J Biomol NMR (2013) 57:83–89
DOI 10.1007/s10858-013-9783-1

COMMUNICATION

HN(CA)N and HN(COCA)N experiments for assignment of large disordered proteins

Xiao Liu · Daiwen Yang

Received: 18 July 2013 / Accepted: 14 September 2013 / Published online: 20 September 2013

Abstract Two new 3D HN-based experiments are proposed for backbone assignment of large disordered proteins. The spectra obtained with the new pulse schemes are free of redundant diagonal peaks (H_iN_i–N_i) and provide sequential correlations (H_iN_i–N_{i+1} and H_iN_i–N_{i-1}) not only between adjacent non-proline residues but also between non-proline and proline residues. The experiments have been demonstrated on an intrinsically disordered protein with 306 amino acids including 64 proline residues. Using the two experiments, we obtained nearly complete assignments of backbone amides and proline ^{15}N spins except for 4 proline and 4 non-proline residues.

Keywords Disordered protein · Backbone resonance assignment · Mechanosensing protein · Proline-rich protein · Multidimensional NMR

An intrinsically disordered protein (IDP) can adopt multiple conformations rather than a well-defined three dimensional (3D) structure in substantial regions or throughout its entire sequence (Chouard 2011; Dyson and Wright 2005). The unique characteristics of IDPs allow them to recognize multiple partners at a reduced affinity with a faster rate of association and dissociation (Chouard 2011; Oldfield et al. 2008), to enhance specificity with their binding partners, and to facilitate regulation by post translational modifications (Lee et al. 2010). This important class of proteins was largely ignored and overshadowed by the success of studies of proteins with specific 3D structures. With discoveries of more and more IDPs that carry essential biological functions, IDPs have gradually attracted more attention from researchers.

Currently, NMR is most suitable for characterizations of IDP structural ensembles, internal dynamics, and interactions with their binding partners (Rezaei-Ghaleh et al. 2012). In such studies, backbone resonance assignment is often the first essential step. Due to poor dispersion of ^{1}H and $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts for IDPs, the sequence-specific assignment methods used for folded proteins usually fail for relatively large IDPs (>200 residues). Nevertheless, the good chemical shift dispersion of ^{15}N and ^{13}CO spins allows the resonance assignment of IDPs (Dyson and Wright 2001). Several experiments generating sequential amide correlations have been demonstrated to be suitable for IDPs, including 3D (H)N(CA)NH (or HNN) (Panchal et al. 2001; Mukrasch et al. 2009) and 3D (H)N(COCA)NH (or HN(C)N) (Panchal et al. 2001) which were also used for folded proteins (Grzesiek et al. 1993; Weisemann et al. 1993; Löhr et al. 2000; Sun et al. 2005; Frueh et al. 2006), 5D APSY-HN(CA)CONH (Kazimierczuk et al. 2010), 6D APSY-HNCOCANH (Fiorito et al. 2006) and 7D APSY-HNCO(CA)CBCANH (Hiller et al. 2007). In the 5D-7D APSY (automated projection spectroscopy) experiments, cross-peaks with information of chemical shifts of 5–7 correlated nuclei are obtained from the acquired 2D projections of spectra of higher dimensionality. The APSY experiments can resolve most assignment ambiguities encountered in the 3D experiments which result from similar HN (^{1}H and ^{15}N) chemical shifts, and are more suitable for IDPs with highly repetitive sequences. However, the 3D HNN experiment is easy to perform and it can provide sufficient information for assigning relatively small IDPs with the conventional assignment methods and software tools. Using

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9783-1) contains supplementary material, which is available to authorized users.

X. Liu · D. Yang (✉)
Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore
e-mail: dbsydw@nus.edu.sg

this experiment in combination with (HA) CANH experiment, the assignment of the full length tau protein (441 residues) has been achieved by dividing the protein into three overlapped fragments ranging from 168 to 198 amino acids (Mukrasch et al. 2009). Here we propose out-and-back 3D HN(COCA)N and HN(CA)N experiments. They provide not only sequential correlations between non-proline residues as the HNN and HN(C)N experiments do, but also correlations between non-proline and proline residues as the previously proposed HNCAN (Löhr et al. 2000) does. Moreover, the experiments eliminate redundant intra-residue correlations and have significantly less spectral complexity, and thus facilitate resonance assignment of large IDPs.

Figure 1a shows the pulse scheme for correlating backbone amide spin pair $^{1}H_i^{15}N_i$ with spin $^{15}N_{i-1}$ in its preceding residue and $^{15}N_{i+1}$ in its following residue. One-bond and two-bond couplings (J_{NH}, $J_{NC'}$, J_{NCa} and $^{2}J_{NCa}$) are utilized to transfer magnetizations from backbone amide spin $^{1}H_{Ni}$ to the target spins via a number of INEPT transfer steps (Fig. 2a). At point *R* in Fig. 1a magnetizations $N_i^z C_{ai}^z N_{i+1}^y$ and $N_i^z C_{ai-1}^z N_{i-1}^y$ contributing to sequential correlations are generated, while magnetizations $N_i^x C_{ai}^{y} C_{ai-1}^x$ and $N_i^x C_{ai}^x C_{ai-1}^y$ corresponding to intra-residue correlation H_iN_i–N_i also exist. In order to eliminate the redundant intra-residue correlation, a semi-constant-time acquisition mode is introduced into the t1 period from point *R* to *S* where $N_i^z C_{ai}^z N_{i+1}^y$ and $N_i^z C_{ai}^z N_{i-1}^y$ are encoded with the chemical shifts of $^{15}N_{i+1}$ and $^{15}N_{i-1}$ respectively. During this period, $N_i^x C_{ai}^y C_{ai-1}^x$ and $N_i^x C_{ai}^x C_{ai-1}^y$ are converted into $N_i^x C_{ai}^x C_i'^z C_{ai-1}^y C_{i-1}'^z$ and $N_i^x C_{ai}^y C_i'^z C_{ai-1}^x C_{i-1}'^z$ which cannot be transferred back to $^{1}H_{Ni}^z$ for detection. During the t1 period, only ^{15}N–^{13}Ca couplings are removed, but ^{15}N–$^{13}C'$ couplings are partially suppressed and reduced from J_{NC} to $(t_{1max} - 2\delta)/t_{1max} * J_{NC'}$, where t_{1max} is the total acquisition

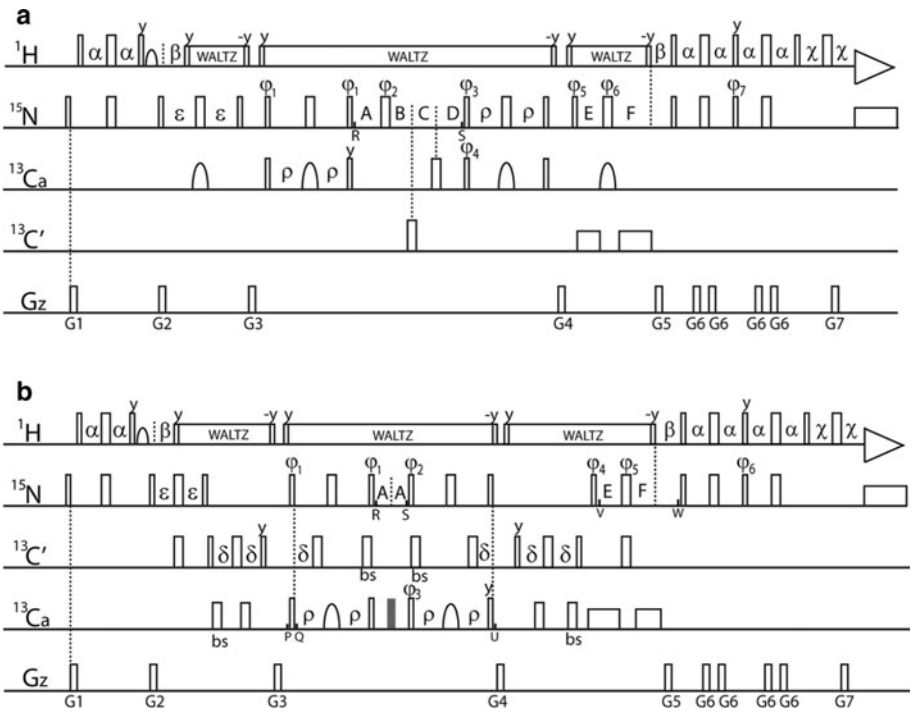

Fig. 1 Pulse schemes for HN(CA)N (**a**) and HN(COCA)N (**b**) experiments. All *narrow* (*wide*) *rectangular* pulses are applied with a flip angle of 90° (180°). The carriers are centered at 4.7 and 119 ppm for ^{1}H and ^{15}N, respectively. The ^{13}C carrier is placed at 56 ppm in (**a**), while it is set at 56 ppm during the period between points *P* and *U* and 176 ppm during other periods in (**b**). The ^{1}H shaped pulse (90°) has a one-lobe sinc profile (1.0 ms). All ^{13}C 90° and 180° pulses (*open rectangles*) are applied with field strengths of $\Delta/\sqrt{15}$ and $\Delta/\sqrt{3}$, respectively, where Δ is the resonant frequency difference of ^{13}Ca and $^{13}C'$ spins. The shaped ^{13}C 180° pulses have a REBURP profile (400 us) centered at 42 ppm. The *shaded rectangle* ^{13}C pulse applied in the middle of the t1 period in (**b**) is a high power composite 180° pulse ($90_x180_y90_x$) with carrier centered at 56 ppm. Bloch-Siegert compensation pulses are applied at the positions indicated by '*bs*'. The ^{1}H decoupling field strength is 5 kHz. Quadrature detections were achieved using State-TPPI of φ_1 for t1 dimension and the enhanced sensitivity pulse field gradient method (Kay et al. 1992) for t2 dimension, respectively. The peak strengths of sine-shaped gradients with duration of 1 ms were: $G1 = 10$ G/cm, $G2 = 30$ G/cm, $G3 = 15$ G/cm, $G4 = 25$ G/cm, $G5 = -40$ G/cm, $G6 = 17.5$ G/cm, $G7 = 4.05$ G/cm. In (**a**), the delays used were: $\alpha = 2.5$ ms, $\beta = 5.4$ ms, $\chi = 1.1$ ms, $\varepsilon = 13$ ms, $\rho = 14$ ms, $A = 4.5$ ms $- 0.5t1'$, $B = 4.5$ ms, $C = 0.5t1 - 0.5t1'$, $D = 0.5t1$, $E = \varepsilon - 0.5t2$ and $F = \varepsilon - \beta + 0.5t2$; the phase cycling used was: $\varphi_1 = x, x, -x, -x$, $\varphi_3 = x, x, x, x, -x, -x, -x, -x$, $\varphi_4 = y, -y$, $\varphi_5 = x$, $\varphi_6 = x, -x, y, -y$, $\varphi_7 = y$, $\varphi_{rec} = x, -x, -x, x$. In (**b**), the delays used were: $\alpha = 2.5$ ms, $\beta = 5.4$ ms, $\chi = 1.1$ ms, $\varepsilon = 15$ ms, $\delta = 4.5$ ms, $\rho = 12.5$ ms, $A = 0.5t1$, $B = 0.5t1$, $E = \varepsilon - 0.5t2$ and $F = \varepsilon - \beta + 0.5t2$; the phase cycling used was: $\varphi_1 = x, x, x, x, -x, -x, -x, -x$, $\varphi_3 = y, -y$, $\varphi_4 = x$, $\varphi_5 = x, -x, y, -y$, $\varphi_6 = y$, $\varphi_{rec} = x, -x, -x, x$

time in the t1 dimension and $\delta = 1/4J_{CaC'}$. When $t_{1max} = 24$ ms and $\delta = 4.5$ ms which were employed here, the reduced coupling is about 10 Hz, much smaller than the acquisition resolution and will not affect the spectral resolution significantly.

Neglecting the relaxation effect during the delays α and χ in Fig. 1, the relative signal intensities for sequential H_iN_i–N_{i+1} (I_f) and H_iN_i–N_{i-1} (I_p) correlations are given by.

$$I_f = \sin^2(2\pi J_{CaN}\varepsilon)\cos^2(2\pi^2 J_{CaN}\varepsilon)\sin^2(2\pi^2 J_{CaN}\rho)\cos^2(2\pi^2 J_{CaN}\rho)\cos^2(2\pi J_{CaCb}\rho)e^{-[R_{2N}(4\varepsilon+4\rho+2\delta)+4R_{2Ca}\rho]} \quad (1)$$

$$I_p = \sin^2(2\pi^2 J_{CaN}\varepsilon)\cos^2(2\pi J_{CaN}\varepsilon)\sin^2(2\pi J_{CaN}\rho)\cos^2(2\pi J_{CaN}\rho)\cos^2(2\pi J_{CaCb}\rho)e^{-[R_{2N}(4\varepsilon+4\rho+2\delta)+4R_{2Ca}\rho]} \quad (2)$$

In Eqs. 1 and 2, ε, ρ and δ are delays used in the pulse sequence (Fig. 1), R_{2N} and R_{2Ca} are the transverse relaxation rates of backbone ^{15}N and ^{13}Ca spins. For many intrinsically disordered proteins, the average R_{2N} values are relative small, e.g., $\sim 3.5\ s^{-1}$ for α-synuclein (15 °C, 800 MHz NMR, pH 6.2) (Kim et al. 2013), $\sim 4\ s^{-1}$ for tau fragments (5 °C, 700 MHz NMR) (Mukrasch et al. 2009) and $\sim 5\ s^{-1}$ for the protein studied here (25 °C, 800 MHz NMR). For a rigid protein with a tumbling time of 2.2 ns, the R_{2N}, R_{2Ca}, R_{2C}, (relaxation rate of carbonyl carbon) and R_{1N} (longitudinal relaxation rate of ^{15}N) on an 800 MHz NMR spectrometer are 5, 11, 6.3 and 2.6 s^{-1}. To compare experimental sensitivities of different pulse schemes, we estimated the relative peak intensities for all the experiments mentioned in this paper using these relaxation rates together with following J coupling constants: $J_{CaN} = 10$ Hz, $^2J_{CaN} = 7$ Hz, $J_{C'N} = 15$ Hz, $J_{CaCb} = 35$ Hz and $J_{CaC'} = 55$ Hz.

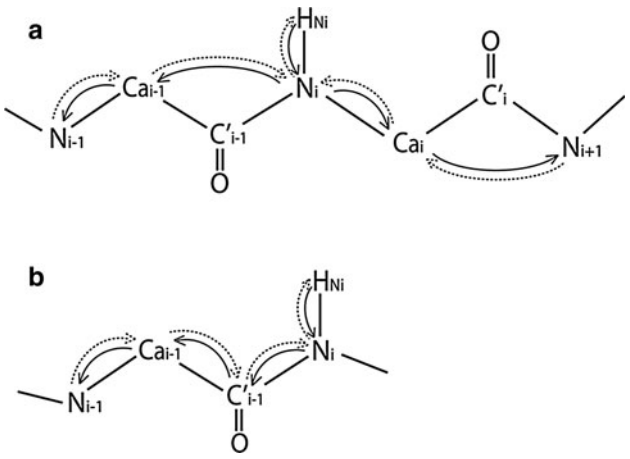

Fig. 2 Diagrams showing the magnetization transfer pathways for HN(CA)N (**a**) and HN(COCA)N (b) experiments. The *solid* and *dashed arrows* indicate the out and back pathways, respectively

I_f and I_p reach maximal intensities at slightly different delays (Fig. S1). Considering both I_f and I_p, the delays ε and ρ were set as 13 and 14 ms respectively, giving rise to $I_f = 0.026$ and $I_p = 0.010$. In the HNN experiment, I_f and I_p are identical and the maximal intensity is 0.058 (Panchal et al. 2001). Although the HN(CA)N experiment proposed here is about two times and six times less sensitive than HNN for H_iN_i–N_{i+1} and H_iN_i–N_{i-1} correlations respectively, the HN(CA)N spectrum is easier to analyze due to its absence of intra-residue correlations. Moreover, unlike the HNN experiment, the HN(CA)N provides sequential correlations of a non-proline residue to proline residues. Thus it is easy to assign PXP and XPX motifs (where X represents any non-proline amino acid) with the HN(CA)N experiment. The out-and-back HNCAN experiment proposed previously (Löhr et al. 2000) also gives sequential correlations between non-proline and proline residues. In the ^{13}Ca–^{15}N INEPT version of this experiment, the relative intensities of H_iN_i–N_{i+1}, H_iN_i–N_{i-1} and H_iN_i–N_i correlations are 0.031, 0.011 and 0.032. In the ^{13}Ca–^{15}N HMQC version of the HNCAN (HMQC–HNCAN) experiment, the optimized intensities of H_iN_i–N_{i+1}, H_iN_i–N_{i-1} and H_iN_i–N_i correlations are 0.037, 0.014 and 0.002 (considering the relaxation effect). It is noteworthy that the magnetization involving C_a^x or C_a^y evolves over 56 ms and 84 − 84 + t_{1max} ms for HN(CA)N and HMQC–HNCAN, respectively, and C_a–C_b coupling is active and can only be decoupled partially during the t1 period for the HMQC–HNCAN experiment, but the coupling is not active for HN(CA)N. Although the HMQC version experiment is about 40 % more sensitive than the experiment proposed here, suppression of the diagonal peaks (or H_iN_i–N_i correlations) is incomplete. The residual diagonal peaks for the residues with high flexibility can be comparable in peak intensities to the sequential cross-peaks for the residues with low flexibility or slow conformational exchange and thus interfere with the identification of sequential correlations when two or more amides have similar HN chemical shifts.

To differentiate H_iN_i–N_{i-1} from H_iN_i–N_{i+1} correlations in an HN(CA)N spectrum and overcome the poor sensitivity for the H_iN_i–N_{i-1} correlations, one can use another out-and-back experiment, HN(COCA)N, to obtain only H_iN_i–N_{i-1} correlations with the pulse scheme shown in Fig. 1b. In this experiment, magnetization transfer is achieved through one-bond J couplings (Fig. 2b). Magnetization $N_i^z C'^z_{i-1} C^z_{ai-1}$ at point *P* in the sequence is generated from $^1H^z_{Ni}$ of residue i via the first three INEPT transfer steps. During the period from *P* to *R*, $N_i^z C'^z_{i-1} C^z_{ai-1}$ evolves to become $N_i^z C^z_{ai-1} N^y_{i-1}$. To maximize the magnetization transfer, the one-bond ^{15}N–^{13}Ca coupling (J_{NCa}) is activated but the two-bond coupling ($^2J_{NCa}$) is deactivated by creating magnetization $N_i^y C'^z_{i-1} C^y_{ai-1}$ at point *Q*, and the transfer time is set to about

$1/J_{CaCb}$ since the one-bond ^{13}Ca–^{13}Cb coupling interaction is also active during this period. During the subsequent t1 period from point *R* to *S*, magnetization $N_i^z C_{ai-1}^z N_{i-1}^y$ is encoded with the chemical shift of $^{15}N_{i-1}$ of residue i−1. The magnetization contributing to the intra-residue correlation is eliminated by phase cycling. The ^{15}N–^{13}Ca and ^{15}N–^{13}CO couplings are removed by employing a ^{13}C composite 180° pulse in the middle of the t1 period. Subsequently, the magnetization is transferred back to $^{15}N_i$ and is encoded with the $^{15}N_i$ chemical shift during the constant-time t2 period from *V* to *W*. Finally, the magnetization is transferred back to $^{1}H_{Ni}$ for detection.

For the HN(COCA)N experiment, the cross-peak intensity (I_p) is given by.

$$I_p = \sin^2(2\pi J_{NC'}\varepsilon)\sin^4(2\pi J_{C'Ca}\delta)\sin^2(2\pi J_{CaN}\rho)\cos^2(2\pi J_{CaN}\rho)\cos^2(2\pi J_{CaCb}\rho)e^{-[R_{2N}(4\varepsilon+4\rho)+4R_{2CO}\delta+4R_{2CA}\rho]} \quad (3)$$

The maximum intensity of the H_iN_i–N_{i-1} correlation is 0.062, smaller than that for the previously proposed HN(C)N experiment (0.145), but similar to that for the HNN (0.058). Despite its lower sensitivity in comparison with HN(C)N, the HN(COCA)N shown here is free of diagonal peaks, which is critical for backbone assignment of large disordered proteins using 3D experiments.

The two experiments proposed here were demonstrated on the substrate domain of rat p130Cas (p130CasSD), which is an intrinsically disordered protein domain. p130CasSD consists of 306 amino acids including 64 proline residues and contains 15 YXXP and 8 PXP motifs (Fig. S2). Upon phosphorylation of the tyrosine in the YXXP motifs, p130CasSD can bind many SH2- or PTB-containing proteins and then triggers downstream biological activities. Unlike other proteins, p130CasSD is phosphorylated much more efficiently under mechanical stretching and thus named as a mechno-sensor protein (Sawada et al. 2006). NMR experiments were performed on a sample of 1 mM ^{13}C,^{15}N-labeled p130CasSD, 20 mM sodium phosphate and 5 % D_2O (pH 6.8). 3D HN(CA)N, HN(COCA)N and HNN were acquired at 25 °C on a Bruker 800 MHz spectrometer equipped with a cryo-probe. Each 3D data set was comprised of 45 × 45 × 640 complex points with spectral widths of 1865, 1865 and 11160 Hz in F1(^{15}N), F2(^{15}N) and F3(^{1}H) dimensions. 8 scans were accumulated for each point in the indirect dimensions with a relaxation delay of 1 s, resulting in a total of experimental time of 22.5 h for each 3D experiment.

Due to the large size and presence of multiple repetitive motifs, the ^{1}H–^{15}N correlations of p130CasSD are highly degenerated (Fig. S3). Thus the HNN spectrum is too crowded to analyze in some regions since each NH gives rise to three correlations (Fig. 3a). On the other hand, the HN(CA)N and HN(COCA)N spectra are less crowded (Figs. 3b and c). In the case where a H_iN_i–N_i correlation overlaps with H_iN_i–N_{i+1} or/and H_iN_i–N_{i-1} correlations

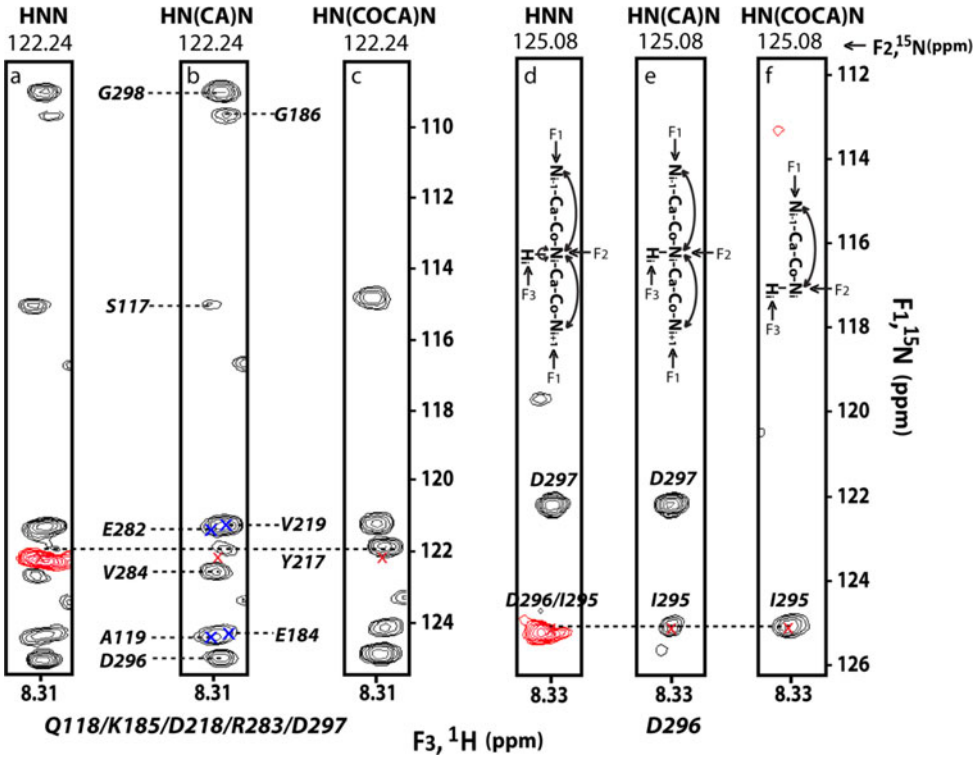

Fig. 3 Representative F1(^{15}N)–F3(^{1}H) slices taken from HNN (**a**, **d**), HN(CA)N (**b**, **e**) and HN(COCA)N (**c**, **f**). The F2(^{15}N) frequency is shown on the *top* of each slice. The peaks in *red* are the H_iN_i–N_i correlations (diagonal peaks) in (**a**) and (**d**). The diagonal peaks in **b**, **c**, **e** and **f** are not present, but their locations are marked by *red* "x". The *blue* "x" represents the center of a sequential correlation when it is overlapped partially with another correlation. Schematic representations of the correlations among three different spins are shown in **d** (for HNN), **e** (for HN(CA)N) and **f** (for HN(COCA)N)

(e.g., overlaps of the diagonal peak of D218 (in red) with the cross-peak D218-Y217 in Fig. 3a and the diagonal of D296 with cross-peak D296-I295 in Fig. 3d), the HNN experiment fails to provide the sequential correlation for residue i (e.g., D218-Y217 and D296-I295), but the HN(CA)N and HN(COCA)N can provide such correlations (Figs. 3a–f). When H_iN_i–N_{i+1} overlaps with H_iN_i–N_{i-1}, H_jN_j–N_{j-1} or H_jN_j–N_{j+1} (e.g., overlaps of Q118-A119 with K185-E184 and D218-V219 with R283-E282 in Fig. 3b), HN(COCA)N is needed to identify sequential correlations for the residues involved in the overlaps (e.g., Q118, K185, D218 and R283).

For p130CasSD, 241 H_iN_i–N_{i+1} and 242 H_iN_i–N_{i-1} correlations are expected. In the HN(COCA)N spectrum, except for 4 residues (H44, H45, S46 and H192) which did not display observable ^1H–^{15}N HSQC cross-peaks, the H_iN_i–N_{i-1} correlations of all other residues (238) were unambiguously detected. In the HN(CA)N spectrum, 213 H_iN_i–N_{i+1} and 214 H_iN_i–N_{i-1} correlations were well resolved, while 48 cross-peaks were involved in peak overlaps. With use of both HN(CA)N and HN(COCA)N, all expected H_iN_i–N_{i+1} correlations (237) were identified except for those from H44-S46 and H192. In the HNN spectrum, 169 H_iN_i–N_{i+1} and 167 H_iN_i–N_{i-1} correlations were unambiguously identified.

Figure 4 shows the intensity ratios of the peaks observed in the HNN to those in the HN(CA)N and HN(COCA)N. Overall, the experimental intensity ratios are consistent with the predicted ratios: ~2.2 for I_f[HNN]:I_f[HN(CA)N], 5.8 for I_p[HNN]:I_p[HN(CA)N], and ~0.9 for I_p[HNN]:I_p[HN(COCA)N]. The large distribution of the intensity ratios results from the variations in J_{NCa} and $^2J_{NCa}$ and dynamics from one residue to another. Although the HN(CA)N is less sensitive than the HNN experiment, more sequential correlations were identified.

With the nearly complete sequential correlations, assignment of backbone amides and proline ^{15}N spins can be achieved using the conventional strategy in following four steps. First, clusters are constructed by grouping the HN(CA)N and HN(COCA)N peaks with identical NH chemical shifts. Second, H_iN_i–N_{i+1} and H_iN_i–N_{i-1} correlations are differentiated by comparing HNCA(N) and HN(COCA)N spectra, and the locations of H_iN_i–N_i diagonal peaks are marked in the spectra. In this step, amino acid types are roughly classified based on ^{15}N chemical shifts: $\delta > 133$ ppm as Pro, $\delta < 110$ ppm as Gly, 110 ppm $< \delta <$ 115 ppm as Thr, Ser or Gly, 125 ppm $< \delta <$ 131 ppm as Ala, Leu, Val or Lys. Only the classification for Pro is accurate. Third, connectivity fragments are built from clusters by matching H_iN_i–N_i and H_iN_i–N_{i+1} in cluster i with H_jN_j–N_{j-1} and H_jN_j–N_j in cluster j respectively. Fourth, the fragments are mapped onto the protein sequence based on amino acid type information. For p130CasSD, we obtained 234 clusters each with one H_iN_i–N_{i-1} and one H_iN_i–N_{i+1} peak (i.e., each cluster corresponds to a single amide) and 2 clusters each with two H_iN_i–N_{i-1} and two H_iN_i–N_{i+1} peaks (i.e., each cluster corresponds to two completely degenerate amides). Although the 2D ^1H–^{15}N HSQC spectrum is very crowded in the middle region (Fig. S3), in most cases the amides in the overlapped 2D HSQC peaks have slight differences in ^{15}N or/and ^1H chemical shifts and their corresponding clusters could be identified manually. For example, five amides (Q118, K185, D218, R283 and D297) overlap as one peak in the HSQC (Fig. S3), four clusters were identified from Fig. 3b and c, and only one of them contained information from two degenerate amides (K185 and D218). Each of the clusters with two amides was divided into four possible sub-clusters by random combinations of two H_iN_i–N_{i-1} with two H_iN_i–N_{i+1} peaks. The wrong combinations were ruled out in steps 3 and 4. Except 21 clusters, other clusters were uniquely formed connectivity fragments. Among the 21 clusters, 7 clusters each could match two other clusters, resulting in assignment ambiguities. The

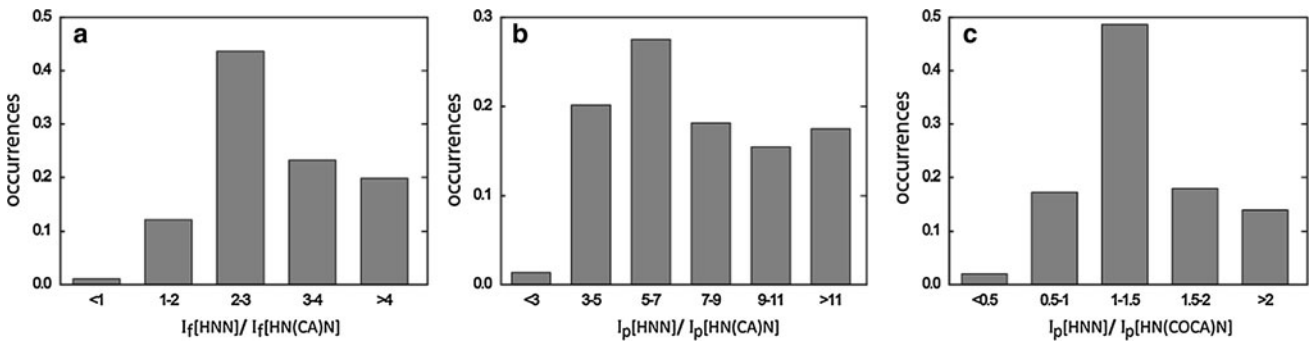

Fig. 4 Distribution of intensity ratios of H_iN_i–N_{i+1} peaks in the HNN to those in the HN(CA)N spectrum (I_f[HNN]/I_f[HN(CA)N], **a**) H_iN_i–N_{i-1} peaks in the HNN to those in the HN(CA)N spectrum (I_p[HNN]/I_p[HN(CA)N], **b**) and H_iN_i–N_{i-1} peaks in the HNN to those in the HN(COCA)N spectrum (I_p[HNN]/I_p[HN(COCA)N], **c**)

connectivity breaks at poly-proline sites ($[P]_n$, $n \geq 2$), the amides that give rise to very weak or no ^{1}H–^{15}N HSQC cross-peaks, and ambiguous sequential linkage sites. Thus long connectivity fragments can be established even for proline-rich proteins. The longest fragment for p130CasSD contained 30 residues including two Pro residues, while the shortest one contained only two residues. Figure 5 shows sequential connectivity of one portion (E288–D296) of a fragment. 14 relatively long fragments each with >10 non-proline residues were mapped onto the sequence in a straightforward manner. Subsequently, short fragments with 6–7 non-proline residues were mapped. Finally, a couple of fragments with only two residues were mapped. The assignment ambiguities were resolved in the process of mapping. Except for H44–S46 and H192, all other backbone amides (238) and 60 proline ^{15}N spins were assigned.

P130CasSD contains a significant amount of Pro (20.9 %), Gly (6.9 %), Ser (6.2 %), Ala (6.2 %) and Thr (4.9 %). Thus in the absence of side-chain ^{13}C chemical shifts, we had no difficulty to map the fragments onto the sequence. If a large IDP consists of small numbers of Pro and Gly, one can record a (H)C(CO)NH-TOCSY (Montelione et al. 1992) experiment to determine the amino acid types of clusters. When the amino acid types for most clusters are identified, even short fragments with 2–3 residues can be easily mapped. We performed the TOCSY experiment and confirmed that our assignments are correct.

The amides in the PXP motifs could not be assigned with the HNN and HN(C)N experiments since only intra-residue correlations H_iN_i–N_i are available for these motifs. They are also difficult to be assigned with the conventional HN-based experiments such as HNCO, HN(CA)CO, HNCA and HN(CO)CA. Using the HN(CA)N and HN(COCA)N experiments, the assignment of such motifs is straightforward (Fig. 6). After obtaining assignments of backbone amides, one can assign side-chains using the TOCSY-based experiments like (H)C(CO)NH-TOCSY and (H)CCH-TOCSY (Fesik et al. 1990). To obtain the assignments of proline residues in poly proline regions,

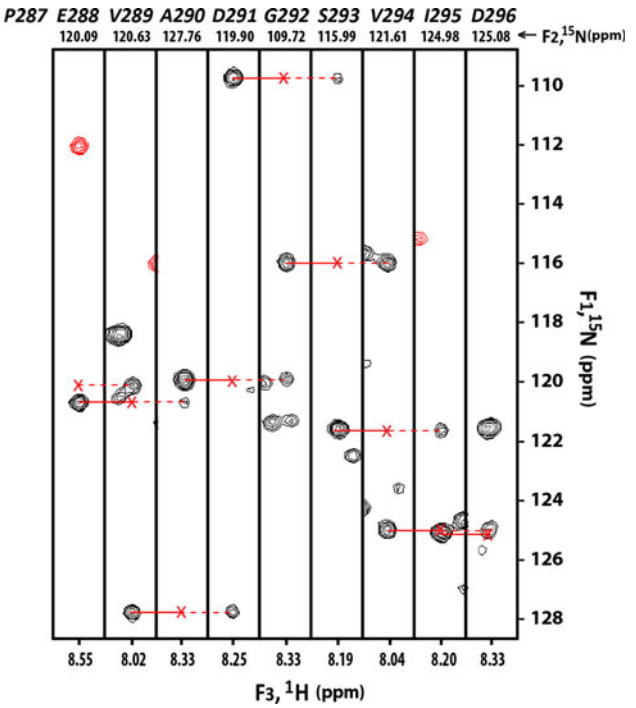

Fig. 5 HN(CA)N F1–F3 slices for residues E288-D296, showing the establishment of sequential connectivity based on the match of H_iN_i–N_i in cluster i with H_jN_j–N_{j-1} in cluster j (*dashed line*) and the match of H_iN_i–N_{i+1} with H_jN_j–N_j (*solid line*) in F1 frequencies. The F2 ^{15}N frequency of each residue is labeled on the *top* of each slice. The peaks in *red*, which are the correlations between amides (N_iH_i) and their adjacent proline ^{15}N (N_{i+1} or N_{i-1}), are aliased by 23 ppm. The locations of diagonal peaks (which are absent in the spectra) are marked by *red* "x"

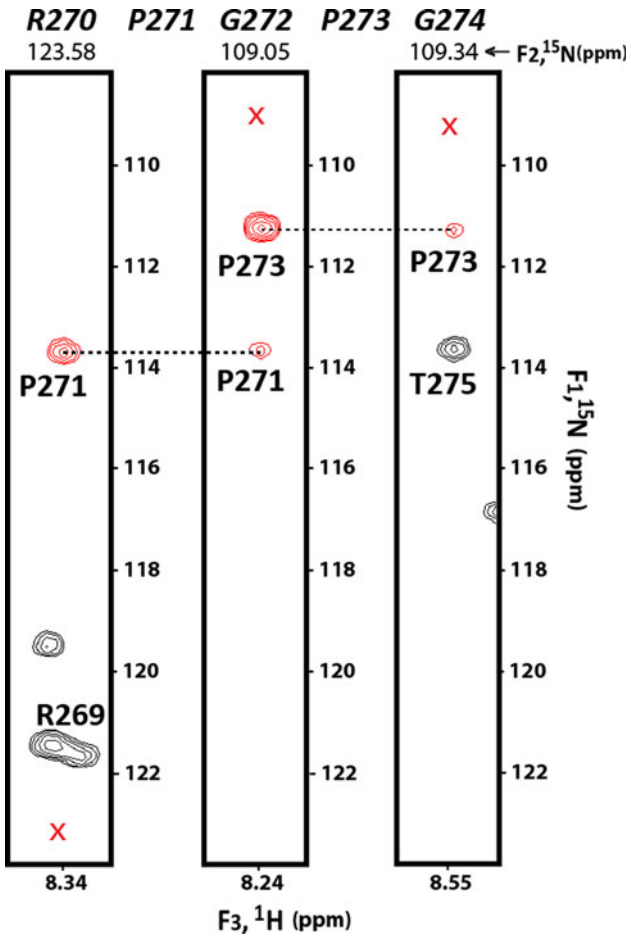

Fig. 6 HN(CA)N F1–F3 slices for R270, G272 and G274, demonstrating the assignment of PXP motifs. The match of H_iN_i–N_{i-1} in cluster i with H_jN_j–N_{j+1} in cluster j is indicated by a *dashed line*. The peaks in *red*, which are the correlations between amides (N_iH_i) and their adjacent proline ^{15}N (N_{i+1} or N_{i-1}), are aliased by 23 ppm. The locations of diagonal peaks (which are absent in the spectra) are marked by *red* "x"

HACA-based experiment (Kanelis et al. 2000) may be employed.

In summary, the HN(CA)N and HN(COCA) experiments proposed here provide significantly more sequential correlations than the previously developed experiments because they suppress the redundant intra-residue diagonal signals. Using these two experiments, nearly complete backbone assignments can be achieved for large disordered proteins even with a high percentage of proline residues.

The chemical shifts of backbone $^{1}H_{N}$ and ^{15}N spins were deposited in the BMRB database under the accession code 19477.

Acknowledgments The authors thank Prof. Sawada in Mechanobiology Institute, National University of Singapore for providing the p130CasSD construct. This work is supported by a grant from Singapore Ministry of Education Academic Research Fund Tier 1 (R154000555112).

References

Chouard T (2011) Breaking the protein rules. Nature 471:151–153

Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–270

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208

Fesik SW, Eaton HL, Olejniczak ET, Zuiderweg ERP, McIntosh LP, Dahlquist FW (1990) 2D and 3D NMR spectroscopy employing carbon-13/carbon-13 magnetization transfer by isotropic mixing. Spin system identification in large proteins. J Am Chem Soc 112:886–888

Fiorito F, Hiller S, Wider G, Wüthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR 35:27–37

Frueh DP, Sun ZY, Vosburg DA, Walsh CT, Hoch JC, Wagner G (2006) Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. J Am Chem Soc 128:5757–5763

Grzesiek S, Anglister J, Ren H, Bax A (1993) 13C Line narrowing by 2H decoupling in 2H/13C/15 N-enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. J Am Chem Soc 115:4369–4370

Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J Am Chem Soc 129:10823–10828

Kanelis V, Donaldson L, Muhandiram D, Rotin D, Forman-Kay J, Kay L (2000) Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. J Biomol NMR 16:253–259

Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

Kazimierczuk K, Zawadzka-Kazimierczuk A, Kozminski W (2010) Non-uniform frequency domain for optimal exploitation of nonuniform sampling. J Magn Reson 205:286–292

Kim S, Wu KP, Baum J (2013) Fast hydrogen exchange affects 15 N relaxation measurements in intrinsically disordered proteins. J Biomol NMR 55:249–256

Lee CW, Ferreon JC, Ferreon AC, Arai M, Wright PE (2010) Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. Proc Natl Acad Sci USA 107:19290–19295

Löhr F, Pfeiffer S, Lin YJ, Hartleib J, Klimmek O, Rüterjans H (2000) HNCAN pulse sequences for sequential backbone resonance assignment across proline residues in perdeuterated proteins. J Biomol NMR 18:337–346

Montelione GT, Lyons BA, Emerson SD, Tashiro M (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 114:10974–10975

Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M (2009) Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol 7:e34

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 9(Suppl 1):S1

Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. J Biomol NMR 20:135–147

Rezaei-Ghaleh N, Blackledge M, Zweckstetter M (2012) Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. Chem Bio Chem 13:930–950

Sawada Y, Tamada M, Dubin-Thaler BJ, Cherniavskaya O, Sakai R, Tanaka S, Sheetz MP (2006) Force sensing by mechanical extension of the Src family kinase substrate p130Cas. Cell 127:1015–1026

Sun ZY, Frueh DP, Selenko P, Hoch JC, Wagner G (2005) Fast assignment of 15 N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. J Biomol NMR 33:43–50

Weisemann R, Rüterjans H, Bermel W (1993) 3D triple-resonance NMR techniques for the sequential assignment of NH and 15 N resonances in 15 N- and 13C-labelled proteins. J Biomol NMR 3:113–120