ residues in the F_2-F_3 planes of the ith residue are dictated by the C^{α} evolution under C^{α}-C^{β} coupling of the $i-1$th residue only. Thus in the F_1-F_3 plane of residue i, the self (i) and sequential ($i+1$) peaks will have different sign combinations, depending on the neighbours of the ith residue. Some of these (i, $i+1$) patterns are, for example: GG_iY: ($-$, $+$), ZG_iY: ($+$, $+$), GX_iY: ($-$, $-$), ZX_iY: ($+$, $-$), etc.; note that here Y can be a G also. Again, whenever there is a proline at the $(i+1)$th position, the corresponding peak will be missing and only one peak will be seen from the ith amide proton.

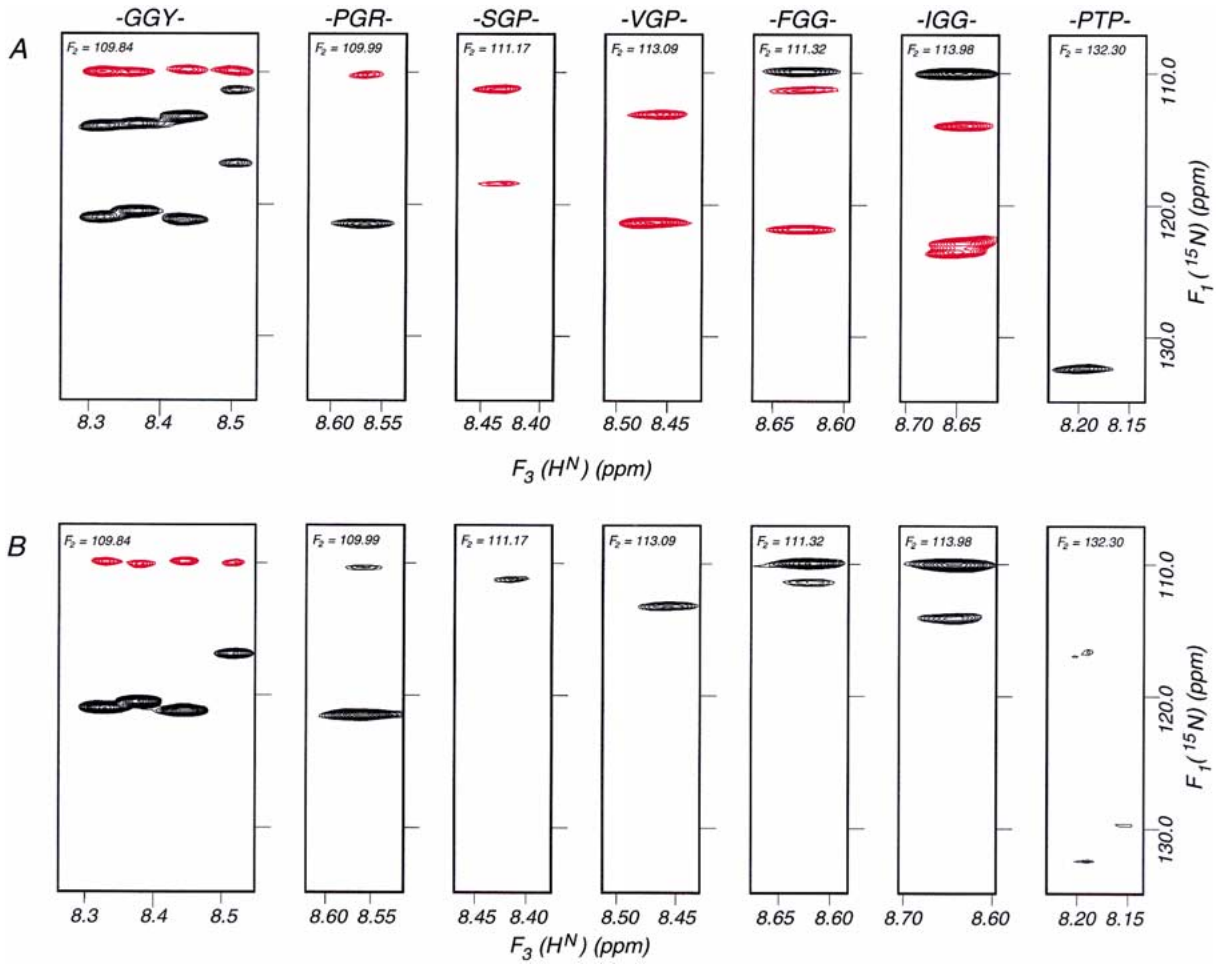


Figure 5. (A) Strips from (F_1, F_3) planes through the HNN spectrum at selected H^N positions which enable direct identification of the triplets of residues as schematically shown in Figure 4. Similar illustrative strips through the HN(C)N spectrum are shown in (B). Black and red contours indicate positive and negative peaks. Experimental parameters were as follows: HNN: $T_N = 15$ ms, $\tau_{CN} = 16$ ms, $t_{1max} = 29.5$ ms, and $t_{2max} = 29.5$ ms, number of complex increments = 68 along both t_1 and t_2; HN(C)N: $T_N = 16$ ms, $\tau_{CN} = 16$ ms, $t_{1max} = 31.3$ ms and $t_{2max} = 31.3$ ms, number of complex increments = 72 along both t_1 and t_2. Both the spectra were recorded with four transients for each FID. The total experimental time was 29.5 h for HNN and 33.5 h for HN(C)N. Temperature = 32 °C.

The above features of the two spectra have been demonstrated with a 22 kDa protein, HIV-1 protease tethered dimer, in 6 M guanidine hydrochloride. Table 1 lists the amino acid sequence and the identifiable special triplet sequences in the protein. All of these have been identified and Figure 5 shows the F_1-F_3 planes from HNN (A) and HN(C)N (B) for all except GIG, which happens to appear in a crowded area. Overall the HN(C)N spectrum has a slightly lower sensitivity compared to HNN. PGR and PTP sequences are unique in the protein and hence can be assigned straightaway from the HNN spectrum. There are four GGY triplet sequences (see Table 1) and all of them can be distinctly seen. Likewise, there are four ZGG triplets, one of which is present in a crowded area and hence is not shown. The two ZGP (SGP, VGP) sequences could be readily identified, again from the HNN spectrum. The specific assignments of FGG, SGP and VGP shown in the figure have been obtained after further analysis of the spectra. These strip plots provide the ^{15}N assignments for the two residues $i-1$ and /or $i+1$ neighbouring to i.

Sequential assignments

The possibility of direct identification of H^N and ^{15}N chemical shifts of neighbouring residues in the HNN and HN(C)N spectra suggests an easy strategy for rapid assignment of resonances along the polypeptide

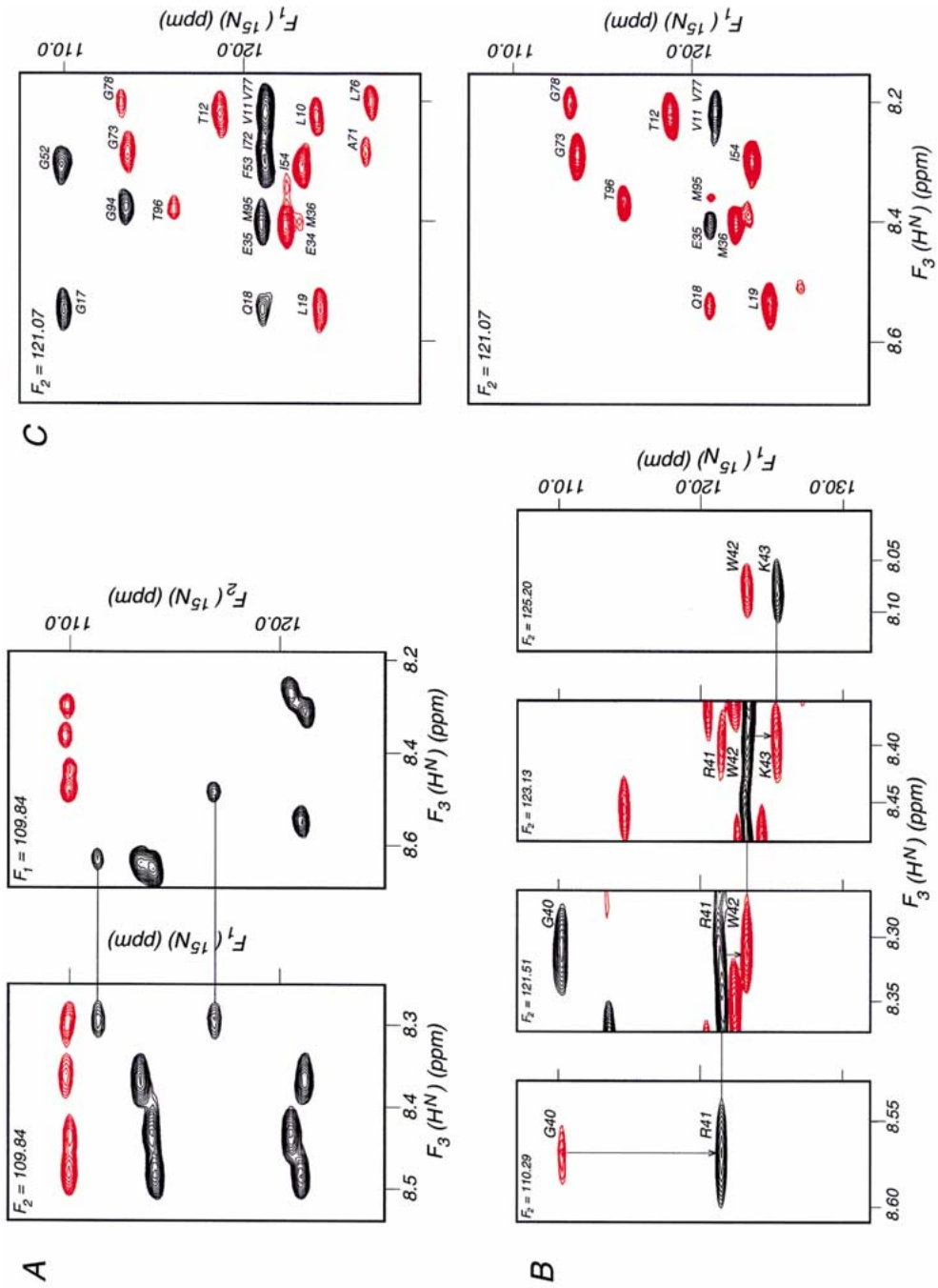


Figure 6. A. Selected $F_1(^{15}N)$-$F_3(^1H^N)$ and $F_2(^{15}N)$-$F_3(^1H^N)$ slices at a ^{15}N chemical shift of 109.84 ppm through the HNN spectrum. The horizontal lines connect the peaks to show the identification of the H^N chemical shifts in the (F_2, F_3) plane from the knowledge of the ^{15}N chemical shifts in the (F_1, F_3) plane. These enable rapid sequential assignments. (B) Sequential connections over a stretch of four residues in the HNN spectrum. Appropriate strips at the correct H^N chemical shifts in the (F_1, F_3) planes have been selected. (C) Selected (F_1, F_3) planes from HNN (top panel) and HN(C)N (bottom panel) at the same chemical shift to show the information in the two spectra. While HNN planes contain diagonal and two neighbouring peaks on either side, the HN(C)N displays the diagonal (i) and only the $i+1$ neighbouring peak. Diagonal peaks and cross peaks to glycines, which have positive sign, are in black and the other cross peaks, which have negative sign, are in red. The assignments of the various peaks are indicated. Note that in the HN(C)N spectrum, the diagonal peaks of F53 and I72 cancel each other due to their opposite signs. Similarly, the diagonal of M95 is very weak.

Table 1. Amino acid sequence of HIV-1 protease tethered dimer[a] and the special triplet sequences in the protein

Triplet type[b]	Triplet present in the HIV-1 protease tethered dimer
ZGP	VGP, SGP
PGY	PGR
GGY	GGQ, GGI, GGF, GGS
ZGG	$(IGG)_3$, FGG
PXP	PTP
GXG	GIG

[a]The amino acid sequence is: PQVTLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMSLPG RWKPKMIGGI GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGMTLNF ***GGSSG*** PQVTLWQRPL VTIKIGGQLK EALLDTGADD TVLEEMSLPG RWKPKMIGGI GGFIKVRQYD QILIEICGHK AIGTVLVGPT PVNIIGRNLL TQIGATLNF.

[b]X, Y, Z can be any residue other than G and P.

chain. While the F_1-F_3 planes help identification of ^{15}N chemical shifts, the F_2-F_3 planes enable identification of the corresponding H^N chemical shifts (Figure 6A). The special features discussed above provide many starting points for the backbone walk. In principle, each of the two experiments is capable of providing complete assignments independently. Nonetheless, a combined use of the two facilitates the assignments, since one of the two, namely HN(C)N, provides directionality to the sequential walk and has fewer peaks, which may help simplification in crowded areas. A short stretch of such sequential connectivities in HNN observed in the HIV-TD protein is shown in Figure 6B. In Figure 6C, a particular plane from both spectra in the crowded area is shown along with the assignments as illustrations. We have indeed obtained full assignments for the protein, which will be described separately along with the other assignments obtained following these H^N and ^{15}N assignments.

Conclusions

We have described in this paper two triple resonance pulse sequences, termed HNN and HN(C)N, to obtain sequential H^N and ^{15}N correlations along the backbone of a protein. In the former, a sequential walk can be performed in both directions along the chain, whereas in the latter, assignment proceeds in the C- to N-terminal direction along the polypeptide chain. We have demonstrated the utility of the two pulse sequences with an unfolded protein, which is clearly very demanding, because of poor chemical shift dispersions of the amide protons in unfolded proteins in general. On the other hand, however, the unfolded or flexible proteins have slower relaxation rates, which is of considerable benefit from the sensitivity point of view. In folded proteins, this is an important concern, but this can be readily overcome by using deuterated proteins (Kay and Gardner, 1997). In such a situation, deuterium decoupling will be required during C^α evolutions to gain sensitivity and the pulse sequences can be readily modified for these inclusions.

The combinations of positive and negative peaks for the various self and sequential peaks in the spectra of the two experiments are very helpful in triplet specific assignments. This facilitates rapid assignment of spectra of large proteins and automatization can be readily envisaged. This is of high significance for high throughput structure elucidations, as required in structural genomics research. Both pulse sequences can be readily extended to the fourth dimension by allowing C^α frequency labeling in a constant time manner during the τ_{CN} periods. This allows quick assignment of H^N, ^{15}N and C^α resonances from a single triple resonance experiment. We have, in fact, implemented these modifications, whose utility will be described elsewhere.

Acknowledgements

We thank the National Facility for High Field NMR at TIFR, supported by the Department of Science and Technology, India, for all facilities.

References

Bax, A. and Ikura, M. (1991) *J. Biomol. NMR*, **1**, 99–104.
Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res*., **26**, 131–138.
Booth, D.R., Sunde, M., Bellotti, V., Robinson, C.V., Hutchinson, W.L., Fraser, P.E., Hawkins, P.N., Dobson, C.M., Radford, S.E., Blake, C.C.F. and Pepys, M.B. (1997) *Nature*, **385**, 787–793.
Bracken, C., Palmer III, A.G. and Cavanagh, J. (1997) *J. Biomol. NMR*, **9**, 94–100.
Dill, K.A. and Shortle, D. (1991) *Annu. Rev. Biochem*., **60**, 795–825.
Grzesiek, S. and Bax, A. (1992a) *J. Magn. Reson*., **99**, 201–207.
Grzesiek, S. and Bax, A. (1992b) *J. Am. Chem. Soc*., **114**, 6291–6293.
Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc*., **115**, 4369–4370.
Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J. and Schwalbe, H. (1999) *J. Mol. Biol.*, **288**, 705–723.

Ikegami, T., Sato, S., Wälchli, M., Kyogoku, Y. and Shirakawa, M. (1997) *J. Magn. Reson.*, **124**, 214–217.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.

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

Kay, L.E. and Gardner, K.H. (1997) *Curr. Opin. Struct. Biol.*, **7**, 722–731.

Kong, X.M., Kong, H.S. and Zhu, G. (1999) *J. Biomol. NMR*, **14**, 133–140.

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

Lin, Y. and Wagner, G. (1999) *J. Biomol. NMR*, **15**, 227–239.

Liu, A., Riek, R., Wider, G., von Schroetter, C., Zahn, R. and Wüthrich, K. (2000) *J. Biomol. NMR*, **16**, 127–138.

Löhr, F., Pfeiffer, S., Lin, Y., Hartleib, J., Klimmek, O. and Rüterjans, H. (2000) *J. Biomol. NMR*, **18**, 337–346.

Logan, T.M., Theriault, Y. and Fesik, S.W. (1994) *J. Mol. Biol.*, **236**, 637–648.

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

Muhandiram, D.R. and Kay, L.E. (1994) *J. Magn. Reson.*, **B103**, 203–216.

Penkett, C.J., Redfield, C., Jones, J.A., Dodd, J., Hubbard, J., Smith, R.A.G., Smith, L.J. and Dobson, C.M. (1998) *Biochemistry*, **37**, 17054–17067.

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

Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J. and Dobson, C.M. (1997) *Biochemistry*, **36**, 8977–8991.

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

Thomas, P.J., Qu, B.H. and Pederson, P.L. (1995) *Trends Biochem. Sci.*, **20**, 456–459.

Weisemann, R., Rüterjans, H. and Bermel, W. (1993) *J. Biomol. NMR*, **3**, 113–120.