ACADEMIC PRESS

Biochemical and Biophysical Research Communications 293 (2002) 427–432

BBRC

www.academicpress.com

A novel protocol based on HN(C)N for rapid resonance assignment in (^{15}N, ^{13}C) labeled proteins: implications to structural genomics☆

Amarnath Chatterjee,[a] Neel S. Bhavesh,[a] Sanjay C. Panchal,[b] and Ramakrishna V. Hosur[a,*]

[a] *Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India*
[b] *Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA*

Received 15 March 2002

Abstract

A novel protocol, based on the HN(C)N experiment, has been developed for rapid assignment of backbone H^N and ^{15}N resonances in (^{15}N, ^{13}C) labeled proteins. The protocol exploits the directly observable ^{15}N and H^N sequential correlations and the distinctive peak patterns in the different planes of the HN(C)N spectrum, depending upon the nature of the residues displaying the correlations. Glycines and prolines, which are responsible for the distinctive features, provide many check/start points for the sequential walks. These features enhance the speed of data analysis and render side chain assignments less crucial for the success of the assignments. The application of the protocol has been demonstrated with FK506 binding protein (FKBP, molecular mass 12 kDa).

Keywords: NMR assignments; HN(C)N protocol; Structural genomics

Structural investigations on proteins by NMR are, currently, almost invariably, carried out using multidimensional multinuclear experiments on doubly (^{15}N, ^{13}C), and, often, triply (^{15}N, ^{13}C, ^{2}H) labeled protein samples, produced by recombinant means (reviewed in [1,2]). Commendable successes have been achieved in the last decade and structures of several hundreds of proteins have been determined to date to very high resolutions. However, in the present post-genomic era, structural genomics which has occupied the center stage of NMR research has placed a very high demand on the speed of structure determination and consequently on the speed of NMR resonance assignment, minimization of experimentation, rapid preparation of isotopically labeled protein samples, etc. This in turn has necessitated the development of better and faster experimental techniques. In this context, we present here a novel protocol, based on the triple resonance three-dimensional experiment, HN(C)N [3], for rapid assignment of backbone ^{15}N and H^N atoms, which then form the basis for obtaining the other assignments in a straightforward manner from the standard procedures [2,3]. The performance of the new protocol has been tested with different proteins, but for demonstration we present here the data on FKBP (FK506 binding protein, molecular mass 12 kDa), a test protein which has been extensively studied in the past [4]. We believe that the new protocol represents a dramatic advancement over the methods based on HNCA [5], HN(CO)CA [6], CBCANH [7], CBCA(CO)NH [8], etc., currently in vogue, and would be the method of choice in the context of structural genomics.

Materials and methods

The doubly labeled (^{15}N, ^{13}C) FKBP protein, used in this paper, was a gift from Prof. M.K. Rosen, Sloan Kettering Cancer Research Center, New York.

NMR spectroscopy. The experiments were performed on a VARIAN Unity Plus 600 MHz spectrometer. FKBP protein concentration

☆ *Abbreviations:* NMR, nuclear magnetic resonance; HSQC, heteronuclear single quantum coherence; TROSY, transverse relaxation optimized spectropscopy.

* Corresponding author. Fax: +91-22-215-2110.
E-mail address: hosur@tifr.res.in (R.V. Hosur).

0006-291X/02/$ - see front matter
PII: S0006-291X(02)00240-1

= 2.2 mM, temperature = 25 °C, and pH = 5.2. For the HN(C)N spectrum, 40 complex points were collected along t_1 and t_2 dimensions and 16 scans were used for each FID. The delay parameters T_N, τ_{CN}, and τ_C were 16, 16, and 4.5 ms, respectively. The total acquisition time was 35 h. The ^1H–^{15}N HSQC [9] spectrum was recorded with 256 complex t_1 increments and four scans for each increment.

Results

The HN(C)N protocol

The present protocol is a consequence of the exploitation of the distinctive peak patterns in the different planes of the HN(C)N spectra. The origin of these patterns has been discussed earlier and we do not wish to repeat that discussion. The emphasis here is to present an efficient tool for routine NMR assignments and with that focus we include only the salient features of the HN(C)N experiment to help easy understanding of the protocol.

The HN(C)N experiment employs the magnetization transfer pathway: $H^N(i)$–$^{15}N(i)(t_1)$–$CO(i-1)$– $C^{\alpha}(i-1)$ –$^{15}N(i,i-1)(t_2)$–$H^N(i,i-1)(t_3)$, as shown schematically in Fig. 1A; t_1, t_2, t_3 refer to the time variables of the three-dimensional experiment. ^{15}N magnetization originating on the residue i (frequency labeled during t_1) appears on ^{15}N of both residues i and $i-1$ (frequency labeled during t_2) and is finally detected on amides of both i and $i-1$ residues during t_3. Similarly magnetization originating on $i+1$ residue is finally detected on both i and $i+1$ residues. Thus, in the three-dimensional spectrum, shown schematically in Fig. 1B, the peaks appear at the coordinates:

$$F_1 = N_i; \quad (F_3, F_2) = (H_i, N_i), \quad (H_{i-1}, N_{i-1}),$$
$$F_2 = N_i; \quad (F_3, F_1) = (H_i, N_i), \quad (H_i, N_{i+1}).$$

The spectrum shows H^N–^{15}N correlations between three consecutive residues $i-1$, i, and $i+1$. Under the choice of standard experimental data acquisition parameters, especially the constant time delay periods (T_N and τ_{CN}) for effecting magnetization transfers (10–16 ms), the diagonal ($F_1 = F_2$) and the sequential ($F_1 \neq F_2$) peaks in the (F_1, F_3) and (F_2, F_3) planes will have different combinations of positive (+) and negative (−) signs depending upon the nature of the residues at i and $i-1$ positions [3]. The four possible combinations of + and −, in the (F_1, F_3) planes and the specific residue-triplets, $(i-1,\ i,\ i+1)$, corresponding to those patterns are shown in Fig. 2A. Clearly, the glycine residue makes a significant contribution to the signs of the peaks. This is a consequence of the fact that the glycine residue lacks the C^{β} carbon. Further, whenever there is a proline at the $i+1$ position, the corresponding peak will be absent since the proline residue lacks the amide proton. This will generate two more patterns depending on whether the ith residue is a glycine or otherwise. These are also shown in Fig. 2A. From these, it is clear that whenever there is a glycine or a proline, there is a distinct change in the sign patterns and the protocol being described here cashes in on these features.

The sequential walk protocol is schematically shown in Fig. 2B. It relies exclusively on the (F_1, F_3) planes of the HN(C)N spectrum and the special peak patterns described above serve as start points and/or check points during the sequential walk. The essential ingredients of the sequential walk protocol are indicated as, 'start', 'continue', 'check', and 'break' in the figure. A vertical line makes a connection from the diagonal ($F_1 = F_2$) of the ith residue to the sequential peak identifying the ^{15}N chemical shift of the $i+1$ residue and then the adjoining horizontal line connecting to the diagonal of the $i+1$ residue identifies the amide chemical shift of that residue.

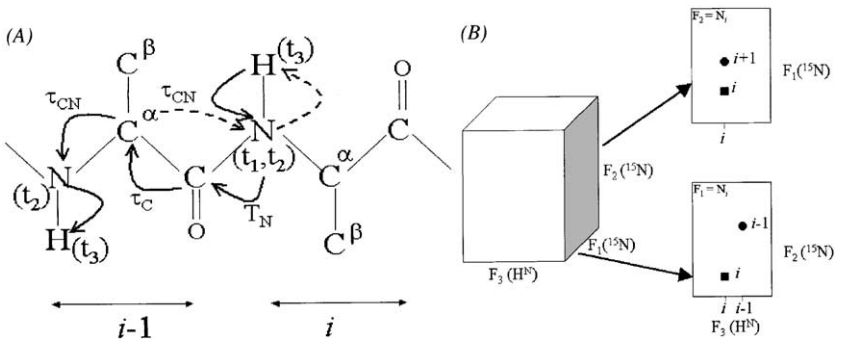

Fig. 1. (A) Magnetization transfer pathway employed in the HN(C)N experiment, shown on the peptide backbone. The magnetization originates on the amide proton of the ith residue and is finally detected on the amide protons of the i and $i-1$ residues. In this sense, the thick and dashed arrows indicate the transfers in the forward and backward directions, respectively, as shown. t_1, t_2, t_3 are the time variables of the three-dimensional experiment. T_N, τ_C, and τ_{CN} are the delay periods for effecting the magnetization transfers. (B) Schematic three-dimensional spectrum and the correlations observed in the F_1–F_3 and F_2–F_3 planes at the ^{15}N chemical shifts of residue i. Squares and circles represent the diagonal and sequential peaks, respectively. Note that in the F_1–F_3 plane, i to $i+1$ correlations are seen, and, in the F_2–F_3 plane, i to $i-1$ correlations are seen.

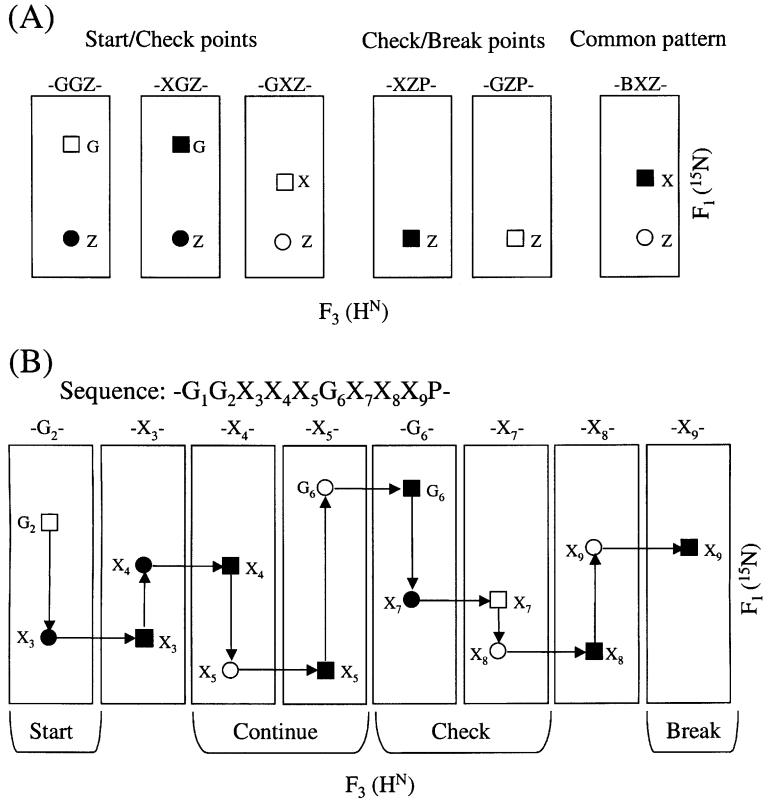

Fig. 2. (A) Schematic peak patterns in the F_1–F_3 planes of HN(C)N spectrum and the corresponding different triplets of residues. In every strip the peaks occur at the H^N chemical shift (F_3) of the central residue in the triplet indicated on the top. B and X can be any residues other than P and G; Z can be any residue other than P. Squares are diagonal ($F_1 = F_2$) peaks and circles are sequential peaks. Filled and open symbols represent the positive and negative signs, respectively. The patterns involving G serve as start and/or check points during a sequential assignment walk through the spectrum. GGP, which is likely to be less frequent than the others, may serve as a unique start point. The patterns involving P identify break points during a sequential walk and thus they also serve as check points. The pattern, which does not involve a G or a P, is the most common one occurring through the sequential walk. (B) The schematic sequential walk protocol through the F_1–F_3 planes of the HN(C)N spectrum of a protein. An arbitrary amino acid sequence is chosen to illustrate the start, continue, check, and break points during the sequential walk. The peak patterns are drawn as per the schematic in A and the residue identified on the top of each strip identifies the central residue of the triplet.

The sequential assignment proceeds from the N-terminal toward the C-terminal of the protein.

Application to FKBP

We have tested the above protocol with different proteins, but, for illustration we present here the data on the 12 kDa protein, FKBP, on which extensive work has already been reported [4] and hence serves as a good test case. The amino acid sequence of the protein, with the glycine and proline positions highlighted, is shown in Fig. 3A. Clearly, a number of 'check points' and 'start points' are possible and the experimental strips for many of these are shown in Fig. 3B. Using these, we have been able to make an almost complete sequential walk through the sequence of the protein. An illustrative stretch is shown in Fig. 3C. A summary of all the observed connectivities is shown in Fig. 4. It is seen that sequential connections are observed for more than 90% of the non-proline residues. The absence of some connections can possibly be due to unfavorable relaxation behaviors of those residues through the pulse sequence. Fig. 5 shows the assignments so obtained in the ^1H–^{15}N HSQC [9] spectrum, the NMR fingerprint of the protein. This allows an easy comparison with the earlier published data [4].

Discussion

The significant features of the present assignment protocol as compared to the currently used protocols based on the triple resonance experiments HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, etc. are as follows: (i) since the sequential ^{15}N chemical shifts are directly displayed in every plane of the HN(C)N spectrum, there is no need for repetitive scanning of the ^{15}N planes to locate the peaks, (ii) two of the three dimensions in the HN(C)N spectrum have ^{15}N, whose chemical shift dispersion is very good under most conditions, (iii) because of the large number of check points that are generally available, explicit side chain assignment would not be very necessary to decide on the correctness of the

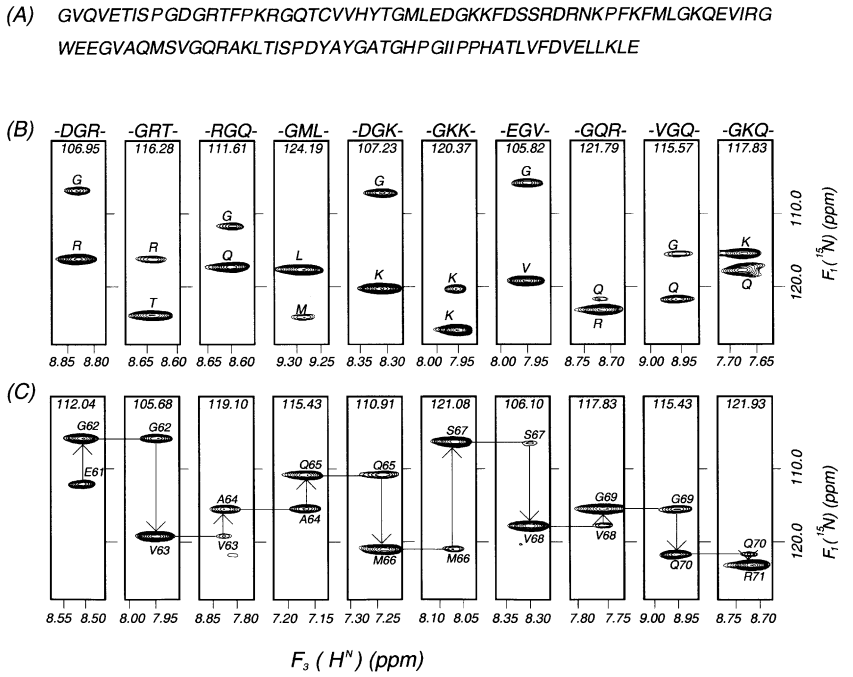

Fig. 3. (A) Amino acid sequence of FKBP with the glycines and prolines highlighted to show their distribution along the sequence. (B) Experimental patterns observed in the HN(C)N spectrum of FKBP; the specific triplet identifications have actually come following complete assignment and these are indicated for clarity. Black and red contours represent the positive and negative peaks, respectively. (C) An illustrative stretch of sequential walk through the HN(C)N spectrum of the FKBP protein. A sequential peak in one plane joins the diagonal peak in the adjacent plane on the right. Note that the panels of G62, V63, G69, and V70 constitute the check points in this sequential walk. The numbers at the top in each panel in B and C identify the F_2 chemical shifts, which help identification of the diagonal peaks.

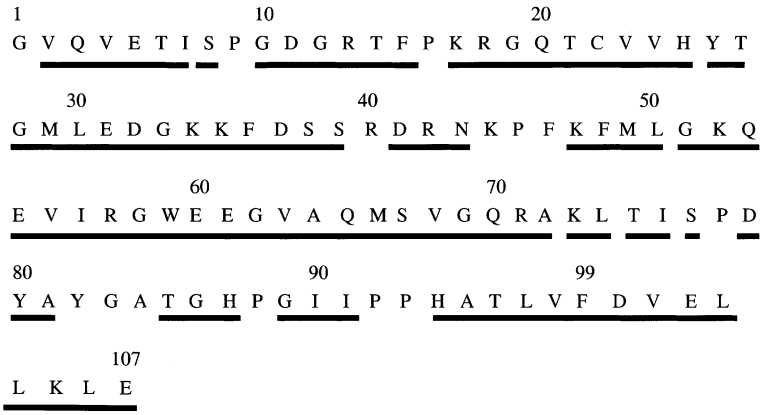

Fig. 4. Summary of all the observed sequential connectivities along the sequence of the protein. The thick line indicates the presence of a connectivity.

amide and ^{15}N assignments, (iv) a single experiment provides direction specific sequential walk; these features make the assignment process very fast and also unambiguous, (v) in large proteins where sometimes a complete assignment by any of the known procedures may be hampered for a variety of reasons, partial contiguous stretches can be assigned unambiguously by the present protocol because of the possibilities of several check points; this is particularly useful for many localized investigations on proteins; and, finally (vi) as the process is very simple and straightforward, ready automation can be envisaged.

The new protocol has also important differences from the HNN [3] based protocol described recently [10], which has also several of the advantages over the conventional methods mentioned above. First of all, in the HNN protocol, where the sequential walk is performed in the HNN spectrum, a HN(C)N spectrum is also essential to discriminate between the two adjacent neighbors of a given residue, to provide directionality to the walk. In contrast, in the present protocol complete sequential walk in a direction specific manner can be achieved entirely in the HN(C)N spectrum itself. This is of great significance in the context of structural

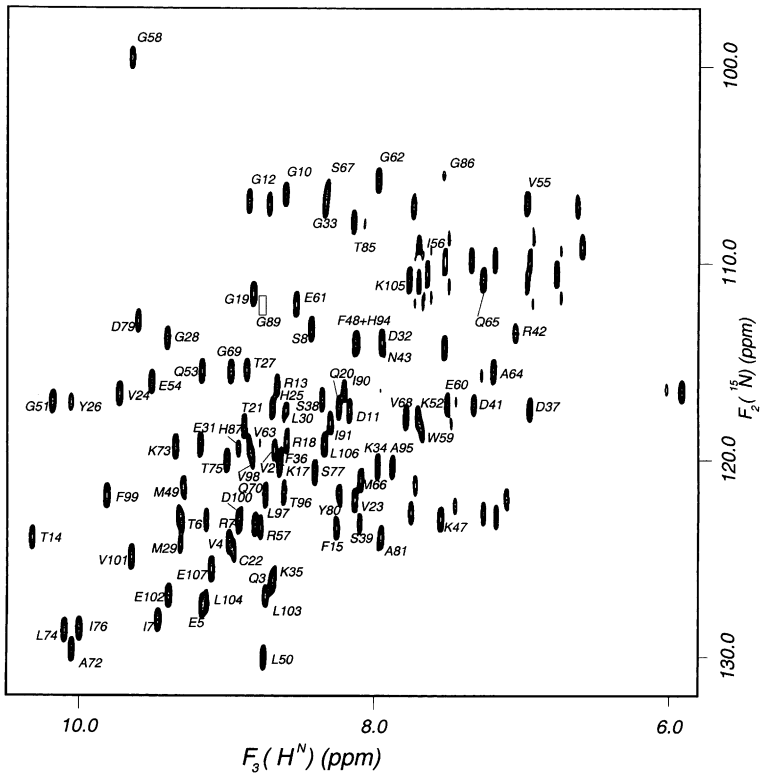

Fig. 5. ^{1}H–^{15}N HSQC spectrum of the FKBP protein showing all the assignments. Some of the unassigned peaks in the spectrum belong to the side chains of asparagines and glutamines. Residues for which connectivities were not found in the HN(C)N spectrum have not been marked. The position of G89 whose peak is seen at lower levels is indicated by a box.

genomics, since, the prospect of success with a single experiment is in tune with the demand for minimum of experimentation. Second, the HN(C)N spectrum has fewer peaks in any plane compared to HNN and thus has less crowding. On the other hand the planes in the HNN spectrum have redundancy of information, since every plane contains correlations to both the 'front' and 'back' neighboring residues. Third, the number of distinctive peak patterns in the HN(C)N spectra is much less compared to that in the HNN spectra. This may be an advantage from the point of view of simplicity, understanding, and reduction of technical jargon.

The success of the present protocol, as with other protocols, for any given protein depends critically on the dispersion of peaks in its ^{1}H–^{15}N HSQC [9] spectrum. Overlap of peaks in this spectrum can lead to cancellation of + and − intensities in the HN(C)N spectrum. This happens under two circumstances: first, when the ^{15}N chemical shifts of neighboring residues are too close, and, second, when the diagonals having + and − intensities have accidentally close ^{15}N chemical shifts. However, we may reiterate that the ^{15}N dispersion is generally very good in most situations, but further improvements, especially for large proteins, can be obtained by deuteration [11], TROSY [12] modifications, and higher magnetic fields.

Finally, a few comments on the sensitivity of the experiment are also in place at this stage. This is important since the experiment employs many steps of magnetization transfer and there could be loss of signal through the pulse sequence. However, our success with different proteins (only one set is shown here) is very encouraging. Further, at the currently available higher magnetic fields, the sensitivities will be much higher. Moreover, the availability of cryoprobes in recent years has pushed the sensitivities even further up by 3–4-fold. With these, both the experimental times and the protein concentrations can be reduced quite significantly. Thus, we believe that the protocol described here will be of very general use.

Acknowledgments

We thank Prof. Michael Rosen for permitting the use of FKBP spectra recorded by Sanjay Panchal in his laboratory. We thank the National Facility for High Field NMR at TIFR, supported by the Government of India, for all the data processing facilities.

References

[1] A.E. Ferentz, G. Wagner, Quart. Rev. Biophys. 33 (2000) 29–65.
[2] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton, Protein NMR Spectroscopy: Principles and Practice, Academic Press, San Diego, CA, 1996.

[3] S.C. Panchal, N.S. Bhavesh, R.V. Hosur, J. Biomol. NMR 20 (2001) 135–147.
[4] M.K. Rosen, S.W. Michnick, M. Karplus, S.L. Schreiber, Biochemistry 30 (1991) 4774–4789.
[5] L.E. Kay, M. Ikura, Tschudin, A. Bax, J. Magn. Reson. 89 (1990) 496–514.
[6] S. Grzesiek, A. Bax, J. Magn. Reson. 96 (1992) 432–440.
[7] S. Grzesiek, A. Bax, J. Magn. Reson. 99 (1992) 201–207.
[8] S. Grzesiek, A. Bax, J. Am. Chem. Soc. 114 (1992) 6291–6293.
[9] G. Bodenhausen, D.J. Ruben, Chem. Phys. Lett. 69 (1980) 185–189.
[10] N.S. Bhavesh, S.C. Panchal, R.V. Hosur, Biochemistry 40 (2001) 14727–14735.
[11] D.M. LcMaster, Quart. Rev. Biophys. 23 (1990) 113–174.
[12] K. Peruvshin, R. Riek, G. Wider, K. Wuthrich, Proc. Natl. Acad. Sci. USA 94 (1997) 12366–12371.