An efficient high-throughput protocol based on 2D-HN(C)N for unambiguous H^N and ^{15}N backbone assignment in small folded proteins in less than a day

Dinesh Kumar[1] and Ramakrishna V. Hosur[1,2,*]

[1]Department of Chemical Sciences,
Tata Institute of Fundamental Research,
1, Homi Bhabha Road, Colaba, Mumbai 400 005, India
[2]UM-DAE Centre for Excellence in Basic Sciences,
University of Mumbai, Kalina Campus, Santa Cruz,
Mumbai 400 098, India

An efficient high-throughput method for sequential assignment of backbone $^1H^N$ and ^{15}N, atoms in less than a day (using 2–4 2D spectra; total data collection in 4–8 h) has been proposed here. This is based on sequential correlations and specific patterns of peaks around the glycines, alanines, serines/threonines (internal check points) observable in the F_1–F_3 projection planes of the 3D-HN(C)N spectral variants. The F_2–F_3 projection planes of the spectra provide unique identification of the check points. The protocol has been demonstrated on two $^{13}C/^{15}N$-labelled proteins: ubiquitin and calbindin-D9k. In either case complete H^N and ^{15}N backbone assignments were obtained in less than a day each. The method would be valuable for NMR structural studies of small, well-folded proteins.

Keywords: Backbone assignment, check points, sequential amide correlation, structural proteomics.

BIOMOLECULAR nuclear magnetic resonance (NMR) spectroscopy has expanded dramatically in recent years and is now a powerful tool for the study of structure, dynamics and the interactions of biomolecules[1]. The only limitation for high-throughput NMR studies in the context of proteomics research is the long time needed to record a set of multidimensional NMR experiments for sequence-specific assignment and thus for structure determination. This also imposes a condition of long-term stability on the protein samples. Additionally, some proteins in solution tend to precipitate in a matter of days, thereby reducing the time available to record NMR data. This has shifted the focus of methodology development in NMR to: (i) reducing the number of NMR experiments to derive the required information, (ii) increasing the speed of data collection, and (iii) developing high-throughput procedures and algorithms/techniques for fast data analysis. There have been efforts to reduce the experimental time by adopting reduced dimensionality techniques[2–4]. Even so, these methodologies require either extensive collection of NMR data[3,4] or the data analysis is not straightforward[2]. In this background, we present here an efficient protocol based on 2D versions of HN(C)N[5,6] to obtain unambiguous sequence-specific H^N and ^{15}N backbone assignment in small, well-folded proteins.

The protocol basically extracts the necessary information (sequential correlations and check points for assignment of $^1H^N$ and ^{15}N atoms) from two experiments, termed herein as (i) 2D-hncNH (or the F_2–F_3 projection plane of the 3D-HN(C)N spectrum[5,6] recorded by avoiding the t_1 evolution) which provides unique identification of the check points, and (ii) 2D-hNcnH (or the F_1–F_3 projection plane of 3D-HN(C)N spectrum[5,6] recorded by avoiding the t_2 evolution) which provides two types of amide correlations on the ^{15}N–^1H HSQC-type spectrum: (i) intra-residue correlation [$^1H^N_i$–$^{15}N_i$] and (ii) inter-residue correlation [$^1H^N_i$–$^{15}N_{i+1}$]. Each of these can be recorded in a few hours time (1–2 h). These acquisition times are much less than what would be required for 3D experiments. Moreover, these can be recorded with larger number of increments and scans per FID without significantly increasing the acquisition time. Moreover, in such 2D-projection experiments, the slight signal attenuation

*For correspondence. (e-mail: hosur@tifr.res.in)

during the additional evolution time of the 3D experiment will be absent.

The basic idea of the protocol described here is as follows. The three-dimensional HN(C)N spectrum (Figure 1 *a*, left panel), shows H^N–^{15}N correlation between three consecutive residues $i-1$, i, and $i+1$ (refs 5, 6). The peaks appear at the following coordinates (Figure 1 *a*, middle panel):

$$F_1 = N_i;\ (F_3, F_2) = (^1H_i, ^{15}N_i), (^1H_{i-1}, ^{15}N_{i-1}),$$

$$F_2 = N_i;\ (F_3, F_1) = (^1H_i, ^{15}N_i), (^1H_i, ^{15}N_{i+1}).$$

The HN(C)N spectrum also exhibits characteristic sign patterns in the peaks as shown in Figure 1 *a* (right panel), considering the peptide sequence abBB′de. Depending upon the type of the HN(C)N experiment[5–8], B and B′ carry different meanings: it represents glycine in the normal HN(C)N experiment, glycine or alanine in the HN(C)N-A experiment, and serine or threonine in the HN(C)N-ST experiment; P is for proline and a, b, d, e represent any residue other than B and P. The triplet-specific peak patterns thus help in the identification of these special residues which serve as starting points and/or check points during the course of sequential walks, and explicit side-chain assignment becomes less crucial for unambiguous backbone assignment.

A simple examination of the spectral features in Figure 1 *a* (middle panel) reveals that a projection down the F_1 axis of the HN(C)N spectrum would produce a 2D-HSQC-type spectrum, containing peaks from glycines and their following residues only. This is because when an F_1 projection is taken, peaks belonging to these residues co-add whereas those belonging to the other residues cancel their intensities; this is illustrated schematically in Figure 1 *b*. It is evident that the peaks of glycines and their sequential following residues will have opposite signs. This 2D spectrum will, hereafter, be referred to as hncNH-G, which can be recorded from the 3D pulse sequence[5] by avoiding the t_1 evolution. Similarly, alanine and serine/threonine check points can be derived from other variants of the experiment, HN(C)N-A and HN(C)N-ST[7,8] respectively. These spectra hereafter will be referred to as hncNH-A and hncNH-ST respectively. Thus depending upon the variant of the HN(C)N experiment, the 2D-hncNH spectra will provide amino-acid type identification of glycines, alanines, serines/threonines (primary check points) and the residues following them in the sequence (secondary check points) directly on a ^1H–^{15}N HSQC type spectrum. These provide a number of initial start points and check points during the sequential walk along the polypeptide chain for obtaining sequence-specific assignment.

Figure 1 *a* (middle panel) also reveals that a projection down the F_2 axis of the HN(C)N spectrum would produce a 2D-HSQC-type spectrum which has both the self (i) and the sequential ($i+1$) peaks aligned at the $^1H^N$ chemical shift of i, for each non-proline residue. They carry with them their specific peak patterns. This is schematically illustrated in Figure 1 *c*. Such 2D spectra will hereafter be referred to as hNcnH-G, hNcnH-A and hNcnH-ST respectively, which can be recorded from the respective 3D pulse sequences[5,7,8], avoiding the t_2 evolution. These, together with the check points identified earlier provide a large number of internal checks during the sequential walk along the polypeptide chain, enabling unambiguous

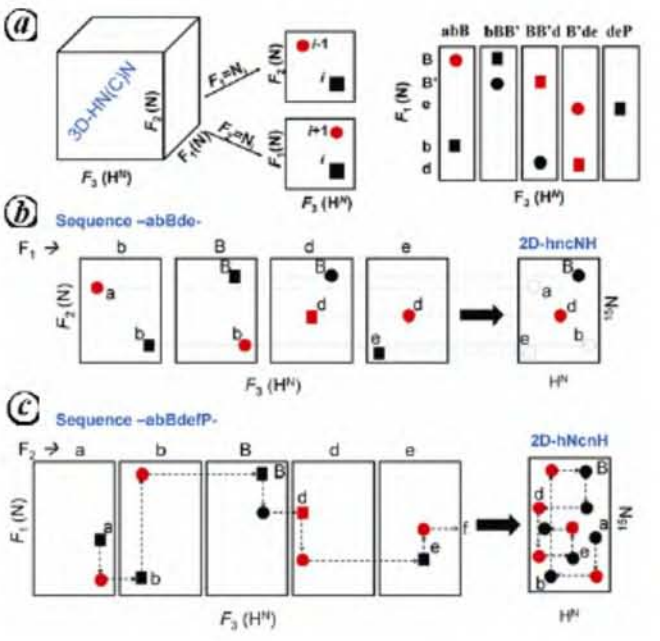

Figure 1. Schematic illustration showing spectral features of (*a*) 3D-HN(C)N, (*b*) 2D-hncNH and (*c*) 2D-hNcnH experiments. The latter two have been derived from the HN(C)N experiment (see text). Squares and circles represent the self and sequential peaks respectively. Black and red colours represent positive and negative signs respectively.

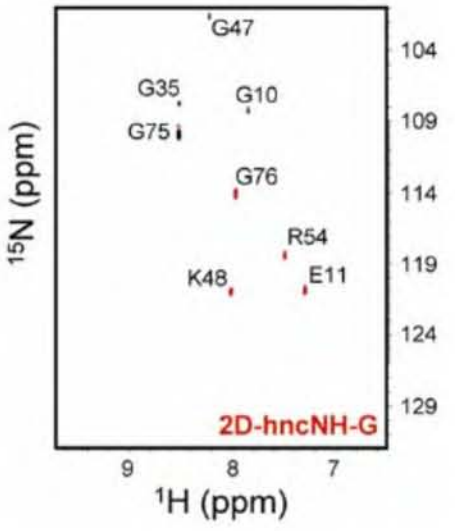

Figure 2. The 2D-hncNH-G spectrum of 1.6 mM ubiquitin (76 aa).

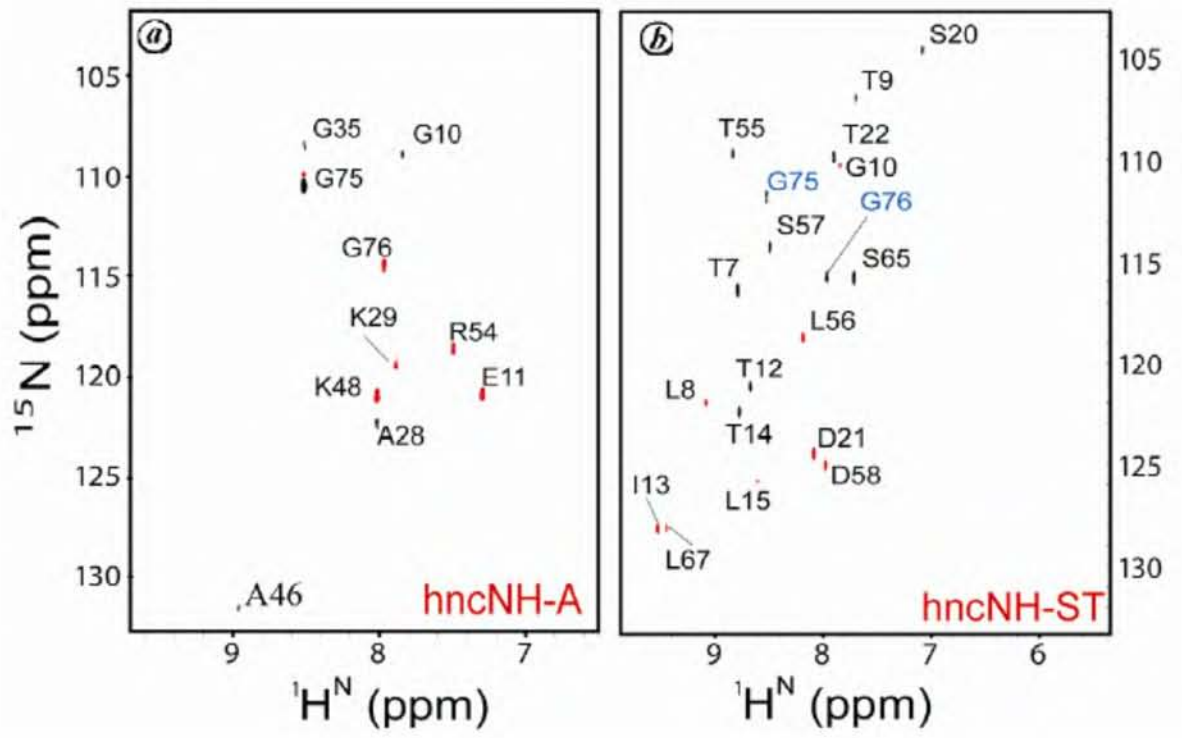

Figure 3. 2D-hncNH-A (*a*) and 2D-hncNH-ST (*b*) spectra of ubiquitin recorded on an 800 MHz Bruker spectrometer equipped with cryoprobe. Each 2D spectrum is recorded on 1.6 mM ubiquitin sample in about 2 h. In each case 64 complex increments were used along the indirect dimension. The residues labels on the peaks have come after the sequential assignment, but are indicated for clarity. In *b*, residual G75 and G76 peaks are seen because of improper cancellations.

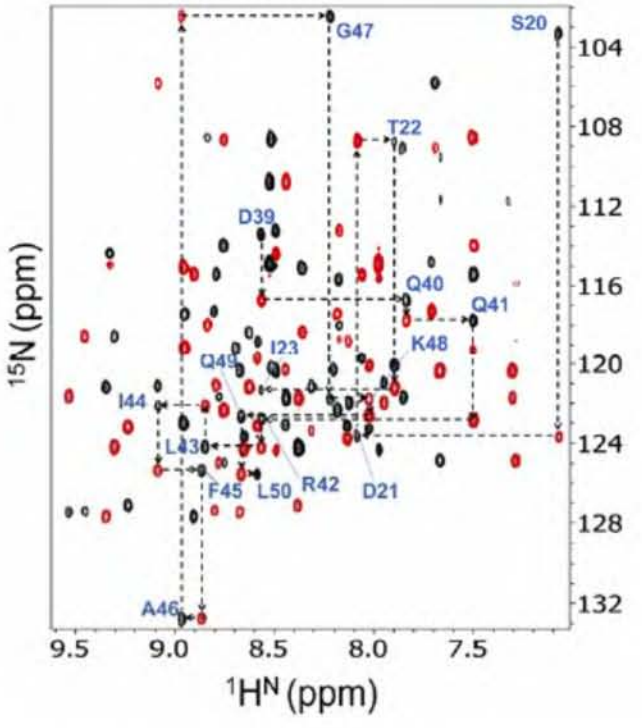

Figure 4. The 2D-hNcnH-G spectrum of 1.6 mM ubiquitin (76 aa) and illustrative stretches of sequential walk on this spectrum (S20–I23 and Asp39–Leu50) shown with dotted lines. Black and red colours represent self and sequential peaks respectively, except for residues which are present next to glycines (check points).

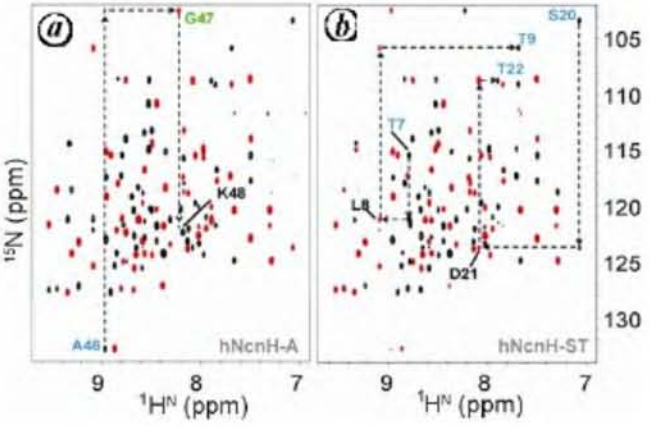

Figure 5. (*a*) The 2D-hNcnH-A and (*b*) 2D-hNcnH-ST spectra of ubiquitin recorded on a 800 MHz Bruker spectrometer with normal, room-temperature TXI probe. The spectra were recorded on 1.6 mM ubiquitin sample in about 2 h time each using 64 complex increments along the indirect dimension.

assignment. All the 2D spectra can be recorded in about 1–2 h, each, on the same (^{15}N/^{13}C)-labelled protein sample. The analysis being straightforward, the assignment can be completed in a few hours time.

A crucial requirement for the above protocol to work well is good dispersion of the amide proton resonances. In well-folded proteins of medium size, this can be expected to occur. Higher magnetic fields will also be

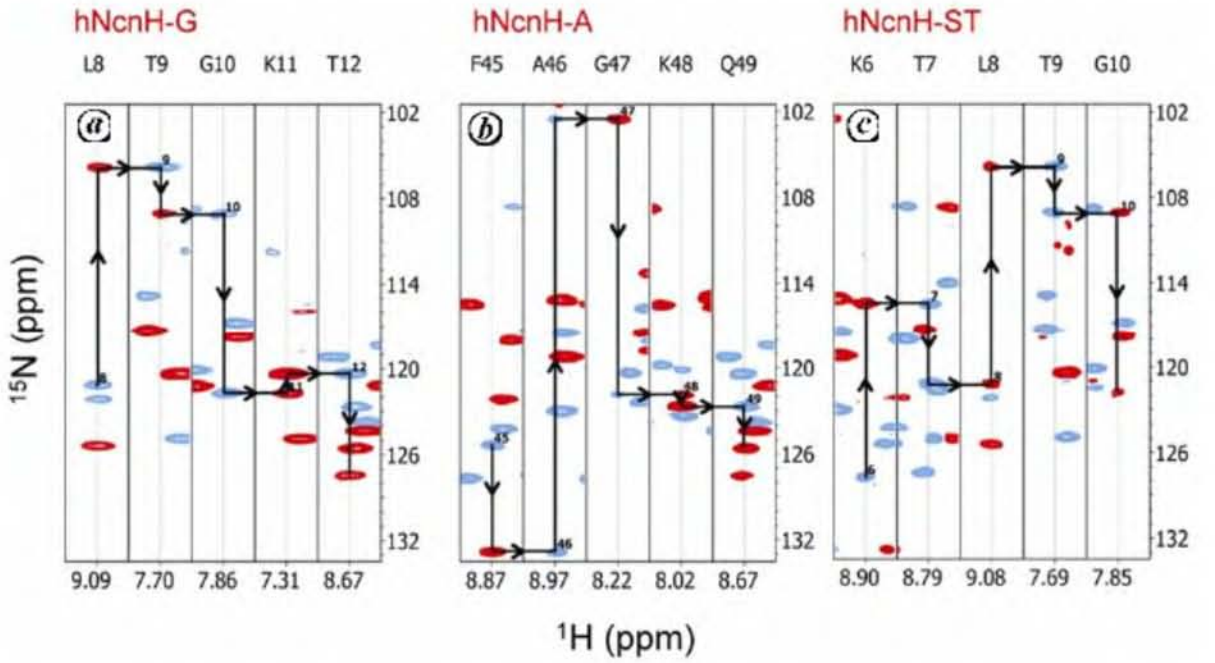

Figure 6. Illustrative stretches of sequential walk through the 2D-hNcnH-G, 2D-hNcnH-A and 2D-hNcnHST spectra. Residue labels on the peaks have come after the sequential assignment, but are indicated for clarity (see text).

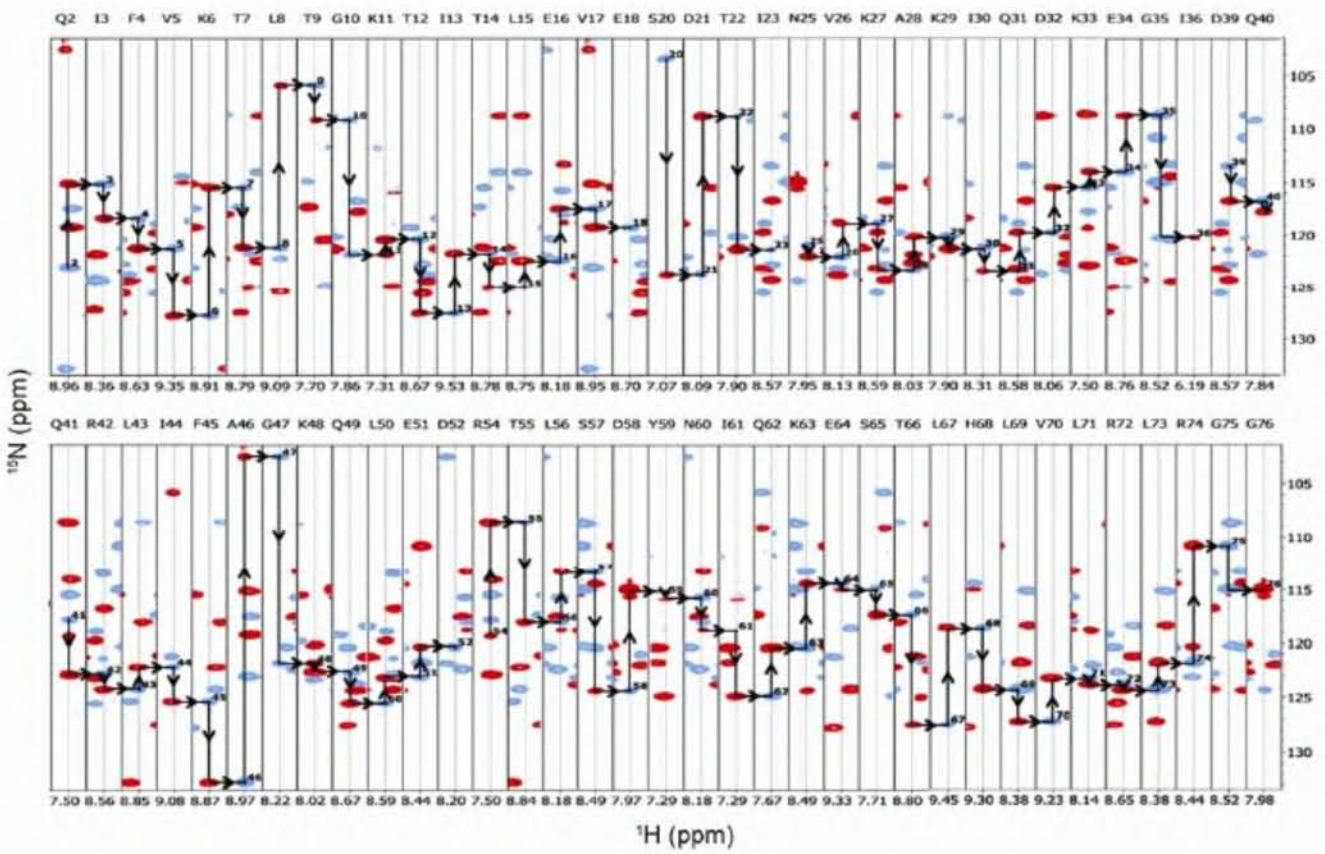

Figure 7. Complete sequential walk through the 2D-hNcnH spectrum of ubiquitin. Cyan and red colours represent positive and negative signs respectively. A sequential peak in one plane joins the self peak in the adjacent plane on the right. The numbers at the bottom in each strip identify the H^N chemical shifts. The self peak in the strip is identified from ^{15}N chemical shift.

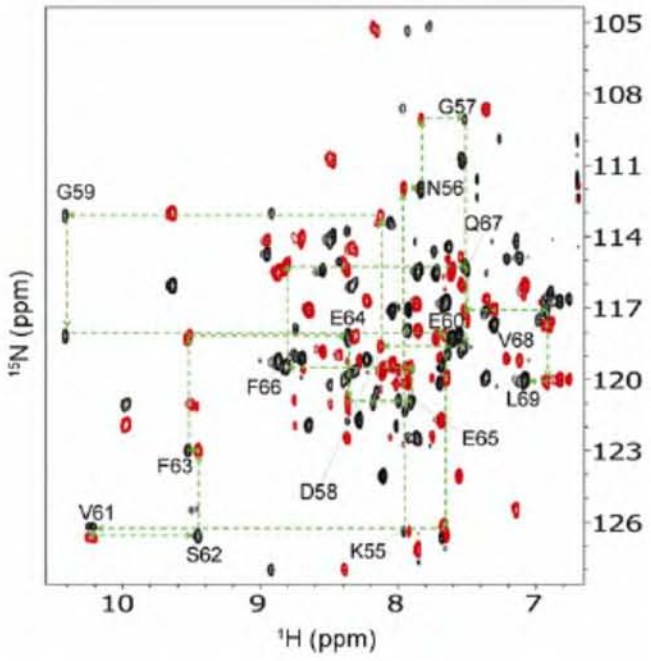

Figure 8. The 2D-hNcnH spectrum of 1.0 mM ^{15}N/^{13}C labelled calbindin-D9k and an illustrative stretch of sequential walk on this spectrum (Lys55–Leu69) shown with dotted lines. Black and red colours represent self and sequential peaks respectively, except for residues which are present next to glycines (check points).

extremely helpful in this regard. Assuming that the amide proton spectral dispersion is approximately 3 ppm, this would correspond to 1800 Hz on a 600 MHz spectrometer, but 2700 Hz on a 900 MHz spectrometer. Thus protons with small chemical-shift differences will get better resolved on a higher field spectrometer.

The method has been tested and demonstrated on two protein samples: (i) 1.6 mM ubiquitin (76 aa) and (ii) calbindin-D9k (75 aa) on an 800 MHz spectrometer equipped with cryoprobe. Figure 2 shows the 2D-hncNH-G spectrum of ubiquitin which allows direct identification of the glycines in the spectrum by their positive sign. In the same spectrum, the residues following these special residues are also seen but in negative phase and thus are readily identifiable by virtue of their opposite signs. An ambiguity arises when there are consecutive glycine residues in the sequence. In a stretch like –G′GX–, the G peak will disappear if X is a non-proline residue. For the terminal diglycine stretch (which is the case with ubiquitin-like proteins) or a stretch like –G′GP–, the G peak will be negative in sign. For proteins with such amino-acid sequences, unambiguous information of the check points can be derived from variants of 2D-(HN)NH experiments[9]. Figure 3 shows 2D-hncNH-A and 2D-hncNH-ST spectra of ubiquitin which display A and S/T related check points. Figure 4 shows a section of the 2D-hNcnH-G spectrum of ubiquitin. The figure also displays illustrative stretches of sequential walk through the 2D-hNcnH-G spectrum following the path from intra-residue correlation [$^{1}H^{N}_{i}$–$^{15}N_{i}$] via inter-residue correlation [$^{1}H^{N}_{i}$–$^{15}N_{i+1}$] to the next intra-residue correlation [$^{1}H^{N}_{i+1}$–$^{15}N_{i+1}$]. The intra- and inter-residue correlations here are identified respectively, by their positive and negative signs. Sequential correlations in 2D-hNcnH-A and 2D-hNcnH-ST spectra are illustrated in Figure 5. Illustrative stretches of sequential walks through 2D-hNcnH-G, 2D-hNcnH-A and 2D-hNcnH-ST spectra are shown in Figure 6 and Figure 7 depicts complete sequential walk through the 2D-hNcnH-G spectrum for ubiquitin. Similarly, as an illustration of calbindin application, the 2D-hNcnH-G spectrum of the protein is shown in Figure 8.

In conclusion, an efficient method for rapid sequential assignment of amide and ^{15}N resonances in small, well-folded proteins has been proposed. This paves the way for fast backbone ^{13}C assignments using 2D versions of hnCOcaNH and hncoCANH experiments[11] described recently, and thereby to structure determination in less than a day using algorithms which rely on chemical shift data alone[12,13]. Since, the experiments do not involve aliphatic protons, the method can also be envisaged to be applicable for deuterium-labelled, large molecular weight proteins. One can also envisage an application of these methods in *in cell* protein NMR, where rapid acquisition of data would be crucial due to problems of maintaining homogeneous cell suspensions for long periods of time. Moreover, for resonance assignment of small but unstable proteins and those which tend to precipitate in solution in a matter of days, the protocol will provide a better option to the biomolecular NMR community. Genome analysis has revealed that a large number of proteins expressed in prokaryotes and archaea are small in size[10]; nearly 22% of the structures in the Protein Data Bank are of proteins having molecular mass less than 15 kDa (http://pdbbeta.rcsb.org). Thus the protocol described here would be the method of choice in the context of structural and functional proteomics research.

1. Raman, S. and Lange, O. F., NMR structure determination for larger proteins using backbone-only data. *Science*, 2010, **327**, 1014–1018.
2. Atreya, H. S. and Szyperski, T., Rapid NMR data collection. *Methods Enzymol.*, 2005, **394**, 78–108.
3. Hiller, S. and Fiorito, F., Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 10876–10881.
4. Kupce, E. and Freeman, R., Projection–reconstruction of three-dimensional NMR spectra. *J. Am. Chem Soc.*, 2003, **125**, 13958–13959.
5. Panchal, S. C., Bhavesh, N. S. and Hosur, R. V., Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J. Biomol. NMR*, 2001, **20**, 135–147.
6. Chatterjee, A. and Bhavesh, N. S., A novel protocol based on HN(C)N for rapid resonance assignment in ((15)N, (13)C) labelled proteins: implications to structural genomics. *Biochem. Biophys. Res. Commun.*, 2002, **293**, 427–432.
7. Chatterjee, A., Kumar, A. and Hosur, R. V., Alanine check points in HNN and HN(C)N spectra. *J. Magn. Reson.*, 2006, **181**, 21–28.

8. Kumar, D., Chugh, J. and Hosur, R. V., Generation of Serine/Threonine Check Points in HN(C)N Spectra. *J. Chem. Sci.*, 2009, **121**, 955–964.
9. Kumar, D., Paul, S. and Hosur, R. V., BEST-HNN and 2D (HN)NH experiments for rapid backbone assignment in proteins. *J. Magn. Reson.*, 2010 (in press).
10. Brocchieri, L. and Karlin, S., Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.*, 2005, **33**, 3390–3400.
11. Kumar, D., Reddy, J. G. and Hosur, R. V., hnCOcaNH and hnco-CANH pulse sequences for rapid and unambiguous backbone assignment in (^{13}C, ^{15}N) labeled proteins. *J. Magn. Reson.*, 2010, **206**, 134–138.
12. Cavalli, A. and Salvatella, X., Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 9615–9620.
13. Shen, Y. and Lange, O., Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 4685–4690.

ACKNOWLEDGEMENT. We thank the Government of India for providing financial support to the National Facility for High Field NMR at the Tata Institute of Fundamental Research, Mumbai.

Received 20 May 2010; accepted 12 October 2010