J Biomol NMR (2014) 59:43–50
DOI 10.1007/s10858-014-9827-1

ARTICLE

New ^{13}C-detected experiments for the assignment of intrinsically disordered proteins

David Pantoja-Uceda · Jorge Santoro

Received: 10 February 2014 / Accepted: 27 March 2014 / Published online: 4 April 2014

Abstract NMR assignment of intrinsically disordered proteins (IDPs) by conventional HN-detected methods is hampered by the small dispersion of the amide protons chemical shifts and exchange broadening of amide proton signals. Therefore several alternative assignment strategies have been proposed in the last years. Attempting to seize that dispersion of $^{13}C'$ and ^{15}N chemical shifts holds even in IDPs, we recently proposed two ^{13}C-detected experiments to directly correlate the chemical shifts of two consecutive $^{13}C'$–^{15}N groups in proteins, i.e. without mediation of other nuclei. Main drawback of these experiments is the interruption of the connection at prolines. Here we present new ^{13}C-detected experiments to correlate consecutive $^{13}C'$–^{15}N groups in IDPs, hacacoNcaNCO and hacaCOncaNCO, that overcome this limitation. Moreover, the experiments provide recognition of glycine residues, thereby facilitating the assignment process.

Keywords Intrinsically disordered proteins · Backbone assignment · ^{13}C-direct detection NMR · Multidimensional NMR experiment

Introduction

Many proteins or protein domains (intrinsically disordered proteins, IDPs) lack a well defined secondary and tertiary structure under functional conditions (Dunker et al. 2008; Fink 2005; Tompa 2002, 2011, 2012; Uversky et al. 2000). NMR spectroscopy is probably the most powerful technique for characterizing IDPs (Dyson and Wright 2001, 2004; Eliezer 2009; Fink 2005; Jensen et al. 2009) as it provides information at atomic resolution on both local and long-range conformational behavior on timescales varying over many orders of magnitude. Sequential assignment is a prerequisite for this NMR-based characterization. Conformational averaging of IDPs results in poorly dispersed amide ^{1}H chemical shifts and in clustering of the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts around the random coil value for each amino acid residue type, difficulting the sequential assignment of IDPs with the methods established for globular proteins (Permi and Annila 2004; Sattler et al. 1999). Therefore, several alternative assignment strategies have been proposed in the last years (Bermel et al. 2005, 2009a, 2012a; Kumar and Hosur 2011; Hiller et al. 2007; Mäntylahti et al. 2010, 2011; Motackova et al. 2010; Novacek et al. 2012; Sahu et al. 2014; Solyom et al. 2013; Wen et al. 2011). Many of these strategies make use of the long relaxation times that IDPs present as a result of the high flexibility of the disordered polypeptide chain, which allows designing experiments involving multiple coherence transfer steps. However, most of these approaches continue to use the poorly dispersed $^{13}C\alpha$, $^{13}C\beta$, or amide proton chemical shifts for the sequential assignment.

Trying to exploit that the ^{15}N and $^{13}C'$ signals remain well dispersed in IDPs (Dyson and Wright 2001; Yao et al. 1997; Zhang et al. 1997) for the sequence specific assignment, we recently have proposed (Pantoja-Uceda and Santoro 2013a) two ^{13}C-direct detected experiments, hNcocaNCO and hnCOcaNCO, that allow to directly correlate two consecutive C'–N groups in proteins, i.e. without mediation of other nuclei. Main shortcoming of these experiments is the interruption of the connection at prolines, since the experiments use the amide protons as the starting point of the magnetization transfer pathway. Shortly thereafter, Bermel et al. (2013)

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

have proposed alternative experiments for correlating successive C′–N groups, 5D h^{N-flip}nCONCACON, 5D hcaCONCACON and 5D hCACONcaCON. The experiments use a bidirectional coherence transfer and generate two kinds of peaks: auto-correlated peaks [$^{13}C'(i-1)$–$^{15}N(i)$–$^{13}C\alpha(i-1)$–$^{15}N(i)$–$^{13}C'(i-1)$ in the first two experiments and $^{13}C\alpha(i-1)$–$^{13}C'(i-1)$–$^{15}N(i)$–$^{15}N(i)$–$^{13}C'(i-1)$ in the third] and sequential peaks [$^{13}C'(i-2)$–$^{15}N(i-1)$–$^{13}C\alpha(i-1)$–$^{15}N(i)$–$^{13}C'(i-1)$ in the first two experiments and $^{13}C\alpha(i-2)$–$^{13}C'(i-2)$–$^{15}N(i-1)$–$^{15}N(i)$–$^{13}C'(i-1)$ in the third], which provide the information for the sequential assignment. Although the experiments involve the chemical shift labeling of the poorly dispersed $^{13}C\alpha$, it is only used to minimize the occurrence of overlap, and 4D versions of the experiments can also be used to establish the sequential connectivity. As in the two last experiments the amide protons are not included in the magnetization transfer, the sequential assignment is not interrupted by the presence of proline residues in the protein sequence, and are therefore better suited to the sequential assignment of IDPs, as many of them are composed of proline-rich polypeptide segments (Marsh and Forman-Kay 2010; Radivojac et al. 2007; Romero et al. 2001). Here we present a new pulse sequence for correlating consecutive C′–N groups that, maintaining this positive feature, offer some advantages over these latter. First, the novel pulse sequence has fewer coherence transfer steps, resulting in an increased sensitivity. Also, our pulse sequence uses a unidirectional coherence transfer pathway. Therefore, as uninformative auto-correlated peaks do not appear in the spectra, the number of observed peaks is reduced, diminishing the spectral crowding.

Materials and methods

NMR experiments were performed at 18.8 T on a Bruker AV spectrometer operating at 800.1 MHz ^{1}H, 201.18 MHz ^{13}C and 81.07 MHz ^{15}N frequencies, equipped with a cryogenically cooled triple-resonance (^{1}H, ^{13}C, ^{15}N) TCI probehead and pulsed z-field gradients. A sample of 0.6 mM uniformly ^{13}C, ^{15}N labeled Nupr1 in 10 mM acetate buffer at pH 4.5 was used to test the experiments. All experiments were acquired at 298 K. Parameters specific to the pulse sequence are reported in the caption of Fig. 2. For ^{13}C band-selective π/2 and π pulses Q5 (or time reversed Q5) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 307.2 and 192 μs, respectively, were used, except for the adiabatic π pulse used to invert both C′ and Cα (smoothed Chirp (Böhlen and Bodenhausen 1993), 500 μs, 20 % smoothing, 80 kHz sweep width, 12.5 kHz rf field strength). The ^{13}C band-selective pulses on Cα and C′ were given at the center of corresponding region and the adiabatic inversion pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz-16 (Shaka et al. 1983) (4.5 kHz) and garp-4 (Shaka et al. 1985) (0.83 kHz) sequences, respectively. All pulsed field gradients employed had 1 ms of duration and a sine shape. The States-TPPI method was applied in all indirect dimensions to achieve quadrature detection. All experiments employ the IPAP approach (Bermel et al. 2005; Bertini et al. 2004) to remove the splitting caused by the Cα-C′ couplings in the direct dimension. The in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded (Bermel et al. 2006, 2008).

3D spectra were recorded with spectral widths of 11 and 35 ppm centered at 173.5 and 121.5 ppm for $^{13}C'$ and ^{15}N, respectively. 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. With these values resolution in the indirect dimensions were 1.25 ppm for ^{15}N and 0.39 ppm for $^{13}C'$. These values were suitable to assign the spectrum of Nupr1, but can be insufficient in other cases. The total acquisition time for each 3D spectrum was 20 h. The programs NMRPipe (Delaglio et al. 1995) and NMRView (Johnson and Blevins 1994) were used for spectral processing and data analysis, respectively.

Results and discussion

Description of the pulse sequence

Main drawback of the HNCOCANCO pulse sequence (Pantoja-Uceda and Santoro 2013a) lies in the absence of correlations in which the magnetization starts at prolines. A naive way to overcome this limitation consists in starting from $^{1}H\alpha(i)$ coherence, transfer the coherence to $^{15}N(i+1)$ via $^{13}C\alpha(i)$ and $^{13}C'(i)$, and follow from there the original sequence. Therefore, the coherence should follow the pathway $^{1}H\alpha(i)$–$^{13}C\alpha(i)$–$^{13}C'(i)$–$^{15}N(i+1)$–$^{13}C'(i)$–$^{13}C\alpha(i)$–$^{15}N(i)$–$^{13}C'(i-1)$. This naive sequence requires several magnetization transfer steps more than the original one, so that a significant loss in sensitivity is expected. However, a careful analysis of this sequence allows its optimization, so that its sensitivity approaches that of the original sequence. Indeed, in the original sequence the passage through the $^{13}C'(i)$ serves two purposes: labeling the signal with the $^{13}C'$ chemical shift and allowing the unidirectional transfer from $^{15}N(i+1)$ to $^{13}C\alpha(i)$. In the modified sequence the labeling can be obtained in the first passage over $^{13}C'(i)$, that also allows the unidirectional transfer from $^{13}C\alpha(i)$ to $^{15}N(i+1)$. Thus, the second passage through $^{13}C'(i)$ can be omitted, if a way to make directly the back transfer from $^{15}N(i+1)$ to $^{13}C\alpha(i)$ is found. This unidirectional transfer is easily

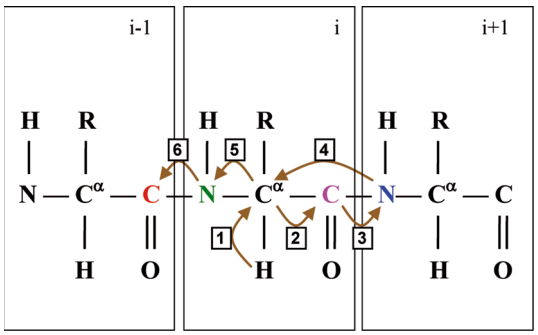

Fig. 1 Schematic representation of the magnetization transfer pathway implemented in the hacaCONcaNCO experiment. The *colored* nuclei are used to correlate residues i and $i + 1$. *Arrows* indicate the magnetization transfer pathway

achieved if the $^{1}J_{C\alpha C'}$ coupling is not refocused in the period in which the coherence resides in $^{13}C'(i)$, so that the $^{15}N(i + 1)$ coherence produced from it is antiphase with respect to both $^{13}C'(i)$ and $^{13}C\alpha(i)$ i.e. of the form $4N_y(i + 1)C'_z(i)C\alpha_z(i)$. Therefore the optimized new sequence must perform the coherence transfers $^{1}H\alpha(i)$–$^{13}C\alpha(i)$–$^{13}C'(i)$–$^{15}N(i + 1)$–$^{13}C\alpha(i)$–$^{15}N(i)$–$^{13}C'(i - 1)$, as is schematically shown in Fig. 1. Although this sequence still has one coherence transfer step more than the original one, as we will see in the description that follows, its sensitivity approaches that of the original sequence.

The pulse sequence that performs the above described coherence transfer steps is given in Fig. 2. The pulse sequence starts with an INEPT coherence transfer from the $^{1}H\alpha$ spin to its directly attached $^{13}C\alpha$ spin, after which the relevant density operator is

$$2H\alpha_z(i)C\alpha_y(i) \tag{1}$$

During the following delay, set to $1/(2J_{C\alpha C\beta})$, the $^{1}J_{C\alpha H\alpha}$ is refocused and the $^{1}J_{C\alpha C'}$ defocused, producing the coherence

$$2C\alpha_y(i)C'_z(i)\Gamma_1 \tag{2}$$

where $\Gamma_1 = n\ \sin(2\pi^1J_{C\alpha H\alpha}\tau_2)\ \cos^{n-1}(2\pi^1J_{C\alpha H\alpha}\tau_2)\ \sin(2\pi^1J_{C\alpha C'}\eta)\ \cos^{2-n}(2\pi^1J_{C\alpha C\beta}\Delta_1)\ \exp[-2\Delta_1/T_2(C\alpha)]$ (n is the proton multiplicity of the α-carbon, i.e. 2 for glycine and 1 otherwise). Then, a pair of 90° pulses applied to $^{13}C\alpha$ and $^{13}C'$, transfer the coherence to the $^{13}C'(i)$ spin. The ensuing period serves to label the coherence with the chemical shift of $^{13}C'(i)$ and to defocus the $^{1}J_{C'N}$ coupling giving the density operator

$$4C\alpha_z(i)C'_y(i)N_z(i + 1)\ \cos(\omega_{C'(i)}t_1)\Gamma_1\Gamma_2 \tag{3}$$

where $\Gamma_2 = \sin(2\pi^1J_{C'N}\Delta_2)\ \exp[-2\Delta_2/T_2(C')]$. Next, the coherence is transferred to ^{15}N transverse magnetization, labeled with the chemical shift of the $^{15}N(i + 1)$ spin in a real time mode and converted back into a $^{13}C\alpha(i)$ coherence, defined by the density operator

$$4C\alpha_y(i)C'_z(i)N_z(i + 1)\ \cos(\omega_{C'(i)}t_1)\ \cos(\omega_{N(i+1)}t_2)\ \exp[-t_2/T_2(N)]\Gamma_1\Gamma_2 \tag{4}$$

Next, the $^{13}C\alpha(i)$ magnetization evolves the $^{1}J_{C\alpha N}$ and $^{2}J_{C\alpha N}$ couplings and refocus the $^{1}J_{C\alpha C'}$ coupling. The density operator terms than can end into observable magentization are

$$[2C\alpha_x(i)N_z(i + 1)\Gamma_3 + 2C\alpha_x(i)N_z(i)\Gamma_4]\ \cos(\omega_{C'(i)}t_1)\ \cos(\omega_{N(i+1)}t_2)\ \exp[-t_2/T_2(N)]\Gamma_1\Gamma_2 \tag{5}$$

where $\Gamma_3 = \sin(2\pi^1J_{C\alpha C'}\eta)\ \cos(2\pi^1J_{C\alpha N}\Delta_3)\ \cos(2\pi^2J_{C\alpha N}\Delta_3)\ \cos^{2-n}(2\pi^1J_{C\alpha C\beta}\Delta_3)\ \exp[-2\Delta_3/T_2(C\alpha)]$ and $\Gamma_4 = -\sin(2\pi^1J_{C\alpha C'}\eta)\ \sin(2\pi^1J_{C\alpha N}\Delta_3)\ \sin(2\pi^2J_{C\alpha N}\Delta_3)\ \cos^{2-n}(2\pi^1J_{C\alpha C\beta}\Delta_3)\ \exp[-2\Delta_3/T_2(C\alpha)]$. These coherences are converted into ^{15}N magnetization that is refocused with respect to the $^{1}J_{C\alpha N}$ or $^{2}J_{C\alpha N}$ coupling and dephased with respect to the $^{1}J_{C'N}$ coupling. During this period the coherences are also labeled with the chemical shift of the ^{15}N spin, and therefore give rise to the density operator

$$2N_y(i + 1)C'_z(i)\ \cos(\omega_{C'(i)}t_1)\ \cos(\omega_{N(i+1)}t_2)\ \cos(\omega_{N(i+1)}t_3)\ \exp[-t_2/T_2(N)]\Gamma_1\Gamma_2\Gamma_3\Gamma_5 + 2N_y(i)C'_z(i - 1)\ \cos(\omega_{C'(i)}t_1)\ \cos(\omega_{N(i+1)}t_2)\ \cos(\omega_{N(i)}t_3)\ \exp[-t_2/T_2(N)]\Gamma_1\Gamma_2\Gamma_4\Gamma_6 \tag{6}$$

where $\Gamma_5 = \cos(2\pi^1J_{C\alpha N}\zeta)\ \sin(2\pi^2J_{C\alpha N}\zeta)\ \sin(2\pi^1J_{C'N}\Delta_4)\ \exp[-2\Delta_4/T_2(N)]$ and $\Gamma_6 = \sin(2\pi^1J_{C\alpha N}\zeta)\ \cos(2\pi^2J_{C\alpha N}\zeta)\ \sin(2\pi^1J_{C'N}\Delta_4)\ \exp[-2\Delta_4/T_2(N)]$. Finally, these coherences are transferred to $^{13}C'$ magnetization and the $^{1}J_{C'N}$ coupling is refocused before $^{13}C'$ detection. The observable magnetization is then

$$C'_x(i)\ \cos(\omega_{C'(i)}t_1)\cos(\omega_{N(i+1)}t_2)\ \cos(\omega_{N(i+1)}t_3)\ \exp[-t_2/T_2(N)]\Gamma_1\Gamma_2\Gamma_3\Gamma_5\Gamma_7 + C'_x(i - 1)\ \cos(\omega_{C'(i)}t_1)\ \cos(\omega_{N(i+1)}t_2)\ \cos(\omega_{N(i)}t_3)\exp[-t_2/T_2(N)]\Gamma_1\Gamma_2\Gamma_4\Gamma_6\Gamma_7 \tag{7}$$

where $\Gamma_7 = \sin(2\pi^1J_{C'N}\Delta_5)\ \exp[-2\Delta_5/T_2(C')]$. To improve resolution in the direct detected dimension, during the final refocusing period the IPAP method of virtual homodecoupling (Bermel et al. 2005; Bertini et al. 2004) is introduced. In summary, there are two pathways that end in observable magnetization, $^{1}H\alpha(i) \rightarrow {}^{13}C\alpha(i) \rightarrow {}^{13}C'(i) \rightarrow {}^{15}N(i + 1) \rightarrow {}^{13}C\alpha(i) \rightarrow {}^{15}N(i + 1) \rightarrow {}^{13}C'(i)$ and $^{1}H\alpha(i) \rightarrow {}^{13}C\alpha(i) \rightarrow {}^{13}C'(i) \rightarrow {}^{15}N(i + 1) \rightarrow {}^{13}C\alpha(i) \rightarrow {}^{15}N(i) \rightarrow {}^{13}C'(i–1)$, giving rise to an auto-correlated and a sequential peak. The intensity of the auto-correlated peak, omitting the small intensity loss associated with the INEPT transfer from $^{1}H\alpha$ to $^{13}C\alpha$, is proportional to

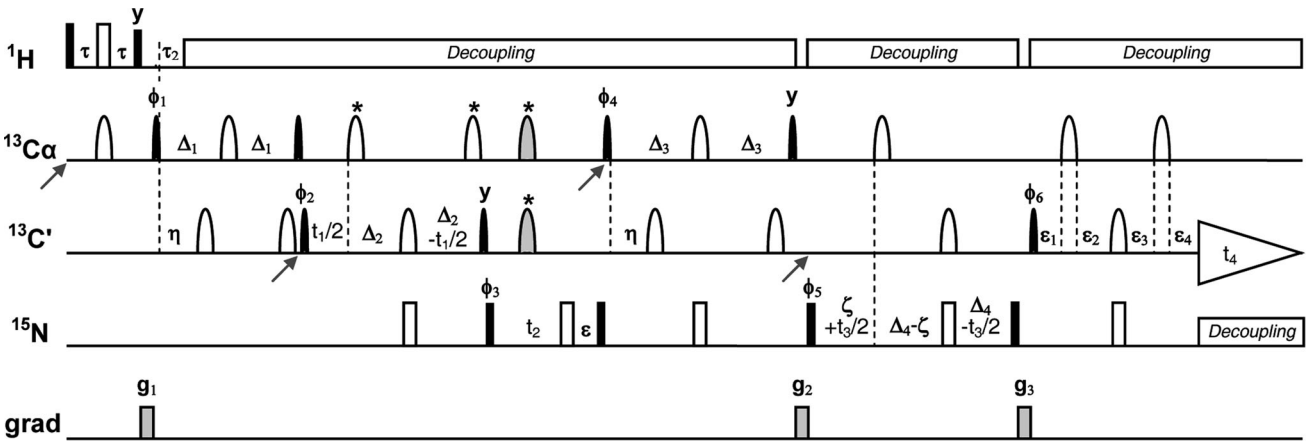

Fig. 2 Scheme of the hacaCONcaNCO pulse sequence. The *arrows* indicate the switching of the ^{13}C carrier frequency. All radiofrequency pulses are applied along the *x-axis* unless indicated. 90 and 180° rectangular pulses are represented by *filled* and *unfilled bars*, respectively. ^{13}C pulses 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 simultaneous $^{13}C\alpha$ and $^{13}C'$ pulses indicated in *grey* correspond to an adiabatic inversion pulse over the $^{13}C'$ and $^{13}C\alpha$ regions, Chirp, 500 μs, 25 % smoothing, 80 kHz sweep, 10.0 kHz strength. Decoupling of ^{1}H and ^{15}N and were achieved with waltz-16 (4.5 kHz) and garp-4 (0.83 kHz), respectively. The delays employed are: $\tau = 1.8$ ms; $\tau_2 = 2.4$ ms; $\eta = 4.5$ ms; $\Delta_1 = 13.5$ ms; $\Delta_2 = 15.0$ ms; $\Delta_3 = 28.0$ ms; $\zeta = 14.0$ ms; $\Delta_4 = 15.0$ ms; $\varepsilon = t_2(0) + p$, where p is the duration of the adiabatic ^{13}C pulse. The delays of the IPAP element are: $\varepsilon_1(IP) = \Delta_5/2$, $\varepsilon_1(AP) = \eta$, $\varepsilon_2(IP) = \Delta_5/2$, $\varepsilon_2(AP) = \Delta_5 - \eta$, $\varepsilon_3(IP) = \Delta_5/2$, $\varepsilon_3(AP) = \Delta_5 - 4$ μs, $\varepsilon_4(IP) = \Delta_5/2$, $\varepsilon_4(AP) = 4$ μs, with $\Delta_5 = 15.0$ ms. Pulsed field gradients g_1 to g_3 of sinusoidal shape are applied along the *z-axis* with a 1 ms length and amplitudes of 50, 30 and 19 % of the maximal intensity of about 50 G/cm. Phase cycle: $\phi_1 = 16(x), 16(-x)$; $\phi_2 = 8(x), 8(-x)$; $\phi_3 = 2(x), 2(-x)$; $\phi_4 = 4(x), 4(-x)$; $\phi_5 = x, -x$; $\phi_6 = x(IP)$ or $-y(AP)$; ϕ(receiver) = x, 2(−x), x, −x, 2(x), 2(−x), 2(x), −x, x, 2(−x), x, −x, 2(x), −x, x, 2(−x), 2(x), 2(−x), x, −x, 2(x), −x. Good results are obtained using only the first 4 steps of the phase cycle. The 3D hacaCOncaNCO experiment is performed by setting t_1 to 0 and omitting the $^{13}C\alpha$ 180° pulses labeled with an *asterisk*. Quadrature detection in the indirect dimensions is achieved by incrementing phases ϕ_2 and ϕ_5 in a States-TPPI manner. To perform the 3D hacacoNcaNCO experiment t_2 is fixed to 0 and the adiabatic inversion pulse labeled with an *asterisk* is omitted. In this case, quadrature detection in the indirect dimensions is achieved by incrementing phases ϕ_3 and ϕ_5 in a States-TPPI manner

$$I_{auto} \propto \Gamma_1\Gamma_2\Gamma_3\Gamma_5\Gamma_7 \quad (8)$$

and that of the sequential peak to

$$I_{seq} \propto \Gamma_1\Gamma_2\Gamma_4\Gamma_6\Gamma_7 \quad (9)$$

Overall transfer efficiencies of 0.006 for the auto-correlated pathway and 0.133 for the sequential pathway have been calculated using values of 140, 15, 53, 10.6, 7.5 and 35 Hz for the $^{1}J_{C\alpha H\alpha}$, $^{1}J_{NC'}$, $^{1}J_{C\alpha C'}$, $^{1}J_{C\alpha N}$, $^{2}J_{C\alpha N}$ and $^{1}J_{C\alpha C\beta}$ couplings, respectively, 200, 100 and 200 ms for the transverse relaxation times of ^{15}N, $^{13}C\alpha$ and $^{13}C'$ spins (Mäntylahti et al. 2011) and the delays given in the caption of Fig. 2. The value used for Δ_3, 28 ms, optimizes I_{seq} and practically suppress the auto-correlation peak, which otherwise would cause unnecessary crowding of the spectra with redundant information. Moreover, the value employed for Δ_1, 13.5 ms, inverts the sign of the peaks when residue *i* is glycine (Feng et al. 1996). This feature allows to easily identify glycine residues, helping in the assignment process. Therefore, the new pulse sequence has advantages over our previous pulse sequence (Pantoja-Uceda and Santoro 2013a), namely the presence of all $^{13}C'$–^{15}N correlations and the recognition of glycine residues. The price to pay is a reduction in the transfer efficiency, 0.133 versus 0.186. The sensitivity of the new experiment could be increased if delays $\Delta_1 = 4.7$ ms and $\tau_2 = 4.7$ ms were employed and the $^{13}C\alpha$ 180° pulse in the center of the $2\Delta_1$ period were replaced by a 180° selective pulse affecting only the $^{13}C\alpha$ chemical shift range in order to avoid the evolution of the $^{1}J_{C\alpha C\beta}$ coupling. In this case the transfer efficiency would increase to 0.160. However, the $^{1}J_{C\alpha C\beta}$ coupling would evolve when residue *i* is serine or threonine, drastically reducing the transfer efficiency for these residues. Therefore, we prefer not to use this alternative implementation. Likewise, it might be thought that the H-flip approach (Bermel et al. 2009b) would increase the sensitivity of the experiment. However, $^{1}H\alpha$ spins of IDPs do not experience significant longitudinal relaxation enhancement by their selective manipulation (Bermel et al. 2013). Furthermore, the H-flip scheme prevents using ^{1}H decoupling during the acquisition period, thereby broadening the $^{13}C'$ resonances by ^{2}J and ^{3}J ^{1}H–$^{13}C'$ couplings. Accordingly, we have not sought to include the H-flip scheme in the experiment.

The described experiment, hacaCONcaNCO, corresponds to a 4D spectrum, which demands a very long experimental

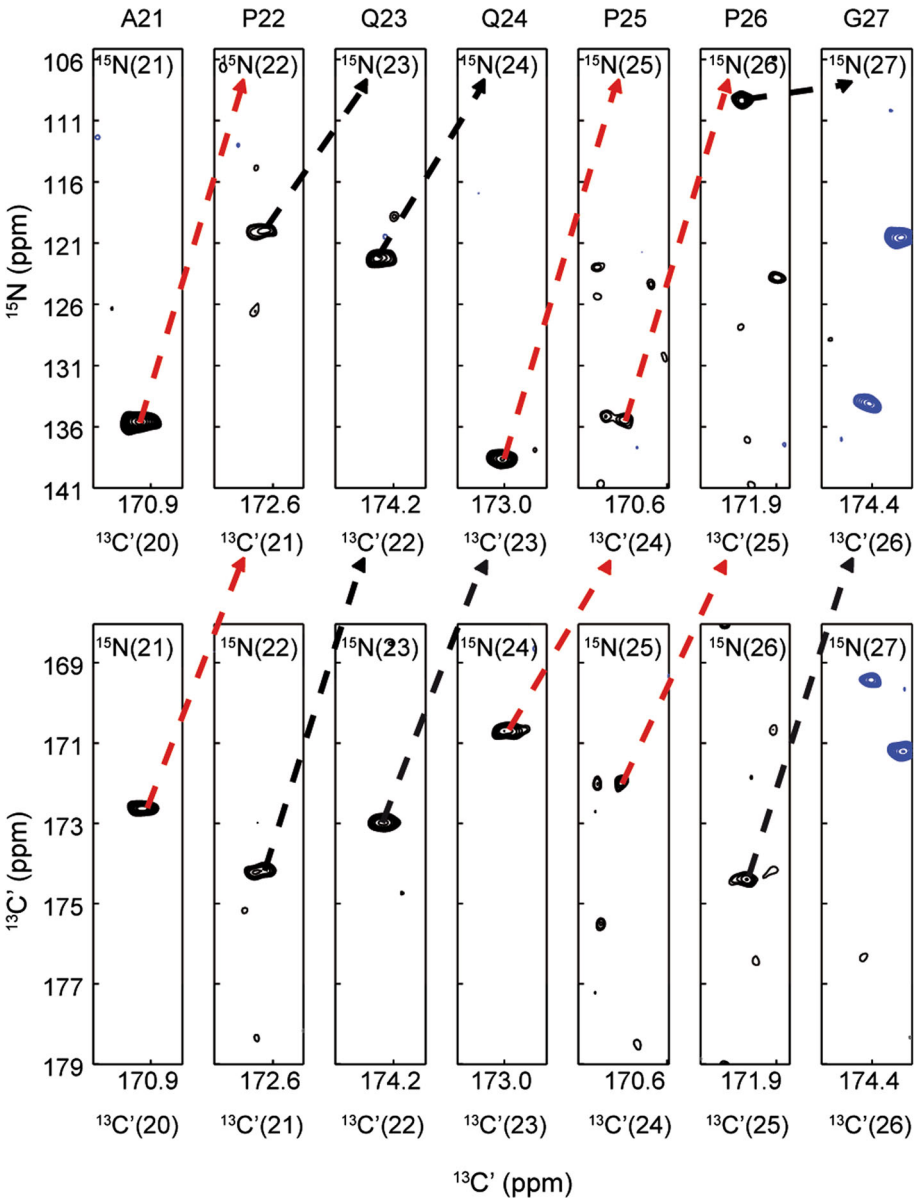

Fig. 3 Illustration of the method for assigning the protein $^{13}C'$ and ^{15}N resonances. Starting from a ^{15}N–$^{13}C'$ correlation observed in a 2D CON spectrum, the ^{15}N and $^{13}C'$ chemical shift values of the succeeding residue are obtained from the sequential cross peaks in the hacacoNcaNCO and hacaCOncaNCO spectra, respectively. Assignment is exemplified using the residues 21–27 of Nupr1. In *top* of each *panel* the ^{15}N frequency that has been fixed to extract the particular F1–F3 plane are reported. Sequential connections Xxx-Pro (highlighted with *red arrows*), absent in the hNcocaNCO and hnCOcaNCO spectra, are here visible. Note also that the negative sign of the sequential correlations observed in the last cross-section identify the residue as a glycine

time to obtain satisfactory resolution, unless non-uniform-sampling methods are used. However, the same information is available from two 3D experiments, hacacoNcaNCO showing the correlation $^{15}N(i+1)$–$^{15}N(i)$–$^{13}C'(i-1)$ and hacaCOncaNCO displaying the correlation $^{13}C'(i)$–$^{15}N(i)$–$^{13}C'(i-1)$. This approach requires a much shorter experimental time, and is the one we have used.

Correlation of consecutive C′–N groups can also be obtained with the 5D (HCA)CONCACON experiment (Bermel et al. 2013). Theoretical calculation for this experiment, using the parameters given above, gives a transfer efficiency of 0.101 for the sequential peak and 0.035 for the auto-correlated peak. Evolution of the $^{1}J_{NC\delta}$ coupling of prolines during the first period in which the magnetization resides in ^{15}N will reduce these figures by about 33 % in the case of the Pro-Xxx sequential peaks and of the proline auto-correlated peaks. Therefore, the new proposed experiment shows a clear sensitivity advantage in

the case of IDPs. Furthermore, the absence of auto-correlated peaks in our experiment produces less congested spectra, which is also advantageous.

Application and assignment protocol

We have applied the novel experiments in concert with a 2D CON spectrum to complete the assignment of Nupr1, an IDP of 93 residues we used to illustrate the hNcocaNCO and hnCOcaNCO experiments. Figure 3 exemplifies the assignment protocol. The procedure starts from a $^{15}N(i)$–$^{13}C'(i-1)$ correlation peak observed in the 2D CON spectrum. The coordinates of this peak are used to locate the chemical shift of $^{15}N(i+1)$ in the 3D hacacoNcaNCO spectrum and the chemical shift of $^{13}C'(i)$ in the 3D hacaCOncaNCO spectrum. These two chemical shifts identify a peak in the 2D CON spectrum, corresponding to the next amide linkage in the protein sequence. Using this process iteratively, employing each time the coordinates of the last incorporated CON peak, a chain of consecutive residues of increasing length is constructed. This daisy chaining stops when one of the following conditions is reached:

a) No following CON peak is obtained. This situation appears when the chain reaches the C-terminal residue, but also may occur because the signals in the 3D spectra are too weak to be properly observed.

b) The method provides more than one possible successor. This case arises when two or more amide linkages have very similar $^{13}C'$ and ^{15}N chemical shifts, so that the corresponding peaks overlap in the 2D CON spectrum. In this situation none of the potential successors is added to the chain.

c) The new CON peak coincides with the first CON peak of a chain previously obtained. In this case both chains are merged.

Once a stopping condition is reached, the process is repeated starting from an unused CON peak. The process of building chains is continued until all peaks in the CON spectrum have been used. Advantages of the new pulse sequences are clearly illustrated in Fig. 3, where the sequential peaks i–$i+1$ absent in the hNcocaNCO and hnCOcaNCO spectra when residue $i+1$ is a proline (Pantoja-Uceda and Santoro 2013a) are visible. The sequential specific assignment is thus not interrupted by the presence of prolines. Furthermore, the sign of the correlation peaks allows to identify glycines. These two advantageous features of the new pulse sequences have allowed us to confirm our previous assignment of Nupr1 (Aguado-Llera et al. 2013; Pantoja-Uceda and Santoro 2013a) and complete it.

In an ideal case the described protocol will end with a single stretch of length equal to the number of residues of the protein minus one. Therefore the matching of this stretch to the protein sequence will be obvious. In other cases, however, the protocol will result in several stretches. So, the matching of the stretches to the protein sequence requires some additional information. In this respect, the identification of glycines provided by the experiment is very useful. Furthermore the prolines, that can be identified by its characteristic ^{15}N chemical shift or by comparing 2D CON spectra starting from ^{1}HN and starting from $^{1}H\alpha$, can be used to check the correctness of the matching or as additional anchor points. If this information is not enough to properly determine the matching, more anchor points can be obtained by running some of the CAS-NMR experiments (Bermel et al. 2012b), that give CON spectra containing correlations arising from a particular type of amino acid residue. Alternatively a β-carbon edited CON experiment (Pantoja-Uceda and Santoro 2013b), which classifies all CON correlations into 6 groups using a single experiment, can be used.

Conclusions

In this manuscript we have presented two novel ^{13}C-detected 3D experiments for the sequence-specific assignment of IDPs. The sequential specific assignment is performed correlating the $^{13}C'$ and ^{15}N chemical shifts of two consecutive amide linkages. The new experiments overcome the limitations of our previous pulse sequence (Pantoja-Uceda and Santoro 2013a), namely the absence of correlations in which the magnetization starts at prolines, at a moderated cost in sensitivity. A further advantage of the new experiments is the identification of glycine residues. The recently published 5D (HCA)CONCACON pulse sequence (Bermel et al. 2013) also yields all the correlations between consecutive C′–N groups. However, our sequence offers several advantages: larger magnetization transfer efficiency and absence of auto-correlation peaks, therefore decreasing spectral crowding.

For the assignment of IDPs larger than the one studied here the use of two 3D spectra may be insufficient. In such cases our experiment can be run in a 4D mode or even incorporate Cα chemical shift labeling to transform it into a 5D experiment. A careful analysis of the 2D CON spectrum will be a good guide to select the more adequate approach in each case.

During the development of our pulse sequence the group of Permi (Hellman et al. 2014) has published a similar pulse sequence, (HCA)CON(CAN)H, which differs from ours in the final coherence transfer step and uses ^{1}H-detection. Although the ^{1}H-detection implies an increase in the intrinsic sensitivity of the experiment, their use is not without disadvantages. First, it is not possible to detect the

correlations in which the magnetization ends at proline. Furthermore, in the case of IDPs the amide protons typically exchange rapidly with the water, thus giving rise to wide or even absent signals (Hsu et al. 2009). Also, the amide proton signals of IDPs show poorer dispersion than those of $^{13}C'$, so the sequence-specific assignment will be more difficult. Finally, the amide proton signal is a doublet due to coupling with $^{1}H\alpha$, while that of the $^{13}C'$ is a singlet if the IPAP virtual decoupling technique is applied. Moreover, the $^{13}C'$ signals are narrower than those of the amide protons. These last facts together with the existence of ^{13}C-detection optimized probes partially compensate the intensity loss.

Pulse sequences for the experiments in Bruker language can be obtained from the web page http://rmnpro.iqfr.csic.es.

Acknowledgments This work was supported by project CTQ2011-22514 from the Spanish Ministerio de Economía y Competitividad. The authors thank J.L. Neira (Universidad Miguel Hernández, Alicante, Spain) and J.L. Iovanna (Centre de Recherche en Cancérologie, Marseille, France) for the Nupr1 sample.

References

Aguado-Llera D, Hamidi T, Doménech R, Pantoja-Uceda D, Gironella M, Santoro J, Velázquez-Campoy A, Neira JL, Iovanna JL (2013) Deciphering the binding between Nupr1 and MSL1 and their DNA-repairing activity. PlosOne 8:e78101

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, Felli IC, Piccioli M, Pierattelli R (2006) ^{13}C detected protonless NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

Bermel W, Felli IC, Kümmerle R, Pierattelli R (2008) ^{13}C direct detection biomolecular NMR. Concepts Magn Reson 32A:183–200

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 experiments. 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). ChemBioChem 13:2425–2432

Bermel W, Felli IC, Gonnelli L, Kozminski W, Piai A, Pieratelli R, Zawadzka-Kazimierczuk A (2013) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361

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

Böhlen JM, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

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

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

Feng W, Rios CB, Montelione G (1996) Phase labeling of C–H and C–C spin-system topologies: application in PFG-HACANH and PFG-HACA(CO)NH triple-resonance experiments for determining backbone resonance assignments in proteins. J Biomol NMR 8:98–104

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

Hellman M, Piirainen H, Jaakola V-P, Permi P (2014) Bridge over troubled proline: assignment of intrinsically disordered proteins using (HCA)CON(CAN)H and (HCA)N(CA)CO(N)H experiments concomitantly with HNCO and i(HCA)CO(CA)NH. J Biomol NMR 58:49–60

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

Hsu ST, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Jensen MR, Markwick PRL, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P, Blackledge M (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 17:1169–1185

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, Hosur RV (2011) hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone ($^{1}H^{N}$, ^{15}N and $^{13}C'$) resonances in $^{15}N/^{13}C$-labeled proteins. Magn Reson Chem 49:575–583

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

Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. Biophys J 98:2383–2390

Motackova V, Novacek J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Zidek L, Sanderova H, Krasny L, Kozminski W, Sklenar 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

Novacek J, Haba NY, Chill JH, Zidek L, Sklenar V (2012) 4D nonuniformly sampled HCBCACON and $^{1}J(NC\alpha)$-selective HCBCANCO experiments for the sequential assignment and

chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

Pantoja-Uceda D, Santoro J (2013a) Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR 57:57–63

Pantoja-Uceda D, Santoro J (2013b) A suite of amino acid residue type classification pulse sequences for ^{13}C-detected NMR of proteins. J Magn Reson 234:190–196

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

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456

Romero P, Obradovic Z, Li XH, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins Struct Funct Genet 42:38–48

Sahu D, Bastidas M, Showalter SA (2014) Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. Anal Biochem 449:17–25

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

Shaka AJ, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

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

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427

Wen J, Wu J, Zhou P (2011) Sparsely sampled high-resolution 4-D experiments for efficient backbone resonance assignment of disordered proteins. J Magn Reson 209:94–100

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