

# New amino acid residue type identification experiments valid for protonated and deuterated proteins

David Pantoja-Uceda · Jorge Santoro

Received: 1 June 2012 / Accepted: 4 August 2012 / Published online: 4 September 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** Two experiments are presented that yield amino acid type identification of individual residues in a protein by editing the  $^1\text{H}$ – $^{15}\text{N}$  correlations into four different 2D sub-spectra, each corresponding to a different amino acid type class, and that can be applied to deuterated proteins. One experiment provides information on the amino acid type of the residue preceding the detected amide  $^1\text{H}$ – $^{15}\text{N}$  correlation, while the other gives information on the type of its own residue. Versions for protonated proteins are also presented, and in this case it is possible to classify the residues into six different classes. Both sequential and intraresidue experiments provide highly complementary information, greatly facilitating the assignment of protein resonances. The experiments will also assist in transferring the assignment of a protein to the spectra obtained under different experimental conditions (e.g. temperature, pH, presence of ligands, cofactors, etc.).

**Keywords** Proteins · Amino acid residue type editing · Sequential assignment · Triple-resonance experiments · HN(CO)CACB · Intra-HNCACB

## Introduction

Identification of spin systems corresponding to different amino acid types can be of great aid in the assignment of NMR spectra of proteins, a prerequisite in all investigations of

proteins by NMR. The amino acid residue type determination can be obtained from NMR data in different ways: (1) by statistical analysis of  $^{13}\text{C}$  chemical shifts (Grzesiek and Bax 1993; Atreya et al. 2000), (2) by the selective isotope labeling (unlabeling) of amino acid residues in an otherwise unlabeled (or labeled) protein (Muchmore et al. 1989; McIntosh and Dahlquist 1990; Lee et al. 1995; Atreya and Chary 2001; Ohki and Kainosho 2008; Tong et al. 2008; Krishnarjuna et al. 2011) or (3) by using NMR experiments which make use of the side chain topology of the different amino acids to detect them selectively (Dötsch and Wagner 1996; Dötsch et al. 1996a, b, c; Feng et al. 1996; Rios et al. 1996; Schubert et al. 1999, 2001a, b, c, 2005; Barnwal et al. 2008).  $^{13}\text{C}$  chemical shifts, used in the first procedure, are not very discriminating in terms of amino acid residue type determination, and only Ala, Gly and Ser/Thr residues can be identified unambiguously. The main drawback of the second approach is that it requires the preparation of multiple differently labeled protein samples, resulting in experimental and economic costs. In contrast, the third approach uses only a single uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled sample, and discriminates the amino acid residue type by using a different experiment for each amino acid or group of amino acids sharing the same topology. Nevertheless this third approach has several weaknesses: it requires many spectra to be registered, which is very time consuming; the multiple transfer steps required for some pulse sequences result in low experimental sensitivity; the appearance of breakthrough peaks produce problems, especially in cases in which spectra for all amino acid residue types are not registered; finally, some of the filters incorporated into the pulse sequences are long and as a consequence these sequences do not work well for large proteins.

To overcome some of these drawbacks, several experiments for amino acid residue type discrimination of all  $^1\text{H}$ – $^{15}\text{N}$  HSQC peaks in a single experiment, and therefore in

**Electronic supplementary material** The online version of this article (doi:10.1007/s10858-012-9665-y) contains supplementary material, which is available to authorized users.

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

a minimal experimental time, have been recently introduced (Lescop et al. 2008; Pantoja-Uceda and Santoro 2008; Lescop and Brutscher 2009; Feuerstein et al. 2012). In these experiments the properties of the  $^{13}\text{C}^\beta$  carbons are used to selectively invert the signals from certain amino acid residue types: odd/even number of aliphatic  $^{13}\text{C}^\gamma$  carbons, odd/even number of  $\text{H}^\beta$  protons, presence/absence of a  $^{13}\text{CO}$  attached to  $^{13}\text{C}^\beta$ , etc. Several of these manipulations are combined in a Hadamard encoding form to produce an experiment whose decoding yields subspectra corresponding to amino acid residues of similar topology. Some breakthrough between subspectra is usually observed in the experiments owing to variations in the coupling constant values and to pulse imperfections. However, the intensity of these artifacts is always much smaller than the intensity of the correlation peak detected in the correct subspectrum. Therefore, the subspectrum with the highest intensity at a given  $^1\text{H}$ ,  $^{15}\text{N}$  frequency pair gives unambiguous amino acid residue type discrimination, and the presence of breakthrough peaks does not constitute a problem. This methodology has been developed to provide information on the type of the amino acid residue preceding the detected amide  $^1\text{H}$ – $^{15}\text{N}$  frequency pair (Lescop et al. 2008; Pantoja-Uceda and Santoro 2008; Lescop and Brutscher 2009) or on the intrasidue amino acid type (Pantoja-Uceda and Santoro 2008; Feuerstein et al. 2012). The main disadvantage of these experiments is the rapid loss in sensitivity with increasing protein molecular weight due to the long delay periods used in the pulse sequences. The remedy that is commonly used in multidimensional NMR experiments to reduce relaxation losses is to work with deuterated proteins. However, the above cited editing pulse sequences cannot be applied to deuterated proteins owing to the fact that the magnetization originates from the side chain  $\text{H}^\beta$  protons of the different residues, which are lost upon deuteration.

Very few experiments for amino acid residue type editing in deuterated proteins have been proposed. Wagner and coworkers (Dötsch et al. 1996c) described the selection of  $^1\text{H}$ – $^{15}\text{N}$  HSQC peaks of residues following Ser, Ala, Cys and Gly on the basis of the fact that these residues do not contain  $\text{C}^\gamma$  carbons. In a variation,  $^1\text{H}$ – $^{15}\text{N}$  peaks of residues following aromatic residues are selected in addition. Muthandiram et al. (1997) proposed a triple-resonance pulse scheme which records  $^1\text{H}$ – $^{15}\text{N}$  correlations of residues that immediately follow a methyl containing amino acid residue. The experiment selects for  $\text{CH}_2\text{D}$  methyl types, therefore requiring fractionally deuterated protein samples. Finally, the group of Kay (Tugarinov and Kay 2003) described the production of Ile, Leu, Val-methyl protonated  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled proteins, which allows the detection of  $^1\text{H}$ – $^{15}\text{N}$  correlations of these residues. These three experiments only allow to determine the amino acid residue type of a limited number of  $^1\text{H}$ – $^{15}\text{N}$  correlation peaks.

In this paper we introduce two experiments for amino acid residue type discrimination based on the same methodology employed in Pantoja-Uceda and Santoro (2008), that can be applied to deuterated proteins, since the magnetization starts on the amide proton. One experiment provides the classification of all the  $^1\text{H}$ – $^{15}\text{N}$  correlation peaks according to the nature of the preceding residue, and the other according to the nature of its own residue. Together, these two experiments provide a comprehensive classification of the  $^1\text{H}$ – $^{15}\text{N}$  HSQC signals, and therefore of the spin systems associated with them, and can be of great aid in the sequence-specific assignment process. Versions of the experiments are presented for both deuterated and protonated proteins.

### Pulse sequences

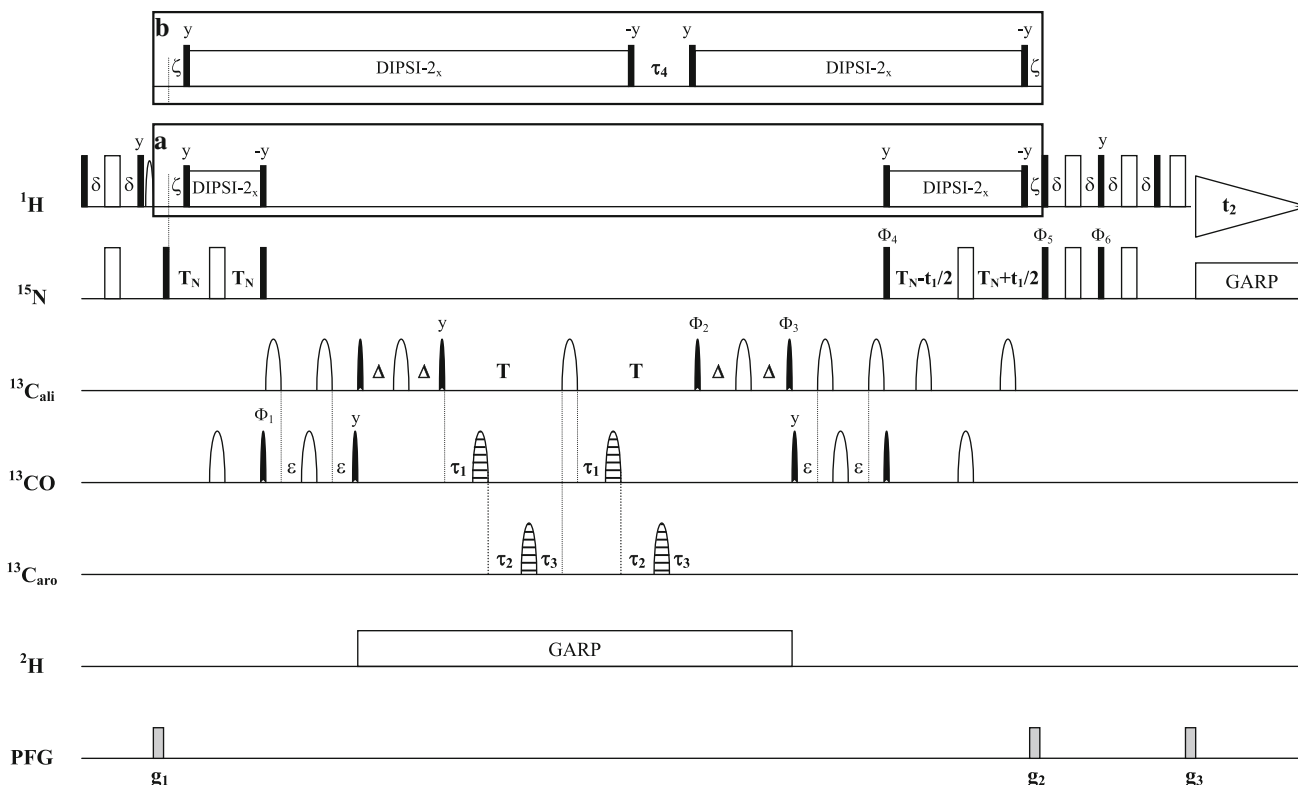
A pulse sequence for amino acid residue type determination valid for deuterated proteins should start with the magnetization of the amide protons and use an ‘out-and-back’ approach to transfer the coherences to the  $^{13}\text{C}^\beta$  carbons, perform the residue type discrimination, and transfer the coherences back to the amide protons for detection. A 2D pulse sequence designed using this approach, based on the HN(CO)CACB experiment (Yamazaki et al. 1994), is shown in Fig. 1. Since in this experiment the magnetization is transferred from the amide proton to the  $^{13}\text{C}^\beta$  carbon of the previous residue, it provides information on the amino acid type of the residue preceding the detected amide  $^1\text{H}$ – $^{15}\text{N}$  correlation. The pulse sequence is identical to the original one, except during the 2T period, in which the magnetization is on the  $^{13}\text{C}^\beta$  carbon. During this period, several spin manipulation techniques are applied to obtain the discrimination of amino acid residue types:

- (a) The aliphatic  $^{13}\text{C}$ – $^{13}\text{C}$  couplings are active during the whole 2T period. At the end of this period the product operator that is converted back into a  $^{13}\text{C}^\alpha$  coherence, thus ending in observable magnetization, is

$$8N_z(i)C_z'(i-1)C_z^\alpha(i-1)C_y^\beta(i-1)\cos(2\pi TJ_{C\beta-C\alpha}) \cdot \cos^{nG}(2\pi TJ_{C\beta-C\gamma}) \quad (1)$$

where  $nG$  is the number of aliphatic  $^{13}\text{C}^\gamma$  carbons attached to the  $^{13}\text{C}^\beta$  carbon. Accordingly, a value of  $T = 1/(2J_{C-C})$  will lead to a sign inversion for amino acid types with a single aliphatic  $^{13}\text{C}^\gamma$  carbon (E, Q, M, K, R, P, L, and T) with respect to the rest of amino acids (F, Y, W, H, D, N, C, S, A, V, and I).

- (b) Insertion into the 2T period of selective  $180^\circ$  pulses affecting only the aromatic carbons makes it possible to evolve the  $J_{C\beta-C\text{aro}}$  couplings. Therefore,  $^{13}\text{C}^\beta$  carbons of aromatic residues will have an additional modulation with the factor  $\cos(2\pi T_X J_{C\beta-C\text{aro}})$ , where  $2T_X$  is the



**Fig. 1** Pulse sequence of the sequential beta-carbon edited HSQC experiment for deuterated proteins. For the 8-step encoding scheme used with protonated proteins the <sup>1</sup>H pulses in the *rectangle (a)* are substituted by those in the *rectangle (b)* and the <sup>2</sup>H decoupling is omitted. All radiofrequency pulses are applied along the *x* axis unless indicated. 90° and 180° rectangular pulses are represented by *filled* and *unfilled bars*, respectively. <sup>13</sup>C pulses have the shape of gaussian cascades Q5 (90°, *black filled shapes*) and Q3 (180°, *open shapes*) with durations of 307 and 192 μs at 800 MHz, respectively. <sup>13</sup>C aliphatic pulses are centered at 41 ppm and <sup>13</sup>CO pulses at 173 ppm. *Striped shapes* correspond to Q3 gaussian cascades of 340 μs duration (at 800 MHz), applied at 120 ppm (<sup>13</sup>Caro) and 180 ppm (<sup>13</sup>CO). The delays are adjusted to  $\delta = 2.3$  ms;  $\zeta = 5.5$  ms;

$T_N = 12.0$  ms;  $\epsilon = 4.0$  ms;  $\Delta = 6.5$  ms;  $T = 12.5$  ms. DIPSI is applied with a field strength of 4,545 Hz and Garp with a strength of 1,250 Hz both in the <sup>2</sup>H and the <sup>15</sup>N channels. For every  $t_1$  value, 4 repetitions (8 repetitions in the case of protonated proteins) of the experiment are separately recorded with the parameter settings given in Table 1. 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 30, 80 and 8.1 % of the maximal intensity. Phase cycle:  $\phi_1 = 8(x), 8(-x)$ ;  $\phi_2 = -y, y$ ;  $\phi_3 = 4(x, -x), 4(-x, x)$ ;  $\phi_4 = x$ ;  $\phi_5 = x, x, -x, -x$ ;  $\phi_6 = -y, -y, y, y$ ;  $\phi_{31}(\text{receiver}) = x, -x, -x, x, 2(-x, x, x, -x), x, -x, -x, x$ . For quadrature detection echo/antiecho data are recorded by sign alternation of the gradient  $g_2$  with simultaneous inversion of phase  $\phi_6$ . After each  $t_1$  increment the phases  $\phi_4$  and  $\phi_{31}$  are inverted

effective evolution time of the  $J_{C\beta-Caro}$  coupling,  $2 \cdot (\tau_1 + p + \tau_2 - \tau_3)$  with the arrangement of Fig. 1. For  $T_X \sim 0$  this cosine term will be positive and approximately equal to 1, while for  $2T_X \sim 1/J_{C\beta-Caro}$  it will be negative and close to  $-1$ . The evolution of the <sup>13</sup>C<sup>β</sup> coherences of residues lacking an aromatic <sup>13</sup>C<sup>γ</sup> carbon is not affected by the aromatic selective pulses. These coherences will evolve as described by Eq. (1), independently of the position of the aromatic selective pulses. Therefore, if two experiments are acquired, one with  $2T_X \sim 0$  and a second one with  $2T_X \sim 1/J_{C\beta-Caro}$ , its difference will show signals only for the aromatic residues, F, Y, W and H. In contrast, the sum of the two datasets will display the signals of all other residues, which are not affected by the aromatic selective pulses. Small signals of aromatic residues can be present in this second subspectrum if the two values of  $2T_X$  do not

differ exactly in  $1/J_{C\beta-Caro}$ . Nevertheless, these breakthrough peaks are easily recognized by comparing the intensity in both subspectra, and therefore do not hinder a correct amino acid residue type classification.

(c) As in the case of aromatic selective pulses, inserting selective 180° <sup>13</sup>CO pulses into the 2T period makes it possible to evolve the coupling between the <sup>13</sup>C<sup>β</sup> carbon and an attached <sup>13</sup>CO for a time that depends on the position of these pulses, giving a modulation term of  $\cos(2\pi T_Y J_{C\beta-CO})$  for the observable magnetization.  $2T_Y$  is the effective evolution time of the  $J_{C\beta-CO}$  coupling,  $2 \cdot (\tau_1 - \tau_2 - p - \tau_3)$  with the arrangement of Fig. 1. Again, the combination of two experiments, one with  $2T_Y \sim 0$  and a second one with  $2T_Y \sim -1/J_{C\beta-CO}$ , allows the residues having a <sup>13</sup>CO attached to the <sup>13</sup>C<sup>β</sup> carbon, D and N, to be distinguished from the rest.

During the 2T period  $^{13}\text{C}^\alpha$  coherences of glycines, and of all other residues will exist if the delay  $\Delta$  is not exactly matched to  $1/(4J_{\text{C}\alpha\text{-C}\beta})$ . Therefore, knowledge of the response of  $^{13}\text{C}^\alpha$  coherences to the above spin manipulations is important in order to classify glycine residues and to avoid interferences in the amino acid residue type sorting process. In the regular HN(CO)CACB experiment, the  $^{13}\text{C}^\alpha$  coherences existing during 2T give signals of opposite sign to the  $^{13}\text{C}^\beta$  coherences. Thus, after manipulation (a), glycines will appear with the same sign as residues having an even number of aliphatic  $^{13}\text{C}^\gamma$  carbons whereas the remaining  $^{13}\text{C}^\alpha$ 's will have the opposite sign. The  $^{13}\text{C}^\alpha$  coherences are not affected by the aromatic selective  $180^\circ$  pulses, but they are affected by the  $^{13}\text{CO}$  selective pulses. Consequently, glycine  $^{13}\text{C}^\alpha$  coherences behave exactly like the  $^{13}\text{C}^\beta$  coherences of D and N do. The remaining  $^{13}\text{C}^\alpha$  coherences also behave like the  $^{13}\text{C}^\beta$  coherences of D and N, but produce signals of the opposite sign.

The combination of the spin manipulations described in (a), (b), and (c) gives rise to the 4-step encoding scheme shown in the four first rows of Table 1. The decoding of this 4-step experiment according to the scheme given in Table 2 yields three subspectra, which allow to classify the signals into 4 groups. The first subspectrum shows the  $^1\text{H}\text{-}^{15}\text{N}$  correlation peaks of residues following aromatic amino acids, F, Y, W, and H, with a sign that we have arbitrarily assigned as positive and used as the reference for the sign of the signals in the other subspectra. The second subspectrum shows positive  $^1\text{H}\text{-}^{15}\text{N}$  correlation peaks of residues following D, N or G, and negative peaks for all residue types, but those succeeding G. These negative peaks, coming from the  $^{13}\text{C}^\alpha$  coherences present during the 2T period, could cancel the D and N signals, preventing their observation and the NH classification. Therefore, removal of  $^{13}\text{C}^\alpha$  coherences, which is achieved using a  $\Delta$  value of  $1/(4J_{\text{C}\alpha\text{-C}\beta})$ , is important. This value also optimizes the transfer to the  $^{13}\text{C}^\beta$  coherences, improving the

sensitivity of the experiment. Finally, the third subspectrum displays positive  $^1\text{H}\text{-}^{15}\text{N}$  correlation peaks for residues following C, S, A, V, or I, and negative peaks for residues following E, Q, M, K, R, P, L, or T.

The above described pulse sequence can be applied to fully or partially deuterated proteins as well as to protonated proteins. In protonated proteins a further manipulation of the spins, evolution with the  $^{13}\text{C}\text{-}^1\text{H}$  couplings, can be introduced into the 2T period. If evolution of these couplings for a period  $1/J_{\text{C-H}}$  is allowed, coherences of carbons with an odd number of attached protons will change its sign. Thus, the sum of two experiments, one with and one without  $J_{\text{C-H}}$  evolution will give rise only signals corresponding to amino acids having  $^{13}\text{C}^\beta$  carbons with an even number of attached protons and to glycine. The difference of these experiments shows the remaining signals, those corresponding to amino acids having  $^{13}\text{C}^\beta$  carbons with an odd number of attached protons. Combination of this new spin manipulation with those previously described produces the 8-step encoding scheme given in Table 1. Decoding of the experiment yields the 4 subspectra given in Table 3. Furthermore, by taking into account the sign of the correlation signals, it is possible to classify the NH signals into 6 different groups. In this case it is not necessary to eliminate the  $^{13}\text{C}^\alpha$  coherences present during the 2T period, since they do not interfere with the signals coming from  $^{13}\text{C}^\beta$ . Nevertheless, it is advisable to use a  $\Delta$  value similar to the one used in the 4-step encoding experiment, since this value also optimizes the transfer to the  $^{13}\text{C}^\beta$  coherences, which improves the experimental sensitivity.

Identification of the amino acid type of the detected residue provides additional information that greatly facilitates the resonance assignment process (Pantoja-Uceda and Santoro 2008; Feuerstein et al. 2012). This identification is especially useful in the case of deuterated proteins, since the above described experiment classifies residues into only four different groups. An obvious way to obtain this additional categorization, valid for deuterated proteins, is

**Table 1** Experimental settings used for the amino acid type encoding

	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$
Exp 1	$T/2 - p^a - 2 \mu\text{s}$	$4 \mu\text{s}$	$T/2 - p - 2u$	$4 \mu\text{s}$
Exp 2	$T/2 - p - 2 \mu\text{s}$	$C1^b - p \times 3/2 + 2 \mu\text{s}$	$T/2 + p/2 - C1$	$4 \mu\text{s}$
Exp 3	$T/2 + p/2 - C2^c$	$C2 - p \times 3/2 + 2 \mu\text{s}$	$T/2 - p - 2u$	$4 \mu\text{s}$
Exp 4	$T/2 + p/2 - C2$	$C1 + C2 - p \times 3$	$T/2 + p/2 - C1$	$4 \mu\text{s}$
Exp 5	$T/2 - p - 2 \mu\text{s}$	$4 \mu\text{s}$	$T/2 - p - 2u$	$7.4 \text{ ms}$
Exp 6	$T/2 - p - 2 \mu\text{s}$	$C1 - p \times 3/2 + 2 \mu\text{s}$	$T/2 + p/2 - C1$	$7.4 \text{ ms}$
Exp 7	$T/2 + p/2 - C2$	$C2 - p \times 3/2 + 2 \mu\text{s}$	$T/2 - p - 2u$	$7.4 \text{ ms}$
Exp 8	$T/2 + p/2 - C2$	$C1 + C2 - p \times 3$	$T/2 + p/2 - C1$	$7.4 \text{ ms}$

For deuterated proteins only the first 4 experiments are used and the column  $\tau_4$  has no meaning. For protonated proteins, the 8 experiments are employed

<sup>a</sup>  $p$  is the duration of the  $180^\circ$  selective pulse; <sup>b</sup>  $C1$  is  $1/(4J_{\text{C}\beta\text{-Caro}}) = 5.70 \text{ ms}$  (Löhr et al. 2000); <sup>c</sup>  $C2$  is  $1/(4J_{\text{C}\beta\text{-CO}}) = 5.40 \text{ ms}$  (Löhr et al. 2000)

**Table 2** Signs of the different amino acid types in the 4-step encoding scheme

	FYWH	DNG	CSAVI/EQMKRPLT
Exp 1	+	+	+/-
Exp 2	-	+	+/-
Exp 3	+	-	+/-
Exp 4	-	-	+/-

Linear combination of the 4 experiments with the signs given in each column yields 3 subspectra

**Table 3** Decoding matrix for the 8-step encoding scheme of Table 1

	FYWH	DNG	AVI/T	CS/EQMKRPL
Exp 1	+	+	+/-	+/-
Exp 2	-	+	+/-	+/-
Exp 3	+	-	+/-	+/-
Exp 4	-	-	+/-	+/-
Exp 5	+	+	-/+	+/-
Exp 6	-	+	-/+	+/-
Exp 7	+	-	-/+	+/-
Exp 8	-	-	-/+	+/-

to include the above described spin manipulations into a 2D version of the HNCACB experiment (Wittekind and Müller 1993). Nevertheless, this approach has some problems. The subspectra can be very crowded, since two  $^1\text{H}$ - $^{15}\text{N}$  correlation peaks per residue will commonly be detected, one in the subspectrum corresponding to the type of the preceding residue and other in the subspectrum corresponding to the residue type of the detected amide proton. If both residues are of the same type, only one peak will appear. In this instance, the amino acid residue type determination will be ambiguous, since this case cannot be distinguished from the case in which the preceding and the detected residues are of different types, but the intrasidial peak is too weak to be detected. Also, if both amino acid residues belong to types that appear in the same subspectrum but with different signs, a great attenuation of the  $^1\text{H}$ - $^{15}\text{N}$  correlation peak will occur, and it may even disappear. To avoid all these drawbacks, the spin manipulations have been included in the intra-HNCACB experiment, derived from intra-HNCA pulse sequence (Nietlispach et al. 2002), that shows exclusively intrasidial peaks. The resulting pulse sequence appears in Fig. 2.

## Results and discussion

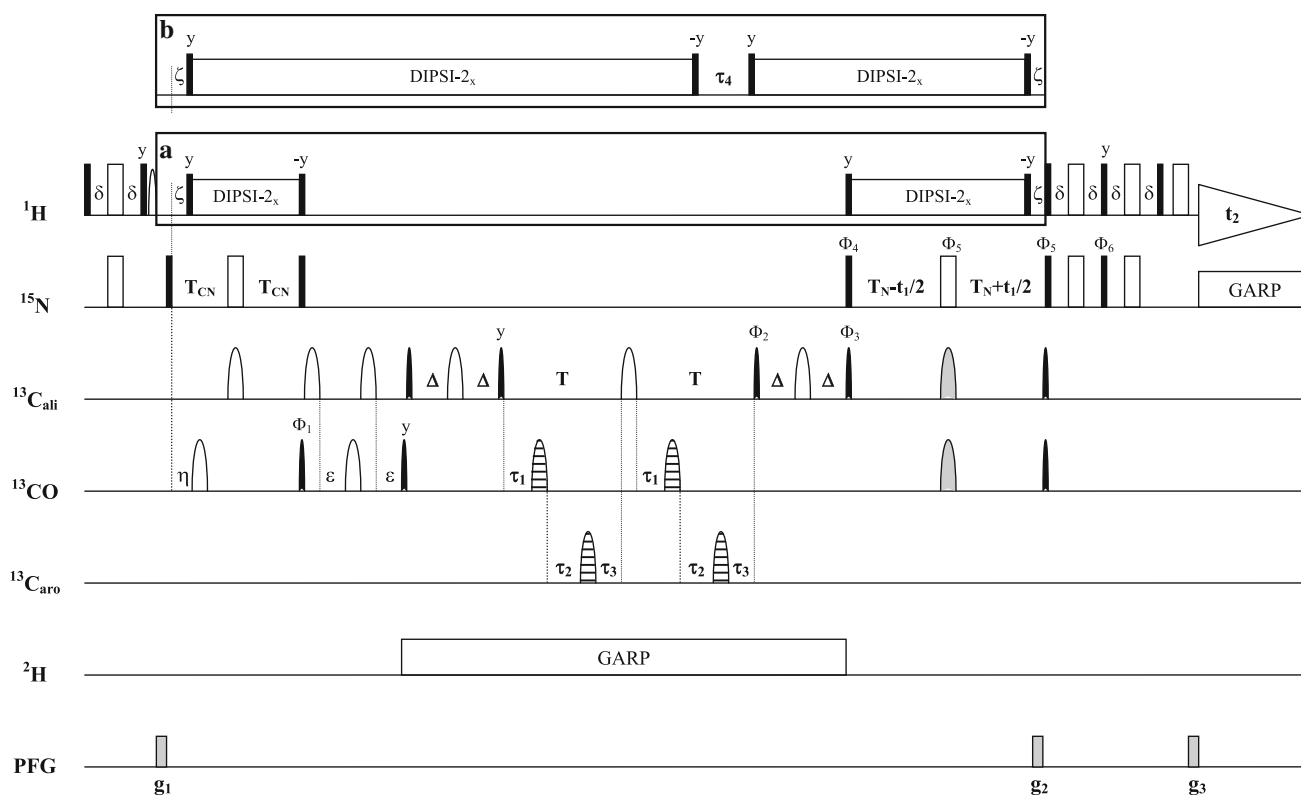
The above set of pulse sequences has been tested with samples of uniformly  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled ubiquitin (1 mM, 50 mM sodium phosphate, pH 7.0, 5 mm tube,

purchased from Giotto Biotech.) and uniformly  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled ubiquitin (1.7 mM, 10 mM potassium phosphate, pH 7.3, 5 mm tube, purchased from ASLA Biotech.). All the NMR experiments were performed at 298 K on a Bruker AV 800 spectrometer equipped with a TCI cryoprobe. The  $\beta$ -carbon edited spectra were acquired using  $64 \times 1,024$  complex data points with spectral widths of 3,243.720 Hz ( $^{15}\text{N}$ ) and 9,615.395 Hz ( $^1\text{H}$ ) and 1 s recycle time. 8 scans were used in the 4-step encoding version of the experiments, and 4 scans in the 8-step encoding. The duration of each experiment was 90 min. NMRPipe (Delaglio et al. 1995), and NMRView (Johnson and Blevins 1994) were used to process and analyze the NMR data, respectively.

Figure 3 shows the subspectra of deuterated ubiquitin obtained with both the sequential and the intrasidial  $\beta$ -carbon edited experiments. The high quality of the spectra allows the classification of all backbone  $^1\text{H}$ - $^{15}\text{N}$  correlations observed in the HSQC spectrum, both in terms of the type of the preceding residue and of the type of the own residue, except in one case, V26. In the HSQC spectrum of ubiquitin, the  $^1\text{H}$ - $^{15}\text{N}$  signal of V26 partially overlaps with that of E16. In the intrasidial experiment, both signals appear in the same subspectrum with opposite signs, which gives rise to a nearly complete cancellation of the V26 signal and a displacement of its maximum, preventing its amino acid residue type determination. In the sequential experiment, however, the  $^1\text{H}$ - $^{15}\text{N}$  signals of E16 and V26 appear in different subspectra, allowing the precise determination of the amino acid type of both preceding residues. Close inspection of the subspectra shows the presence of several breakthrough peaks. However, the intensity of these artifactual peaks is always low, less than 10 % of the intensity of the peak appearing in the correct subspectrum, and do not hamper a proper amino acid type classification.

Together with the signals coming from the backbone NH's, weak signals coming from the side chain amides of Asn and Gln appear in the DNG and EQMKRPLT subspectra of the sequential experiment (see Figure S1 in Supplementary Material). The  $\text{NH}_2$  moieties of Asn and Gln side chains should not produce signals at all, since the evolution time of the  $J_{\text{NH}}$  coupling,  $\zeta$ , is adjusted to maximize the transfers between  $^{15}\text{N}$  and  $^1\text{H}$  for a NH group, thereby abolishing the transfers for moieties of other multiplicities. However, proteins are usually dissolved in  $\text{H}_2\text{O}$  with 5–10 % of  $\text{D}_2\text{O}$ , resulting in the formation of detectable NHD groups in the Asn and Gln side chains. For Asn the pathway that yields observable magnetization starts on the NHD proton, progresses to the  $^{13}\text{C}^\alpha$  carbon, where the editing sequence is applied, and goes back to the NHD proton for detection. Since the editing is made on a carbon bounded to a  $^{13}\text{CO}$  group, Asn amide side chains give signals in the DNG subspectrum. For Gln, the editing





**Fig. 2** Pulse sequence of the intra-residual beta-carbon edited HSQC experiment for deuterated proteins. For protonated proteins the  $^1\text{H}$  pulses in the *rectangle* (a) are substituted by those in the *rectangle* (b) and the  $^2\text{H}$  decoupling is omitted. All radiofrequency pulses are applied along the  $x$  axis unless indicated.  $90^\circ$  and  $180^\circ$  rectangular pulses are represented by filled and unfilled bars, respectively.  $^{13}\text{C}$  pulses have the shape of gaussian cascades Q5 ( $90^\circ$ , black filled shapes) and Q3 ( $180^\circ$ , open shapes) with durations of 307 and 192  $\mu\text{s}$  at 800 MHz, respectively.  $^{13}\text{C}$  aliphatic pulses are centered at 41 ppm and  $^{13}\text{CO}$  pulses at 173 ppm. Striped shapes correspond to Q3 gaussian cascades of 340  $\mu\text{s}$  duration (at 800 MHz), applied at 120 ppm ( $^{13}\text{Caro}$ ) and 180 ppm ( $^{13}\text{CO}$ ). The simultaneous  $^{13}\text{C}_{\text{ali}}$  and  $^{13}\text{CO}$  pulses indicated in grey correspond to an adiabatic inversion pulse over the  $^{13}\text{CO}$  and  $^{13}\text{C}_{\alpha}$  regions, Chirp, 500  $\mu\text{s}$ , 25 % smoothing, 80 kHz sweep, 10.0 kHz strength. The delays are adjusted

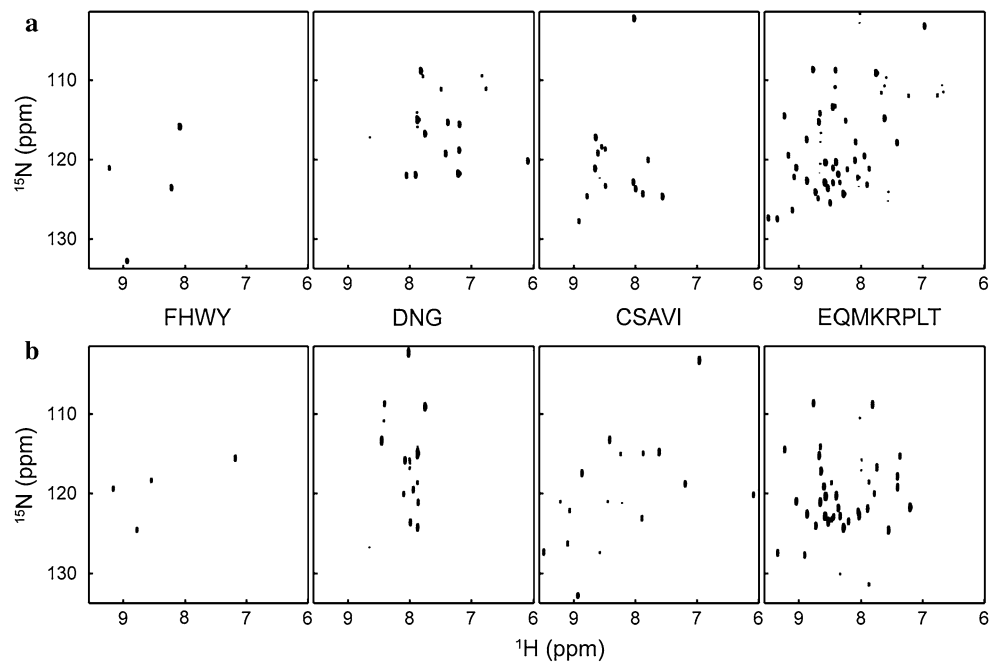
to  $\delta = 2.3$  ms;  $\zeta = 5.5$  ms;  $\eta = 16.5$  ms;  $T_{\text{CN}} = 26.0$  ms;  $\varepsilon = 4.2$  ms;  $\Delta = 6.5$  ms;  $T = 12.5$  ms;  $T_{\text{N}} = 15.8$  ms. DIPSI is applied with a field strength of 4,545 Hz and Garp with a strength of 1,250 Hz both in the  $^2\text{H}$  and the  $^{15}\text{N}$  channels. For every  $t_1$  value, 4 repetitions (8 repetitions in the case of protonated proteins) of the experiment are separately recorded with the parameter settings given in Table 1. 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 30, 80 and 8.1 % of the maximal intensity. Phase cycle:  $\phi_1 = 8(x), 8(-x)$ ;  $\phi_2 = -y, y$ ;  $\phi_3 = 2(x, -x), 2(-x, x)$ ;  $\phi_4 = y$ ;  $\phi_5 = x, x, -x, -x$ ;  $\phi_6 = -y, -y, y, y$ ;  $\phi_{31}(\text{receiver}) = x, -x, -x, x, 2(-x, x, x, -x), x, -x, -x, x$ . For quadrature detection echo/antiecho data are recorded by sign alternation of the gradient  $g_2$  with simultaneous inversion of phase  $\phi_6$ . After each  $t_1$  increment the phases  $\phi_4$  and  $\phi_{31}$  are inverted

is made when the magnetization is at the  $^{13}\text{C}^\beta$  carbon. Consequently, signals of Gln amide side chains appear in the EQMKRPLT subspectrum. Therefore, the sequential experiment not only permits the classification of the backbone NH's into different groups but also the distinction between N and Q side chain NHD resonances. The intraresidual experiment is based on the generation of a nitrogen coherence antiphase with respect to the two adjacent  $^{13}\text{C}^\alpha$  carbons and to its attached  $^{13}\text{CO}$ . Since no similar coherence can be generated by the side chain amide nitrogen's of N and Q, no side chain signals appear in the intraresidual subspectra.

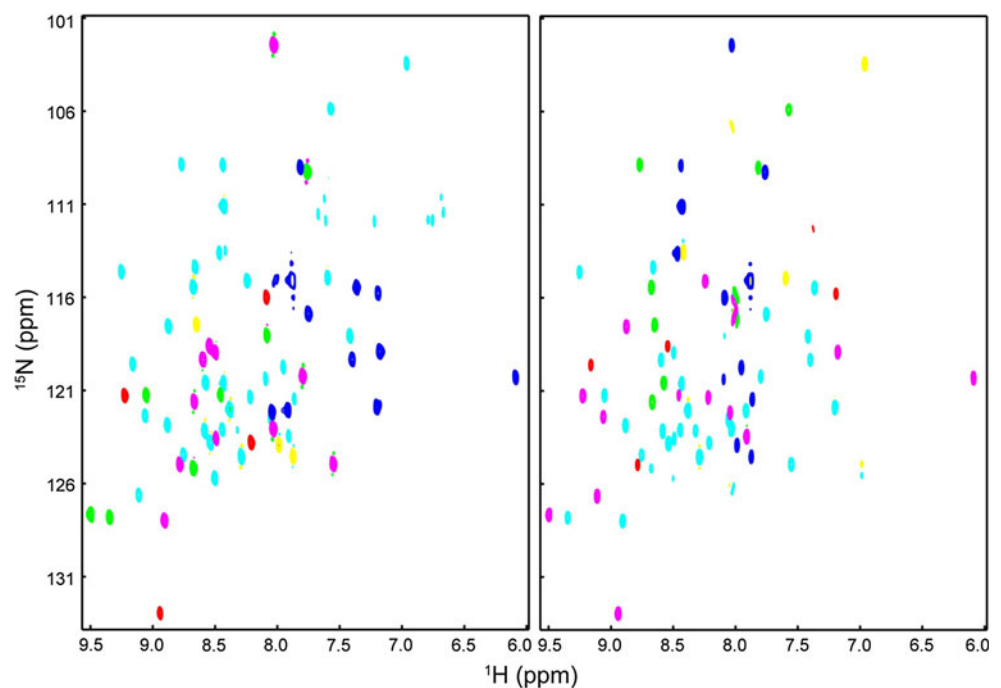
Results for the  $^{13}\text{C}, ^{15}\text{N}$  labeled ubiquitin sample using the 8-step encoding scheme are shown in Fig. 4. Once again the spectra show excellent quality, and all backbone  $^1\text{H}-^{15}\text{N}$

correlations observed in the HSQC spectrum can be properly classified, both in terms of the sequential and of the own residue. As with the four steps encoding scheme, weak signals coming from the side chain NHD moieties appear in the sequential subspectra. For Gln side chain, the editing is made when the magnetization is on the  $^{13}\text{C}^\beta$  carbon, which has two protons attached. Therefore, Gln side chain signals appear in the EQMKRPL subspectrum. For Asn, the editing of the side chain signal is made at  $^{13}\text{C}^\alpha$ , having one proton attached. Thus, Asn side chain signals do not appear in any of the subspectra corresponding to the backbone amides, and its observation requires a special decoding scheme. This decoding scheme is identical to that given in Table 3 for obtaining the DNG subspectrum, except the sign of the last four experiments is changed.

**Fig. 3** Sequential (*top*) and intraresidual (*bottom*)  $\beta$ -carbon edited ubiquitin subspectra of  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled ubiquitin. In the subspectrum corresponding to the residue type EQMKRPLT of the intraresidual experiment prolines are obviously not detected



**Fig. 4** Sequential (*left*) and intraresidual (*right*)  $\beta$ -carbon edited subspectra of  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled ubiquitin. The six subspectra corresponding to the different amino acid type classes are color coded and superposed on the same graph. Color code: FHWY, red; DNG, blue; AVI, magenta; T, green; CS, yellow; EQMKRPL (EQMKRL in the intraresidual spectrum), cyan. For interpretation of the references to color, the reader is referred to the web version of this article



The new sequential experiment yields the same amino acid type groups as the experiment in which the magnetization starts from the side chain  $\text{H}^\beta$  protons (Pantoja-Uceda and Santoro 2008). Therefore, it is interesting to compare the performances of both experiments. This comparison has been made by analyzing the intensity of the cross peaks of ubiquitin in the two experiments recorded with exactly the same parameters. Some cross peaks appear with higher intensity in the new experiment, while others show lower intensity. These results can be rationalized considering that

there are two different opposing effects. On one hand, the new pulse sequence is longer and the magnetization resides more time on an aliphatic carbon than in our previous sequence. Therefore, relaxation losses are more important in the new pulse sequence. On the other hand, the modulation term used in the new experiment to distinguish residues with an odd number of aliphatic  $^{13}\text{C}^\gamma$  carbons from the rest, Eq. (1), is larger than the modulation term of the experiment that starts from the side chain  $\text{H}^\beta$  proton (Pantoja-Uceda and Santoro 2008), especially for residues

with one or two aliphatic gamma carbons. These two effects explain the results observed in ubiquitin, namely a decreased intensity for aliphatic residues without aliphatic gamma carbons, more or less the same intensity for residues with one aliphatic gamma carbon and greater intensity for residues with two aliphatic  $^{13}\text{C}^\gamma$  carbons. Therefore, the new sequence is of similar quality to the former one for protonated proteins of the size of ubiquitin. For larger protonated proteins, where relaxation losses are more important, the former sequence should be better. For smaller proteins, and probably also for disordered proteins, the new sequence will be preferred.

The two carbon-edited  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments proposed identify the amino acid residue type of two sequential residues,  $i$  and  $i - 1$ , and thus a sequential pair of amino acid residues. With 4 types of sequential residues and 4 types of the NH's own residue, 16 different sequential pair types can occur in deuterated proteins. However, not all these sequential pair types will necessarily be present in a particular protein, and more importantly, some of the pairs will appear only once. So, the ubiquitin sequence contains 4 unique sequential pairs out of the 14 types present. Therefore, the experiments allow the assignment of some of the  $^1\text{H}$ - $^{15}\text{N}$  correlation peaks without using any additional information. Furthermore, the number of possible assignments of the remaining peaks is drastically reduced, since many pair types appear only a limited number of times in the protein sequence. In the case of protonated proteins, the experiments provide more complete information, since there are 36 possible sequential pair types. So, for ubiquitin, 9  $^1\text{H}$ - $^{15}\text{N}$  correlation peaks are assigned unambiguously. The experiments also reduce notably the number of possible sequential assignments, since two NH's,  $i$  and  $j$ , can correspond to consecutive residues in the protein sequence,  $k$  and  $k + 1$ , only if they meet the following two conditions: the intraresidual amino acid type of  $i$  and the sequential amino acid type of  $j$  must coincide; the triplet “*sequential type of i*”–“*intraresidual type of i*”–“*intraresidual type of j*” must be compatible with the protein sequence. An example of using these conditions to reduce ambiguities in the sequential assignment of ubiquitin using only the matching of  $^{13}\text{C}^\alpha$  resonances is given in the Supplementary Material. Finally, the experiments facilitate the transfer of the assignment of a protein to the spectra obtained in different experimental conditions (temperature, pH, presence of ligands, etc.).

The experiments reported here have been performed as 2D, but they can also be implemented as 3D, by including an additional  $^{13}\text{C}$  evolution period to resolve overlapping  $^1\text{H}$ - $^{15}\text{N}$  correlations in a third spectral dimension. The third dimension can correspond either to  $^{13}\text{CO}$ ,  $^{13}\text{C}^\alpha$  or  $^{13}\text{C}^\beta$ . In the latter two cases not only the overlapping signals can be separated, but in addition the intraresidual and sequential

$^{13}\text{C}^\alpha$  or  $^{13}\text{C}^\beta$  chemical shifts are obtained, thus allowing the sequential assignment of the protein without resorting to other experiments. However, we have to remark here that the  $\beta$ -carbon edited experiments are less sensitive than their respective standard triple resonance experiments. Therefore, the general strategy for the assignment should be the combination of the 2D  $\beta$ -carbon edited experiments with the standard 3D HNC experiments. The amendments required to transform the  $\beta$ -carbon edited pulse sequences into 3D, and spectra of ubiquitin obtained using  $^{13}\text{C}^\alpha$  in the third dimension are given in the Supplementary Material.

The sequential experiment here proposed has some similarities with the one proposed by Dötsch et al. (1996c) to select  $^1\text{H}$ - $^{15}\text{N}$  HSQC peaks of residues following Ser, Ala, Cys and Gly. Both experiments are based on the HNCOCACB pulse sequence. In the experiment of Dötsch, the selection of amino acid residues without a  $\gamma$  carbon is based on the evolution of the  $^{13}\text{C}^\beta$ - $^{13}\text{C}^\gamma$  coupling during a time  $1/(2J_{C\beta,C\gamma})$ . Only coherences of residues lacking this coupling, Ala, Cys, Ser and Gly, end in observable magnetization. Deviations of the coupling constant from the value used for the evolution time result in breakthrough peaks which, for flexible residues, can be similar in intensity to real peaks, thus leading to a misclassification of the amino acid residue type. Our experiment, however, is based on the evolution of a single quantum coherence of the  $\text{C}^\beta$  spin with the aliphatic carbon–aliphatic carbon coupling constants during a time  $1/J_{C\beta,C\gamma}$ . Coherences of all amino acid residue types end into observable magnetization, with a sign that depends on the parity of the number of aliphatic  $\text{C}^\gamma$  carbons. Deviations of the coupling constant from the value used for the evolution time do not result in breakthrough peaks, but in a small intensity loss. This procedure is combined with a Hadamard encoding of the evolution with the  $J_{C\beta,Caro}$  and the  $J_{C\beta,CO}$  coupling constants, which allows to classify the  $^1\text{H}$ - $^{15}\text{N}$  correlations into 4 types. Some breakthrough between the subspectra obtained by the Hadamard decoding can appear, owing to variations in the coupling constant values and to pulse imperfections. However, the intensity of these artifacts is always much smaller than the intensity of the correlation peak detected in the correct subspectrum and, therefore, does not lead to classification errors. Thus, our experiment presents two advantages over the experiment of Dötsch: it allows the classification of all  $^1\text{H}$ - $^{15}\text{N}$  correlations; the presence of breakthrough peaks does not produce errors in the amino acid residue type classification.

In summary, we have introduced two new pulse sequences to obtain information on the amino acid type of the residue preceding the observed backbone NH and of its own residue. These pulse sequences are applicable to both protonated and deuterated proteins, and therefore larger proteins can be studied compared to previous experiments. Together, these two experiments drastically reduce the



ambiguities in the sequential assignment of the individual amide groups and therefore of their associated spin systems. Although normal versions of the experiments have been presented, inclusion of TROSY is straightforward and would improve the performance of the experiments in the case of large deuterated proteins. Pulse sequences in Bruker language, and programs to perform the linear combinations to separate the subspectra can be obtained from the web page <http://rmn.iqfr.csic.es>.

**Acknowledgments** This work was supported by projects CTQ2008-00080 and CTQ2011-22514 from the Spanish *Ministerio de Ciencia e Innovación*.

## References

- Atreya HS, Chary KVR (2001) Selective ‘unlabeling’ of amino acids in fractionally C-13 labeled proteins: an approach for stereospecific NMR assignments of CH<sub>3</sub> groups in Val and Leu residues. *J Biomol NMR* 19:267–272
- Atreya HS, Sahu SC, Chary KVR, Govil G (2000) A tracked approach for automated NMR assignments in proteins (TATAPRO). *J Biomol NMR* 17:125–136
- Barnwal RP, Rout AK, Atreya HS, Chary KVR (2008) Identification of C-terminal neighbors of amino acid residues without an aliphatic <sup>13</sup>C<sub>γ</sub> as an aid to NMR assignments in proteins. *J Biomol NMR* 41:191–197
- 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
- Dötsch V, Wagner G (1996) Editing for amino-acid type in CBCACONH experiments based on the <sup>13</sup>Cβ–<sup>13</sup>C<sub>γ</sub> coupling. *J Magn Reson B* 111:310–313
- Dötsch V, Oswald RE, Wagner G (1996a) Amino-acid-type-selective triple-resonance experiments. *J Magn Reson B* 110:107–111
- Dötsch V, Oswald RE, Wagner G (1996b) Selective identification of threonine, valine, and isoleucine sequential connectivities with a TVI-CBCACONH experiment. *J Magn Reson B* 110:04–308
- Dötsch V, Matsuo H, Wagner G (1996c) Amino-acid-type identification for deuterated proteins with a β-carbon-edited HNCOCACB experiment. *J Magn Reson B* 112:95–100
- Feng W, Rios CB, Montelione GT (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
- Feuerstein S, Plevin MJ, Willbold D, Brutscher B (2012) iHADA-MAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. *J Magn Reson* 214:329–334
- Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly <sup>13</sup>C/<sup>15</sup>N-enriched proteins. *J Biomol NMR* 3:185–204
- Johnson BA, Blevins RA (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614
- Krishnarajuna B, Jaipuria G, Thakur S, D’Silva P, Atreya HS (2011) Amino acid selective unlabeled for sequence specific resonance assignments in proteins. *J Biomol NMR* 49:39–51
- Lee KM, Androphy EJ, Baleja JD (1995) A novel method for selective isotope labeling of bacterially expressed proteins. *J Biomol NMR* 5:93–96
- Lescop E, Brutscher B (2009) Highly automated protein backbone resonance assignment within a few hours: the BATCH strategy and software package. *J Biomol NMR* 44:43–57
- Lescop E, Rasia R, Brutscher B (2008) Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J Am Chem Soc* 130:5014–5015
- Löhr F, Pérez C, Köhler R, Rüterjans H, Schmidt JM (2000) Heteronuclear relayed E.COSY revisited: determination of 3 J(Hz, C<sub>γ</sub>) couplings in Asx and aromatic residues in proteins. *J Biomol NMR* 18:13–22
- McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of <sup>15</sup>N and <sup>13</sup>C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* 23:1–38
- Muchmore DD, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol* 177:44–73
- Muhandiram DR, Johnson PE, Yang D, Zhang O, McIntosh LP, Kay LE (1997) Specific <sup>15</sup>N, NH correlations for residues in <sup>15</sup>N, <sup>13</sup>C and fractionally deuterated proteins that immediately follow methyl-containing amino acids. *J Biomol NMR* 10:283–288
- Nietlispach D, Ito Y, Laue E (2002) A novel approach for the sequential backbone assignment of larger proteins: selective intra-HNCA and DQ-HNCA. *J Am Chem Soc* 124:11199–11207
- Ohki S, Kainosho M (2008) Stable isotope labeling methods for protein NMR. *Prog Nucl Magn Reson Spectrosc* 53:208–226
- Pantoja-Uceda D, Santoro J (2008) Amino acid type identification in NMR spectra of proteins via beta- and gamma-carbon edited experiments. *J Magn Reson* 195:187–195
- Rios CB, Feng W, Tashiro M, Shang Z, Montelione GT (1996) Phase labeling of C–H and C–C spin-system topologies: application in constant-time PFG-CBCA(CO)NH experiments for discriminating amino acid spin-system types. *J Biomol NMR* 8:345–350
- Schubert M, Smalla M, Schmieder P, Oschkinat H (1999) MUSIC in triple-resonance experiments: amino acid type-selective H-1-N-15 correlations. *J Magn Reson* 141:34–43
- Schubert M, Oschkinat H, Schmieder P (2001a) MUSIC, selective pulses, and tuned delays: amino acid type-selective H-1-N-15 correlations, II. *J Magn Reson* 148:61–72
- Schubert M, Oschkinat H, Schmieder P (2001b) MUSIC and aromatic residues: amino acid type-selective H-1-N-15 correlations, III. *J Magn Reson* 153:186–192
- Schubert M, Oschkinat H, Schmieder P (2001c) Amino acid type-selective backbone 1H–15N-correlations for Arg and Lys. *J Biomol NMR* 20:379–384
- Schubert M, Labudde D, Leitner D, Oschkinat H, Schmieder P (2005) A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments. *J Biomol NMR* 31:115–127
- Tong KI, Yamamoto M, Tanaka T (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in *E. coli*. *J Biomol NMR* 42:59–67
- Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* 125:13868–13878
- Wittekind M, Müller L (1993) HNCACB a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson B* 101:201–205
- Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994) A suite of triple resonance NMR experiments for the backbone assignment of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H labeled proteins with high sensitivity. *J Am Chem Soc* 116:11655–11666