ARTICLE

Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins

David Pantoja-Uceda · Jorge Santoro

Abstract Two novel 3D ^{13}C-detected experiments, hNcocaNCO and hnCOcaNCO, are proposed to facilitate the resonance assignment of intrinsically disordered proteins. The experiments correlate the ^{15}N and ^{13}C′ chemical shifts of two consecutive amide moieties without involving other nuclei, thus taking advantage of the good dispersion shown by the ^{15}N–^{13}C′ correlations, even for proteins that lack a well defined tertiary structure. The new pulse sequences were successfully tested using Nupr1, an intrinsically disordered protein of 93 residues.

Introduction

Over the last 20 years, experimental and computational structural biologists have accumulated evidences that many proteins or protein domains (intrinsically disordered proteins, IDPs) lack a well defined tertiary structure under functional conditions (Dunker et al. 2008; Fink 2005; Tompa 2002, 2011, 2012). This has brought an increased interest in characterizing these proteins. In these studies NMR plays a central role (Dyson and Wright 2004; Eliezer 2009), since it can provide residue-level parameters carrying local structure information, like chemical shifts, residual dipolar couplings or relaxation rates.

D. Pantoja-Uceda · J. Santoro (✉)
Instituto de Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain
e-mail: jsantoro@iqfr.csic.es

Every protein NMR study starts with the sequence specific assignment. For globular proteins a well established strategy exists (Permi and Annila 2004; Sattler et al. 1999), that mainly consists in connecting two consecutive NH's through their correlations with one or more of the ^{13}C spins located between them, ^{13}Cα, ^{13}Cβ and ^{13}C′. In the case of IDPs the application of this assignment strategy is seriously compromised, owing to poorly dispersed amide ^{1}H peaks. Besides, the ^{13}Cα and ^{13}Cβ chemical shifts are clustered around the random coil value for each amino acid residue type, further complicating the assignment process. Therefore alternative assignment strategies have been proposed. Bermel et al. (2005, 2009a) developed an approach based on ^{13}C detection, while Mäntylahti et al. (2010, 2011) proposed a strategy based on Hα detection. Although both strategies make use of the ^{15}N and ^{13}C′ signals, that remain well dispersed in IDPs (Dyson and Wright 2001; Yao et al. 1997; Zhang et al. 1997), also include other nuclei with poorer dispersion, ^{13}Cα, ^{13}Cβ or Hα, which may adversely affect the assignment process. To overcome these difficulties high–dimensionality experiments, 4D or 5D, have been proposed (Bermel et al. 2012a; Motáčková et al. 2010; Novácek et al. 2011, 2012). As a simpler alternative, we propose here a new pulse sequence that makes full use of the good dispersion of the ^{15}N and ^{13}C′ chemical shifts, since it correlates two consecutive ^{13}C′–^{15}N groups in a protein. The novel pulse sequence, which requires (because of the multiple coherence transfer steps involved) a slow transverse relaxation, is designed for IDPs and is superior in terms of peak dispersion and the easiness of determining sequential connectivity's. The experiment is the result of a modification of the HNcocaNH experiment. (Grzesiek et al. 1993; Panchal et al. 2001; Pantoja-Uceda and Santoro 2009; Sun et al. 2005). The main modification consists of replacing the last

^{15}N → ^{1}H transfer by a ^{15}N → ^{13}C′ transfer. Also, unlike usual HNcocaNH experiments, in which the initial ^{1}H and ^{15}N chemical shifts are labeled, in the new experiment the chemical shifts of the initial ^{15}N and ^{13}C′ spins are labeled. Kumar et al. (2010) have also proposed the labeling of ^{13}C′ in an hnCOcaNH experiment. However, this experiment involves the amide ^{1}H, whose poor dispersion in IDPs makes it not very suitable for their study.

Materials and methods

NMR measurements were performed at 18.8 T on a Bruker AV 800 spectrometer equipped with a cryogenically cooled triple-resonance (^{1}H, ^{13}C, ^{15}N) TCI probe, and pulsed z-field gradients. The proposed pulse sequences were tested on a sample of ^{13}C, ^{15}N labeled Nupr1 (1 mM, 10 mM acetate buffer, pH 4.5), an IDP of 93 residues. All spectra were recorded with spectral widths of 14 and 40 ppm centered at 173.5 and 122 ppm for ^{13}C′ and ^{15}N, respectively. For the 2D CON spectrum 512 and 128 complex data points were acquired in t_1 (^{13}C′) and t_2 (^{15}N) dimensions. For each t_1 increment, 8 scans were accumulated and the total experimental time was 44 min. The 3D spectra were acquired with 512 complex points in the direct dimension and 28 complex points in both indirect dimensions, using 8 transients per FID and a recycle delay of 1 s. The total acquisition time for each 3D spectrum was 10 h. The programs NMRPipe (Delaglio et al. 1995) and NMRView (Johnson and Blevins 1994) were used for spectral processing and data analysis, respectively. Although a 16-step phase cycle is described in the legend of Fig. 2, clean 3D spectra have been obtained using shorter phase cycles of 8 or even 4 steps.

Results and discussion

Figure 1 shows schematically the magnetization transfer pathway implemented in the hNCOcaNCO experiment and Fig. 2 shows the corresponding radio-frequency pulse sequence. Correlating two consecutive ^{13}C′–^{15}N groups with the hNCOcaNCO pulse sequence would require a 4D spectrum, what involves a very long experimental time to obtain satisfactory resolution. Nevertheless, the same information can be obtained from two 3D spectra, one with the correlation N(i + 1)–N(i)–C′(i − 1) and another one with the correlation C′(i)–N(i)–C′(i − 1). This approach requires a much shorter experimental time, and is the one we have used and will be described below. The pulse sequence starts with an INEPT transfer of magnetization from the amide proton of residue i + 1 to its directly bonded ^{15}N, generating the coherence $-2H_z(i + 1)N_y(i + 1)$. In the next period, the
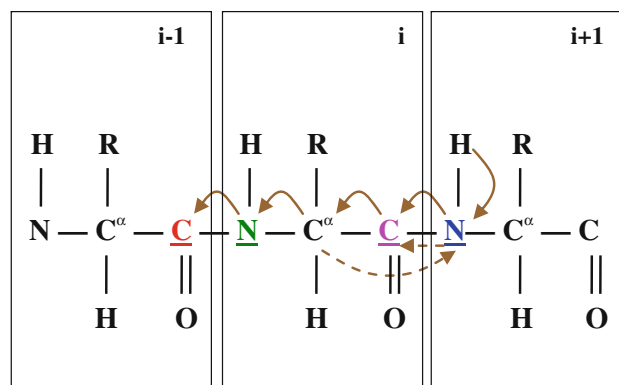

Fig. 1 Schematic representation of the magnetization transfer pathway implemented in the HNCOCANCO experiment. The *colored underlined* nuclei are frequency labeled and used to correlate residues i and i + 1. *Arrows* indicate the magnetization transfer pathway. The pathway with *broken arrows* can be suppressed with an appropriate delay setting

^{1}J$_{NH}$ coupling is refocused and the ^{15}N magnetization evolves the ^{1}J$_{NC'}$ coupling to become anti-phase with respect to its attached carbonyl carbon, $-2N_y(i + 1)O_z(i)$. In the 3D hNcocaNCO experiment this period is also used to label the coherence with the chemical shift of the ^{15}N(i + 1) spin in a constant time manner. The following pair of 90° pulses transfer the coherence to ^{13}C′(i), $2O_y(i)N_z(i + 1)$. During the next time period the ^{1}J$_{C\alpha C'}$ coupling evolves, so that the ^{13}C′(i) coherence results in anti-phase to both ^{15}N(i + 1) and ^{13}Cα(i), $4O_x(i)A_z(i)N_z(i + 1)$. In the hnCOcaNCO experiment, this period is also used to label the coherence with the chemical shift of the carbonyl carbon in either a constant time or a semi-constant time mode. After that, a pair of 90° pulses, applied to ^{13}C′ and ^{13}Cα, transfer the coherence to ^{13}Cα(i), $4A_x(i)O_z(i)N_z(i + 1)$. Then, the ^{1}J$_{C\alpha C'}$ coupling is refocused and the ^{1}J$_{C\alpha N}$ and ^{2}J$_{C\alpha N}$ couplings evolve during a 2λ period. These evolutions give rise to two coherences ending in observable magnetization, $2A_y(i)N_z(i + 1)$ and $2A_y(i)N_z(i)$. The most interesting coherence is the second one, in which the nitrogen for which the coherence is anti-phase has changed. If the pulse applied in the middle of the 2λ period affects both ^{13}Cα and ^{13}Cβ spins, it is necessary to also consider the evolution of ^{1}J$_{C\alpha C\beta}$, which modulates the intensity of the above mentioned coherences. The coherences are subsequently relayed to ^{15}N spins, $2N_y(i + 1)A_z(i)$ and $2N_y(i)A_z(i)$. The ensuing period is used for labeling the ^{15}N chemical shift, while the J$_{C\alpha N}$ couplings are refocused and the ^{1}J$_{NC'}$ is defocused. These evolutions give rise to the $2N_y(i + 1)O_z(i)$ and $2N_y(i)O_z(i − 1)$ coherences. Finally, these coherences are converted into ^{13}C′ coherences, $2O_y(i)N_z(i + 1)$ and $2O_y(i − 1)N_z(i)$, and the ^{1}J$_{NC'}$ is refocused before ^{13}C′ detection. To improve resolution in the direct detected dimension, during this refocusing period, the IPAP method of virtual homodecoupling (Bermel et al. 2005;
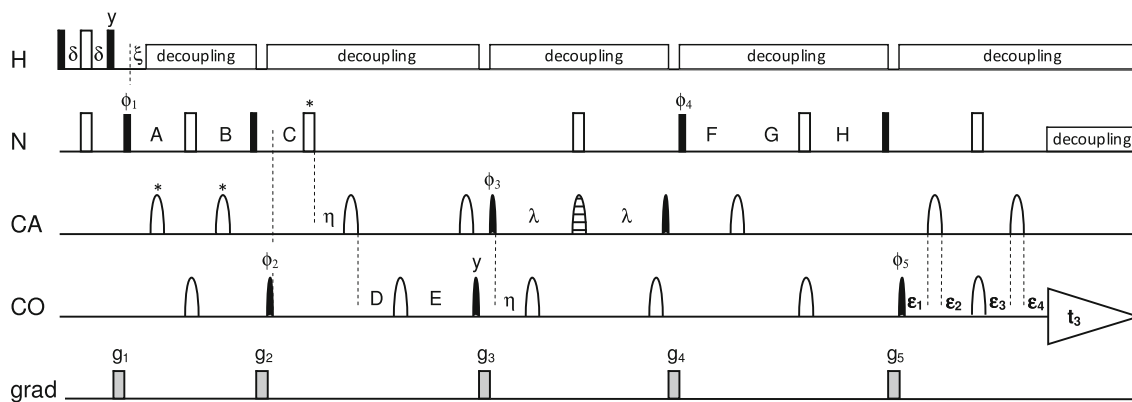
Fig. 2 Scheme of the hNCOcaNCO pulse sequence. All radiofrequency pulses are applied along the *x-axis* unless indicated. 90° and 180° *rectangular pulses* are represented by *filled* and *unfilled bars*, respectively. ^{13}C pulses have the *shape of gaussian cascades* Q5 (90°, *black filled shapes*) and Q3 (180°, *open shapes*) with durations of 307 and 192 ms at 800 MHz, respectively. The striped shape correspond to a Q3 *gaussian cascade* of 768 ms duration (at 800 MHz), applied at 54 ppm. Decoupling of ^{1}H and ^{15}N and were achieved with dipsi-2 (2.9 kHz) and garp (1.25 kHz), respectively. The delays employed for the 3D hNcocaNCO experiment are (differing delays of the 3D hnCOcaNCO experiment are given in *parentheses*): $\delta = 2.3$ ms; $\xi = 5.5$ ms; $\eta = 4.5$ ms; $\Delta_1 = 12.5$ ms; $\Delta_2 = 13.5$ ms; $\Delta_3 = 16.0$ ms; $\lambda = 25.0$ ms; $A = \Delta_1 + t_1/2$ (Δ_1); $B = \Delta_1 - t_1/2$ (Δ_1); $C = 0$ ($t_1/2$); $D = 0$ ms ($t_1/2 - \eta t_1/div$; $div = \max(2\eta, t_1^{max})$); $E = \eta$ ($\eta(1 - t_1/div)$); $F = \Delta_2 + t_2/2$; $G = \Delta_3 - \Delta_2$; $H = \Delta_3 - t_2/2$.

The delays of the IPAP element are: $\varepsilon_1(IP) = \Delta_3/2$, $\varepsilon_1(AP) = \eta$, $\varepsilon_2(IP) = \Delta_3/2$, $\varepsilon_2(AP) = \Delta_3 - \eta$, $\varepsilon_3(IP) = \Delta_3/2$, $\varepsilon_3(AP) = \Delta_3 - 4$ µs, $\varepsilon_4(IP) = \Delta_3/2$, $\varepsilon_4(AP) = 4$ µs. Pulsed field gradients g_1 to g_5 of sinusoidal shape are applied along the *z-axis* with a 1 ms length and amplitudes of 80, 60, 50, 30 and 19 % of the maximal intensity of about 50 G/cm. Phase cycle: $\phi_1 = 2(x)$, $2(-x)$; $\phi_2 = 2(x)$, $2(-x)$, $2(x)$, $4(-x)$, $2(x)$, $2(x)$, $2(x)$; $\phi_3 = y$, $-y$; $\phi_4 = 4(x)$, $4(-x)$; $\phi_5 = x(IP)$ or $-y(AP)$; $\phi(receiver) = 2(x, -x)$, $4(-x, x)$, $2(x, -x)$. Quadrature detection in t_1 and t_2 in the hNcocaNCO experiment is achieved by incrementing ϕ_1 and ϕ_4 according to States-TPPI and in the hnCOcaNCO experiment by incrementing ϕ_2 and ϕ_4. The ^{15}N pulse labeled with an *asterisk* is omitted in the hNcocaNCO experiment, and, likewise, the two ^{13}C pulses labeled with an *asterisk* can be omitted in the hnCOcaNCO experiment

Bertini et al. 2004) is introduced. In summary, there are two pathways that end in observable magnetization, ^{1}H(i + 1) → ^{15}N(i + 1) → ^{13}C′(i) → ^{13}Cα(i) → ^{15}N(i + 1) → ^{13}C′(i) and ^{1}H(i + 1) → ^{15}N(i + 1) → ^{13}C′(i) → ^{13}Cα(i) → ^{15}N(i) → ^{13}C′(i − 1), giving rise to an auto-correlated and a sequential peak. The sequential peak will have the coordinates ^{15}N(i + 1), ^{15}N(i), ^{13}C′(i − 1) in the hNcocaNCO experiment, and ^{13}C′(i), ^{15}N(i), ^{13}C′(i − 1) in the hnCOcaNCO experiment, thus correlating the chemical shifts of two consecutive ^{13}C′–^{15}N groups in the protein. The corresponding coordinates of the uninteresting auto-correlated peak are ^{15}N(i + 1), ^{15}N(i + 1), ^{13}C′(i) and ^{13}C′(i), ^{15}N(i + 1), ^{13}C′(i).

The intensity of the auto-correlated peak is proportional to

$$I_{auto} \propto \Gamma_1 \Gamma_2 \Gamma_3 \Gamma_{4a} \Gamma_{5a} \Gamma_6 \qquad (1)$$

and that of the sequential peak to

$$I_{seq} \propto \Gamma_1 \Gamma_2 \Gamma_3 \Gamma_{4a} \Gamma_{5a} \Gamma_6 \qquad (2)$$

where

$$\Gamma_1 = \sin(2\pi^1 J_{NH}\delta) \exp(-2\delta/T_2(^1H)) \qquad (3)$$

$$\Gamma_2 = \sin(2\pi^1 J_{NH}\xi) \sin(-2\pi^1 J_{NC'}\Delta_1) \exp(-2\Delta_1/T_2(^{15}N)) \qquad (4)$$

$$\Gamma_3 = \sin(2\pi^1 J_{CαC'}\eta) \exp(-2\eta/T_2(^{13}C')) \qquad (5)$$

$$\Gamma_{4a} = \sin(2\pi^1 J_{CαC'}\eta) \cos(2\pi^1 J_{CαN}\lambda) \cos(2\pi^2 J_{CαN}\lambda) \\ \cos(2\pi^1 J_{CαCβ}\lambda) \exp(-2\lambda/T_2(^{13}Cα)) \qquad (6)$$

$$\Gamma_{4b} = -\sin(2\pi^1 J_{CαC'}\eta) \sin(2\pi^1 J_{CαN}\lambda) \sin(2\pi^2 J_{CαN}\lambda) \\ \cos(2\pi^1 J_{CαCβ}\lambda) \exp(-2\lambda/T_2(^{13}Cα)) \qquad (7)$$

$$\Gamma_{5a} = \sin(2\pi^1 J_{NC'}\Delta_3) \cos(2\pi^1 J_{CαN}\Delta_2) \sin(2\pi^2 J_{CαN}\Delta_2) \\ \exp(-2\Delta_3/T_2(^{15}N)) \qquad (8)$$

$$\Gamma_{5b} = \sin(2\pi^1 J_{NC'}\Delta_3) \sin(2\pi^1 J_{CαN}\Delta_2) \cos(2\pi^2 J_{CαN}\Delta_2) \\ \exp(-2\Delta_3/T_2(^{15}N)) \qquad (9)$$

$$\Gamma_6 = \sin(2\pi^1 J_{NC'}\Delta_3) \exp(-2\Delta_3/T_2(^{13}C')) \qquad (10)$$

are the transfer functions during the six evolution periods, and δ, ξ, η, Δ_1, λ, Δ_2 and Δ_3 are the delays given in Fig. 2. To give maximum generality to the equations, the transfer functions corresponding to the evolution during the 2λ period, Γ_{4a} and Γ_{4b}, include the effect of the ^{1}J$_{CαCβ}$ coupling. The overall transfer efficiencies, Eqs. (1) and (2), are shown in Fig. 3, as a function of the important delay λ. For this representation values of 93, 15, 53, 10.6, 7.5 and 35 Hz were used for the ^{1}J$_{NH}$, ^{1}J$_{NC'}$, ^{1}J$_{CαC'}$, ^{1}J$_{CαN}$, ^{2}J$_{CαN}$ and ^{1}J$_{CαCβ}$ couplings, respectively, and relaxation times of 100,
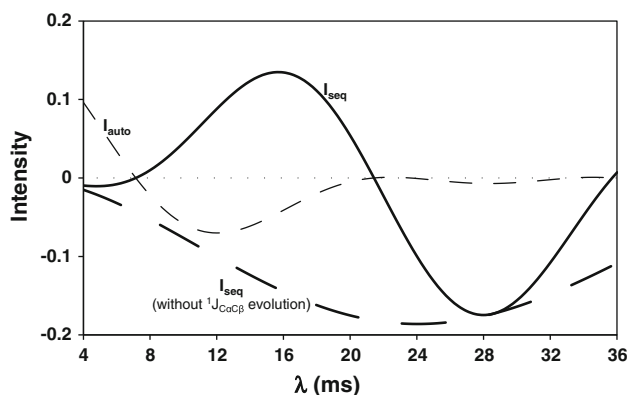
Fig. 3 Representation of the overall transfer functions for the sequential (I_{seq}) and auto-correlated (I_{auto}) peaks as a function of the delay λ. For the sequential peak the transfer function omitting the effect of the $^1J_{C\alpha C\beta}$ is also represented. Delays δ, ξ, η, Δ_1, Δ_2 and Δ_3 used in the calculation are given in the caption of Fig. 2. Other parameters, coupling constant and relaxation time values, are given in the text

200, 100 and 200 ms for ^1H, ^{15}N, $^{13}C\alpha$ and $^{13}C'$ spins. These relaxation times correspond to the values expected for an IDP of about 100 residues (Mäntylahti et al. 2011). The transfer efficiency of the sequential peak is also represented without considering the effect of the $^1J_{C\alpha C\beta}$ coupling. As is evident in the figure, the overall transfer function of the sequential peak has maxima (in magnitude) at the values $\lambda \sim 15$ ms and $\lambda \sim 28$ ms. A value of $\lambda \sim 15$ ms leads to the additional appearance of relatively intense auto-correlated peaks with phase opposite to that of the sequential peaks. Also, peaks for which the magnetization passes through $^{13}C\alpha$ of glycines, therefore lacking the $^1J_{C\alpha C\beta}$ coupling modulation, appear with opposite sign to the other peaks of the same type, thereby allowing identification of glycine residues. The maximum of I_{seq} at $\lambda \sim 28$ ms is larger than that at $\lambda \sim 15$ ms. Tuning λ at a slightly smaller value, $\lambda = 25$ ms, results in near maximal sequential peak intensity and a practically absent auto-correlated peak. Therefore, the number of observed peaks is reduced, diminishing the spectral crowding and avoiding the risk of mutual cancellation of the sequential and auto-correlated peaks in case of near degeneracy in ^{15}N or $^{13}C'$ chemical shifts of two consecutive residues. The intensity of the peaks, except those in which the transfer pathway goes through $^{13}C\alpha$ of glycines, serines or threonines, can be increased by using a 180° selective pulse affecting only the $^{13}C\alpha$ chemical shift range in the middle of the 2λ period to remove the dependence on the $^1J_{C\alpha C\beta}$ coupling. Consequently, in our experiments we have chosen a value $\lambda = 25$ ms and used a highly selective 180° pulse in the middle of the 2λ period. These conditions optimize I_{seq} and suppress the auto-correlated peak, which would otherwise lead to unnecessary crowding of the spectra with redundant information.

Correlation of consecutive C′–N groups can also be obtained with the (H)CANCO experiment (Bermel et al. 2009b). In this experiment two consecutive C′–N groups correlate with the $^{13}C\alpha$ located between them. However, the poor dispersion of $^{13}C\alpha$ in the case of IDPs makes the (H)CANCO experiment less appropriate for their assignment than ours. Theoretical calculation for the (H)CANCO experiment, using the parameters given above, gives a transfer efficiency of 0.186 for the $^{13}C\alpha(i)$, $^{15}N(i)$, $^{13}C'(i-1)$ peak and 0.064 for the $^{13}C\alpha(i)$, $^{15}N(i+1)$, $^{13}C'(i)$ peak. Therefore, the new experiments, with a transfer efficiency of 0.184, also show a sensitivity advantage in the case of IDPs. Nevertheless, as the relaxation rates grow the sensitivity advantage is getting lost, so that the (H)CANCO experiment becomes better suited for the study of folded proteins of 100 residues or more.

In our approach, the sequential connectivity is obtained using the 3D hNcocaNCO and hnCOcaNCO spectra in concert with the 2D CON spectrum. The process is illustrated in Fig. 4, where the assignment of a stretch of residues of Nupr1 is presented. The assignment starts from a

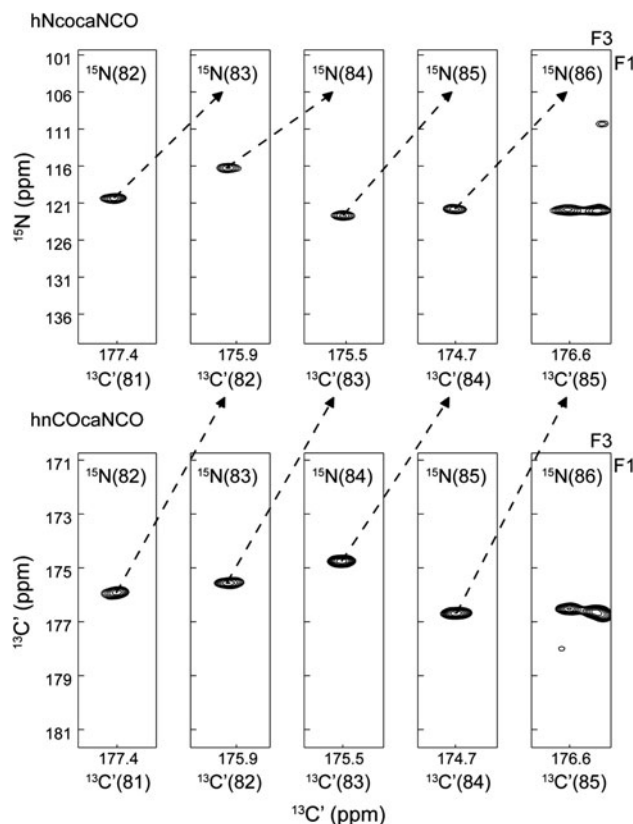**Fig. 4** Illustration of the method for assigning the protein $^{13}C'$ and ^{15}N resonances. Starting from a ^{15}N–$^{13}C'$ correlation, the ^{15}N and $^{13}C'$ chemical shift values of the succeeding residue are obtained from the sequential cross peaks in the hNcocaNCO and hnCOcaNCO spectra, respectively. Assignment is exemplified using the residues Q82–R86 of Nupr1

$^{15}N(i)$–$^{13}C'(i-1)$ correlation peak in the 2D CON spectrum. By selecting the $^{15}N(i)$ plane in the hNcocaNCO spectrum, the cross peak $^{15}N(i+1)$–$^{15}N(i)$–$^{13}C'(i-1)$, appearing in a F1–F3 strip centered at $^{13}C'(i-1)$, provides the $^{15}N(i+1)$ chemical shift. Similarly, the cross peak $^{13}C'(i)$–$^{15}N(i)$–$^{13}C'(i-1)$, appearing in the $^{15}N(i)$ plane of the hnCOcaNCO spectrum, provides the $^{13}C'(i)$ chemical shift. Together, these two chemical shifts identify the CON cross peak of the following amide bond in the protein sequence, and can be used as starting point to locate the next ^{15}N–$^{13}C'$ correlation. This 'backbone CON walk' has been automated in NMRView. Marking a cross peak in the 2D CON spectrum brings out the corresponding plane strips in the hnCOcaNCO and hNcocaNCO spectra, centered at the coordinates of the marked CON peak. The peaks observed in these strips provide in a very simple visual manner the coordinates of the following CON cross peak, as is shown in Fig. 5. The process is repeated iteratively, defining a stretch of consecutive residues, until it

terminates either at a residue preceding proline (we number the ^{15}N–$^{13}C'$ correlations with the residue number of the nitrogen) or at the C-terminal residue. An unassigned ^{15}N–$^{13}C'$ correlation in the CON spectrum is then selected to start the assembly of a new stretch of consecutive residues. The process is repeated until all ^{15}N–$^{13}C'$ correlations have been used. Since stretches of connected residues should start at residue 2 or at a proline, it is a good practice to select ^{15}N–$^{13}C'$ correlations of prolines, that are easily distinguished by their singular ^{15}N chemical shift, as starting points.

For protein Nupr1 6 stretches starting at proline were identified with lengths of 40, 24, 5, 3, 2, and 2 residues. The two remaining prolines do not show any correlation and correspond to the first residues of the two PP groups present in the protein. Additionally 2 stretches starting at non-Pro residues, with lengths of 6 and 4 residues, were established. Connections of 4 non-proline residues having practically the same $^{13}C'$ chemical shift and very similar ^{15}N chemical shift could not be discriminated. Once the stretches of consecutive residues have been obtained, the next step in the assignment process consists in mapping the stretches to the protein sequence. In general, the process will be problematic if information from the hNcocaNCO and hnCOcaNCO experiments is merely used, since only prolines can be identified and employed as anchor points. Nevertheless, in our case the 4 longer stretches starting at proline can be mapped to the protein sequence, resulting in the assignment of 72 residues (78 % of the protein), as seen in Fig. 6. It is also possible to propose likely assignments for the two stretches not starting at proline. The fragment of 6 residues lacks a connection to a following residue and has a connection to a preceding residue with one of the 4 residues that could not be resolved in the 3D spectra. Therefore, most likely ends in a residue preceding proline. The fragment of 4 residues lacks a connection to a preceding residue and has a following connection with one of the 4 residues that could not be resolved in the 3D spectra. Thus, most likely starts at the second residue of the protein. Incorporating these proposals increases the assignment percentage to 89 %. In less favorable cases, the assignment requires knowledge of the amino acid residue type of some members of each stretch. For this purpose, performing the experiments with $\lambda = 15$ ms can be helpful, since glycines and residues following glycine are identified. Furthermore, working with this delay and tuning the offset and bandwidth of the 180° pulse in the middle of the 2λ period, it is

Fig. 5 A diagram showing the 'backbone CON walk' as implemented in NMRView. Starting from a ^{15}N–$^{13}C'$ peak of the 2D CON spectrum, the ^{15}N–$^{13}C'$ correlation peak corresponding to the succeeding amide moiety is obtained from the coordinates of the sequential cross peaks in the hNcocaNCO and hnCOcaNCO spectra

Fig. 6 Mapping of the stretches starting at proline, obtained in the analysis of the hNcocaNCO and hnCOcaNCO spectra, to the sequence of Nupr1. ^{15}N–$^{13}C'$ correlations identified as coming from Pro residues, corresponding to the first member of a stretch, appear in *red* and *underlined*. Additionally, the most probable assignment of the two stretches starting at non-Pro residues is given

possible to identify alanine (Chatterjee et al. 2006) or serine/threonine residues (Chugh et al. 2008). However, we believe that it is more efficient to perform the experiments with $\lambda = 25$ ms, which optimizes the intensity of the sequential peak, eliminates the risk of signal cancelations and reduce the spectral crowding, and get information about the amino acid residue types from auxiliary experiments. The identification of two or three types should suffice to perform the mapping process successfully. In this regard, the recently published CAS–NMR experiments (Bermel et al. 2012b), that give ^{15}N–$^{13}C'$ correlations arising only from a particular amino acid residue type, can be of great utility.

In summary, we have proposed a pair of 3D correlation experiments for sequential assignment of highly flexible protein systems such as IDPs. The performance of this novel assignment protocol has been demonstrated with Nupr1, an intrinsically disordered protein of 93 residues, for which a nearly complete assignment was obtained. The pulse sequences can be also applied to facilitate the assignment of small to medium-size globular proteins with highly crowded ^{1}H–^{15}N HSQC spectra, like α-helical proteins. Main limitation of the experiments is the absence of the N_{i+1}–N_i–C'_{i-1} and C'_i–N_i–C'_{i-1} correlations when $i + 1$ is a proline residue.

References

Bermel W, Bertini I, Duma L, Emsley L, Felli IC, Pierattelli R, Vasos PR (2005) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew Chem Int Ed 44:3089–3092

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009a) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bermel W, Bertini I, Felli IC, Pierattelli R (2009b) Speeding up ^{13}C direct detection biomolecular NMR spectroscopy. J Am Chem Soc 131:15339–15345

Bermel W, Bertini I, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, Stanek J (2012a) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Bermel W, Bertini I, Chill J, Felli IC, Haba N, Kumar V, Pieratelli R (2012b) Exclusively heteronuclear 13C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDP). Chem Biochem 13:2425–2432

Bertini I, Felli IC, Kümmerle R, Luchinat C, Pieratelli R (2004) ^{13}C–^{13}C NOESY: a constructive use of ^{13}C–^{13}C spin-diffusion. J Biomol NMR 30:245–251

Chatterjee A, Kumar A, Hosur RV (2006) Alanine check points in HNN and HN(C)N spectra. J Magn Reson 181:21–28

Chugh J, Kumar D, Hosur RV (2008) Tuning the HNN experiment: generation of serine-threonine check points. J Biomol NMR 40:145–152

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Curr Opin Struct Biol 18:756–764

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 (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622

Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 19:23–30

Fink AL (2005) Natively unfolded proteins. Curr Opin Struct Biol 15:35–41

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

Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614

Kumar D, Reddy JG, Hosur RV (2010) hnCOcaNH and hncoCANH pulse sequences for rapid and unambiguous backbone assignment in (^{13}C, ^{15}N) labeled proteins. J Magn Reson 206:134–138

Mäntylahti S, Aito O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. J Biomol NMR 47:171–181

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Motáčková V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Žídek L, Sanderová H, Krásny L, Kozminski W, Sklenár V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution enhanced 5D experiments. J Biomol NMR 48:169–177

Nováček J, Zawadzka-Kazimierczuk A, Papouskova V, Žídek L, Sandervá H, Krásny L, Kozminski W, Sklenár V (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Nováček J, Haba NY, Chill JH, Žídek L, Sklenár V (2012) 4D Non-uniformly sampled HCBCACON and 1 J(NCα)-selective-HCB-CANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

Panchal SC, Bavesh NS, Hosur RV (2001) 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. J Biomol NMR 20:135–147

Pantoja-Uceda D, Santoro J (2009) Aliasing in reduced dimensionality spectra: (3,2)D HNHA and (4,2)D HN(COCA)NH as examples. J Biomol NMR 45:351–356

Permi P, Annila A (2004) Coherence transfer in proteins. Prog Nucl Magn Reson Spectrosc 44:97–137

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158

Sun ZJ, 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

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533

Tompa P (2011) Unstructural biology coming of age. Curr Opin Struct Biol 21:419–425

Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37:509–516

Yao J, Dyson HJ, Wright PE (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partially folded proteins. FEBS Lett 419:285–289

Zhang O, Forman-Kay JD, Shortle D, Kay LE (1997) Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. J Biomol NMR 9:181–200