J Biomol NMR (2015) 63:389–402
DOI 10.1007/s10858-015-9998-4

ARTICLE

CH_3-specific NMR assignment of alanine, isoleucine, leucine and valine methyl groups in high molecular weight proteins using a single sample

Rime Kerfah[1,2,3] · Olivier Hamelin[4] · Jérôme Boisbouvier[1,2,3] · Dominique Marion[1,2,3,5]

Received: 24 September 2015 / Accepted: 20 October 2015 / Published online: 13 November 2015

Abstract A new strategy for the NMR assignment of aliphatic side-chains in large perdeuterated proteins is proposed. It involves an alternative isotopic labeling protocol, the use of an out-and-back ^{13}C–^{13}C TOCSY experiment ((H)C-TOCSY-C-TOCSY-(C)H) and an optimized non-uniform sampling protocol. It has long been known that the non-linearity of an aliphatic spin-system (for example Ile, Val, or Leu) substantially compromises the efficiency of the TOCSY transfers. To permit the use of this efficient pulse scheme, a series of optimized precursors were designed to yield linear ^{13}C perdeuterated side-chains with a single protonated CH_3 group in these three residues. These precursors were added to the culture medium for incorporation into expressed proteins. For Val and Leu residues, the topologically different spin-systems introduced for the *pro-R* and *pro-S* methyl groups enable stereospecific assignment. All CH_3 can be simultaneously assigned on a single sample using a TOCSY experiment. It only requires the tuning of a mixing delay and is thus more versatile than the relayed COSY experiment. Enhanced resolution and sensi-tivity can be achieved by non-uniform sampling combined with the removal of the large J_{CC} coupling by deconvolution prior to the processing by iterative soft thresholding. This strategy has been used on malate synthase G where a large percentage of the CH_3 groups could be correlated directly up to the backbone Ca. It is anticipated that this robust combined strategy can be routinely applied to large proteins.

Keywords NMR spectroscopy · Proteins · Isotopic labeling · Stereospecific assignment · Aliphatic side-chains

Introduction

Resonance assignment is a prerequisite for most NMR studies of biomacromolecules (Wüthrich et al. 1982). For proteins, the assignment of the backbone resonances is generally the easiest step while the more tedious assignment of side-chains is rarely completed. To provide the link to the backbone, J-correlated experiments are preferred to the ambiguous through-space connectivities via the nuclear Overhauser effect (nOe).

From a historical perspective, it is of interest to note that the design of J_{HH}-based NMR experiments for side-chain assignment has evolved over the years: COSY spectra were first replaced by relayed-COSY (Eich et al. 1982) and ultimately by TOCSY (Hartman–Hahn polarization transfer) experiments (Braunschweiler and Ernst 1983; Davis and Bax 1985). Despite the fact that the last two experiments yield similar correlations between non-adjacent spins, the TOCSY pulse scheme, though proposed only a few months later, has superseded the relayed-COSY because it is more efficient (two-times faster coherence transfer) and easier to implement in some ways (no delays

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9998-4) contains supplementary material, which is available to authorized users.

✉ Dominique Marion
Dominique.Marion@ibs.fr

[1] Université Grenoble Alpes, IBS, 38044 Grenoble, France

[2] CNRS, IBS, 38044 Grenoble, France

[3] CEA, IBS, 38044 Grenoble, France

[4] Chemistry and Biology of Metals Laboratory, University Grenoble Alpes, 38027 Grenoble, France

[5] Biomolecular NMR Spectroscopy Group, Institut de Biologie Structurale, 71 Avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9, France

need to be adjusted). Obviously, relaxation remains a limitation that cannot be lifted by purely spectroscopic means.

When uniformly ^{13}C labeled proteins became available, J_{HH}-based correlation experiments were replaced by similar ones involving the carbon–carbon network. The efficiency of the magnetization transfer is improved due to the occurrence of larger (35 Hz) couplings that are moreover less conformation dependent. Similarly to the proton approach, HCCH–COSY or HCCH–TOCSY (Bax et al. 1990) experiments were proposed. In a 3D (H)CCH–TOCSY or HC(C)H–TOCSY an entire set of ^{13}C resonances can be identified and linked together. Eaton et al. (1990) have analyzed the time-dependence of the ^{13}C–^{13}C TOCSY transfer for the various side-chain spin-systems. Under the assumption that all J_{CC} couplings are identical, the transfer function only depends on the coupling topologies. For branched aliphatic side-chains such as Leu, Val or Ile, a sharp decrease in the transfer efficiency was reported. To achieve isotropic mixing in aliphatic side chains, it is necessary to apply strong rf fields (>7 kHz) for several tens of milliseconds to the probe, a stringent requirement now compatible with the most recent cryoprobes.

In larger proteins (typically above 50 kDa), sensitivity and resolution become critical. Alternate labeling strategies have been designed wherein the complete protein is nearly completely deuterated except for some CH_3 groups (Gardner and Kay 1997; Goto et al. 1999). However, the non-linearity of branched residues such as Leu, Val and Ile still compromises the efficiency of any ^{13}C–^{13}C TOCSY based approach. Tugarinov et al. have reported a strategy for linearizing Leu and Val residues (Tugarinov and Kay 2003a) (but not for Ile) and designed a set of COSY-type experiments to correlate the CH_3 resonances to the backbone of these samples (Tugarinov and Kay 2003a, b). The ^{13}C coherence is step-wise transferred from one carbon to its neighbor using suitably designed building blocks and selective pulses are applied to avoid the magnetization leakage in two separate directions at the branch points. In the early design, two distinct pulse sequences for Ile/Leu and for Val were necessary (Tugarinov and Kay 2003a) because of their different topologies. A single experiment (Tugarinov et al. 2013) also reaping benefit from the methyl-TROSY effect was introduced more recently. Using similar COSY-based experiments, the overlapping Val and Leu methyl groups could also be edited on the basis of the chemical shift difference between Val Cα and Leu Cβ (Hu et al. 2012).

For proteins containing diastereotopic groups ($C\beta H_2$, Val- and Leu-CH_3,…) a complete resonance assignment involves stereospecific discrimination (Güntert et al. 2001). Stereospecific assignment can be obtained either by NMR triangulation (Hyberts et al. 1987; Vuister et al. 1993; Goto et al. 1999; Tugarinov and Kay 2004; Tang et al. 2005), computational prediction (Pristovšek and Franzoni 2006), or stereo-specific labeling (Neri et al. 1989; Tugarinov and Kay 2003a, 2004; Plevin et al. 2011; Mas et al. 2013). It is well established that the quality of the NMR-based structures improves dramatically with stereo-specific assignment (Kainosho et al. 2006).

The first labeling method (Neri et al. 1989) uses 10 % of U-[^{13}C] glucose and 90 % of unlabeled glucose as the sole carbon source and leads to isolated Leu-δ_2/Val-γ_2 $^{13}CH_3$ (*pro-S*) and to Leu-δ_1/Val-γ_1 groups (*pro-R*) that are J-coupled to the $^{13}C\gamma$ and $^{13}C\beta$ respectively. This strategy based on the presence or absence of a ^{13}C–^{13}C J-coupling suffers from poor NMR sensitivity owing to the fractional labeling and to the high level of protonation. This low ^{13}C enrichment could be significantly enhanced using the previously described methyl labeled-aceto-acetate (*pro-S*) (Gans et al. 2010) together with ^{13}C-glucose (Plevin et al. 2011). Unfortunately, the need to perform the protein expression in protonated water cannot be avoided. If the protein over-expression were carried out in $^{2}H_2O$, then some protons of the pyruvate arising from the added glucose would be exchanged (Rosen et al. 1996). Moreover, several isotopomers of the Val and Leu methyl groups (i.e. $^{13}CH_3$, $^{13}CH_2^2H$, $^{13}CH^2H_2$, $^{13}C^2H_3$) would coexist and be detrimental to experimental sensitivity and resolution. Thus, this strategy is not fully compatible with the perdeuteration imperative for high molecular weight proteins.

In perdeuterated proteins, protonated and linearized Leu and Val methyl groups can be obtained using α-ketoisovalerate as precursor (Tugarinov and Kay 2003a). However, the obtainable patterns [both $^{13}CH_3$ or a racemic mixture of (LV)-*pro-R* and (LV)-*pro-S*] retain no stereospecific information. In contrast, the use of methyl-labeled acetolactate (Gans et al. 2010) leads to straightforward prochiral differentiation.

Although this strategy increases sensitivity as compared to fractional labeling approaches proposed earlier (Neri et al. 1989), some practical limitations remain. Both *pro-R*- and *pro-S*-moieties have the same spin topology and thus, if mixed, the stereoselective information gets lost. Hence, duplicate experiments on two samples are necessary.

An alternate approach for assigning stereospecifically branched aliphatic residues has been proposed recently by Mas et al. (2013): it aims at producing two labeling patterns that can be distinguished spectro-scopically. For this approach, two samples are thus produced as follows:

- Sample (1) is labeled with Val-[2,3-$^{2}H_2$; 1,2,3-$^{13}C_3$;-[$^{13}C^1H_3$] *pro-R*/[$^{12}C^2H_3$] *pro-S*] and identified as Val-γ_1 or as *linear labeling*.
- Sample (2) is labeled with Val-[2,3-$^{2}H_2$; 3–^{13}C; [$^{13}C^2H_3$] *pro-R*/[$^{13}C^1H_3$] *pro-S*] and identified as Val-γ_2 or as *V-shaped labeling*.

The Val-γ_1 methyl group is linked to the backbone resonances using sample (1) and subsequently the Val-γ_2 CH_3 is connected to Val-γ_1 CH_3 using sample (2).

In this communication we explore an alternate strategy to achieve the full and stereospecific assignment for Ile, Val, Leu and Ala using a single sample with the minimal number of experiments. The ^{13}C, ^{2}H labeling strategy for Ile, Val, Leu and Ala side-chains should meet the following requirements: (a) The entire side-chains are deuterated with the exception of a single protonated CH_3 group; (b) A linear ^{13}C labeling arrangement is required to permit an efficient ^{13}C–^{13}C TOCSY transfer from the methyl group to the backbone; and (c) In the case of Ile, Leu and Val, the spin-systems containing each of the two CH_3 have to be topologically or spectroscopically different. For both Ile-δ_1 and -γ_2 CH_3, a linear chain to the backbone can be selected. In the case of Val and Leu, the discrimination of *pro-R* and *pro-S* CH_3 is achieved by connecting the Leu–δ_1/Val–γ_1 methyl group to the backbone (*linear labeling*) while the Leu-δ_2/Val-γ_2 CH_3 is connected to the *pro-R methyl* directly and not via the backbone (this labeling will be later referred to as the *V-shaped labeling*) (Mas et al. 2013). The proposed methodology has been applied on malate synthase G (MSG), a 723-residue monomeric enzyme with molecular weight of 81 kDa (Tugarinov et al. 2005).

Materials and methods

Precursor preparation

Precursor for the linearized Ile-δ_1 [$^{13}CH_3$] labeling

A precursor for Ile, 2-hydroxy-2-ethyl-3-keto-butanoic acid, can be obtained from generic small molecules enzymatically, i.e. using the aceto-hydroxy-acid synthase II (AHAS II) from *E. coli*. Its overexpression and purification have been carried out according to the previously published protocol (Hill et al. 1997; Chipman et al. 1998). As reported elsewhere (Kerfah et al. 2014), AHAS II was used to condense a molecule of 1,2,3,4-[$^{13}C_4$]-3,3-[$^{2}H_2$]-2-ketobutyrate with 2-[^{13}C]-3,3,3-[$^{2}H_3$]pyruvate. The product of this reaction is the 2-hydroxy-2-(1′,1′-[$^{2}H_2$], 1′,2′-[$^{13}C_2$])ethyl-3-oxo-1,2,3-[$^{13}C_3$]-4,4,4-[$^{2}H_3$]butanoic-acid (compound (3) in Fig. 1), a precursor of Ile. Through this labeled precursor, the $^{13}CH_3$ in the Ile δ_1 position can be correlated to the backbone via a linear ^{13}C spin system since the Ile-γ_2 methyl group remains ^{12}C and deuterated. The reaction was carried out by adding the AHAS II at 6 mM (420 ng/mL) in an equimolar mixture (33 mM each) of deuterated pyruvate and ^{13}C, ^{2}H labeled 2-keto-butyrate in a $^{2}H_2O$ buffer of potassium phosphate 50 mM at pH 7.8 (uncorrected), $MgCl_2$ 10 mM, Thiamine diphosphate 1 mM, FAD 20 μM. Perdeuteration of pyruvate was achieved by treatment of unlabeled pyruvate in $^{2}H_2O$ at pH 10.7 during 72 h. Condensation was monitored by 1D NMR. Labeled initial chemicals were purchased from Sigma-Aldrich.

Precursors for the linearized Ala-β, Ile-γ_2, Leu-δ_1/Val-γ_1 and V-shaped Leu-δ_2/Val-γ_2 labelings

These compounds were obtained from NMR-Bio (www.nmr-bio.com). The chemical synthesis of 2-hydroxy-2-(1′,1′,2′,2′,2′-[$^{2}H_5$]) ethyl -3-oxo-(1,2,3,4-[$^{13}C_4$])butanoic acid (linearized Ile-γ_2 precursor, compound (4) in Fig. 1) has been described by Ayala et al. (2012). 2-hydroxy-2-[$^{2}H_3$]methyl-3-keto-1,2,3,4-[$^{13}C_4$]butanoic acid (linearized Leu-δ_1/Val-γ_1 precursor) and 2-hydroxy-2-[^{13}C]methyl-3-keto -3,4-[$^{13}C_2$]- 4,4,4-[$^{2}H_3$]butanoic acid (V-shaped Leu-δ_2/Val-γ_2) have been synthesized as described by Gans et al. (Gans et al. 2010).

Protein production and purification

MSG was expressed and purified as described in Godoy-Ruiz et al. (2010). However, to achieve the desired AILV labeling of MSG, a first solution containing an equimolar concentration (2 × 150 mg/L) of compounds (1) and (2) (Leu and Val precursor, c.f. Figure 1a) was added to the U-[^{13}C,^{15}N, ^{2}H] culture 1 h before the induction of MSG expression. In order to optimize the co-incorporation of the ensemble of used compounds, a second mixture (30 mg/L of compound (3) and 60 mg/L of (4) Ile precursors, cf Fig. 1a) was added only 20 min before induction together with (800 mg/L) U-[^{13}C], 2-[^{2}H] Ala. As compound (4) was obtained by chemical synthesis, a racemic mixture (R + S) at twice the concentration is added but only the S diasteroisomer is metabolized. The final labeling pattern for ILV is described in Fig. 1b. The analyzed MSG samples were prepared in $^{2}H_2O$ buffer containing 25 mM MES (pH 7.0 uncorrected), 20 mM $MgCl_2$ and 5 mM DTT. Two samples were used at 400 mM (3 mm Shigemi tube, 80 mL) and 1 mM (5 mm Shigemi tube, 300 mL). The additional costs of 1 L of culture with the precursors for AILV are typically less than 1 k€. An ILV-labeled sample of ubiquitin (1.44 mM, 20 mM Tris, 20 mM NaCl, pH 7.4) was also used to optimize the pulse sequences and evaluate their relative sensitivity.

Nuclear magnetic resonance

The NMR data were recorded at 37 °C on spectrometers operating at ^{1}H frequencies of 600 MHz (Agilent) or 700 MHz (Bruker) both equipped with cryogenic probes. The pulse sequence was adapted from the HCCH–TOCSY

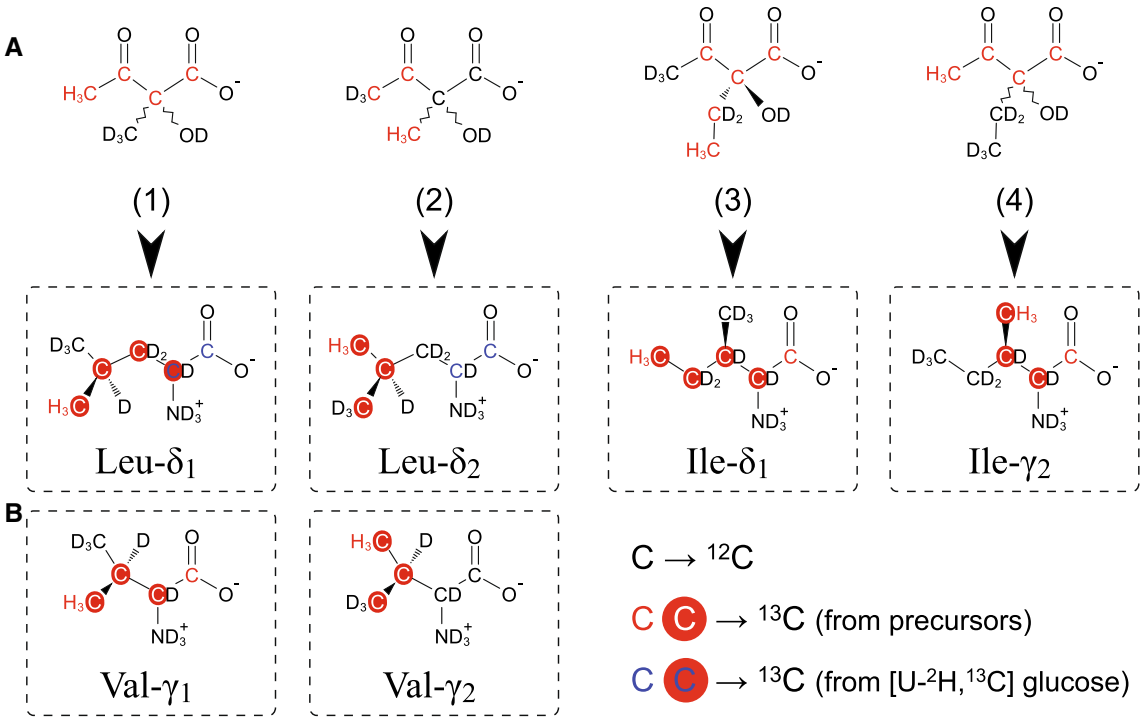

Fig. 1 Precursors and resulting residues for the labeling of Leu, Val and Ile side-chains of MSG. In (**a**) the 2D structure of the used precursors are depicted: (*1*) 2 hydroxy- 2-[$^{2}H_3$]methyl-3-keto-1,2,3,4-[$^{13}C_4$]butanoic acid, (*2*) 2-hydroxy- 2-[^{13}C]methyl- 3-keto -3,4-[$^{13}C_2$]- 4,4,4-[$^{2}H_3$]butanoic acid, (*3*) (S)-2-hydroxy- 2-(1′,1′-[$^{2}H_2$], 1′,2′-[$^{13}C_2$])ethyl- 3-oxo-1,2,3-[$^{13}C_3$]- 4,4,4-[$^{2}H_3$]butanoic acid and (*4*) 2-hydroxy- 2-(1′,1′, 2′,2′,2′-[$^{2}H_5$])ethyl- 3-keto -1,2,3,4-[$^{13}C_4$]butanoic acid. All precursors except (3) were obtained by chemical synthesis and are therefore provided as racemic mixtures. Protocols for the individual incorporation of compounds (1) and (2) have been reported by Gans et al. (Gans et al. 2010) and for compound (3) and (4) by Ayala et al. (2012). In (**b**), the isotopic labeling patterns obtained for ILV residues are depicted. In the figure, these patterns are named according to the $^{13}CH_3$ moiety present in the residue. The following alternate nomenclature has been used in the litterature (Kerfah et al. 2014): Val-γ_1 = Val-*pro-R*, Val-γ_2 = Val-*pro-S* = Val-*pro-V* (Mas et al. 2013), Leu-δ_1 = Leu-*pro-R* and Leu-δ_2 = Leu-*pro-S* = Leu-*pro-V*. ^{12}C carbons are in black, ^{13}C carbons are in red when the labeling stems from the precursors and in blue when coming from the [U–^{2}H,^{13}C] glucose added to the culture. The filled red circles identify the 3- or 4-spin systems along which the magnetization is transferred in the out-and-back HC–CH TOCSY pulse sequence (cf Fig. 2). Linear 4-spin chains are obtained for Leu-δ_1 and Ile-δ_1 and 3-spin chains for the four other residues. Note that except for Val-γ_2, the carbonyl carbons are also ^{13}C labeled. Their decoupling may improve the spectral resolution of the adjacent Cα

(Bax et al. 1990). On deuterated samples with protonated CH_3, an "out-and-back" design is required for sensitivity reasons. Starting from the CH_3 moiety, the magnetization migrates along the side-chain (possibly up to the Cα), is then labeled during t_1 and migrates back to the CH_3 according to the following scheme:

$$H \rightarrow C(t_1) \rightarrow \text{TOCSY} \rightarrow C(t_2) \rightarrow \text{TOCSY} \rightarrow C \rightarrow H(t_3).$$

The (H)C-TOCSY-C-TOCSY-(C)H (abbreviated later on as "HC–CH TOCSY") pulse sequence is depicted in Fig. 2a. The 3D data set was acquired using non-uniform sampling (NUS) along the two indirect dimensions: the t_1 and t_2 dimensions contained 140 and 150 complex data points, respectively, of which only 4260 (t_1, t_2) point pairs were recorded, corresponding to 20 % undersampling. The slower transverse relaxation of the mobile methyl groups as compared to the rest of the side-chain justifies the choice of $t_1^{max} = 35.88$ ms larger than $t_2^{max} = 12.82$ ms. A relaxation delay of 1.5 s was used along with 8 scans/FID, giving rise to a net acquisition time of 65 h. The sampling schedule was generated with *ScheduleTool*, a Java program written by Jeff Hoch and colleagues at the University of Connecticut (USA) (see caption to Fig. S1).

For a quantitative comparison of the performance of HC–CH TOCSY, the relayed COSY approach described by Tugarinov and Kay (2003a) was implemented and adapted as described in Fig. 2b. As our sample contains only linearized spin systems, several selective pulses, required to direct the coherence flow in branched systems, can be omitted. In contrast to the HC–CH TOCSY sequence, the t_1 evolution period can be merged with the first C–C transfer delay using a constant-time evolution on a multiple-quantum coherence. Details of the pulse sequences are given in the caption to Fig. 2.

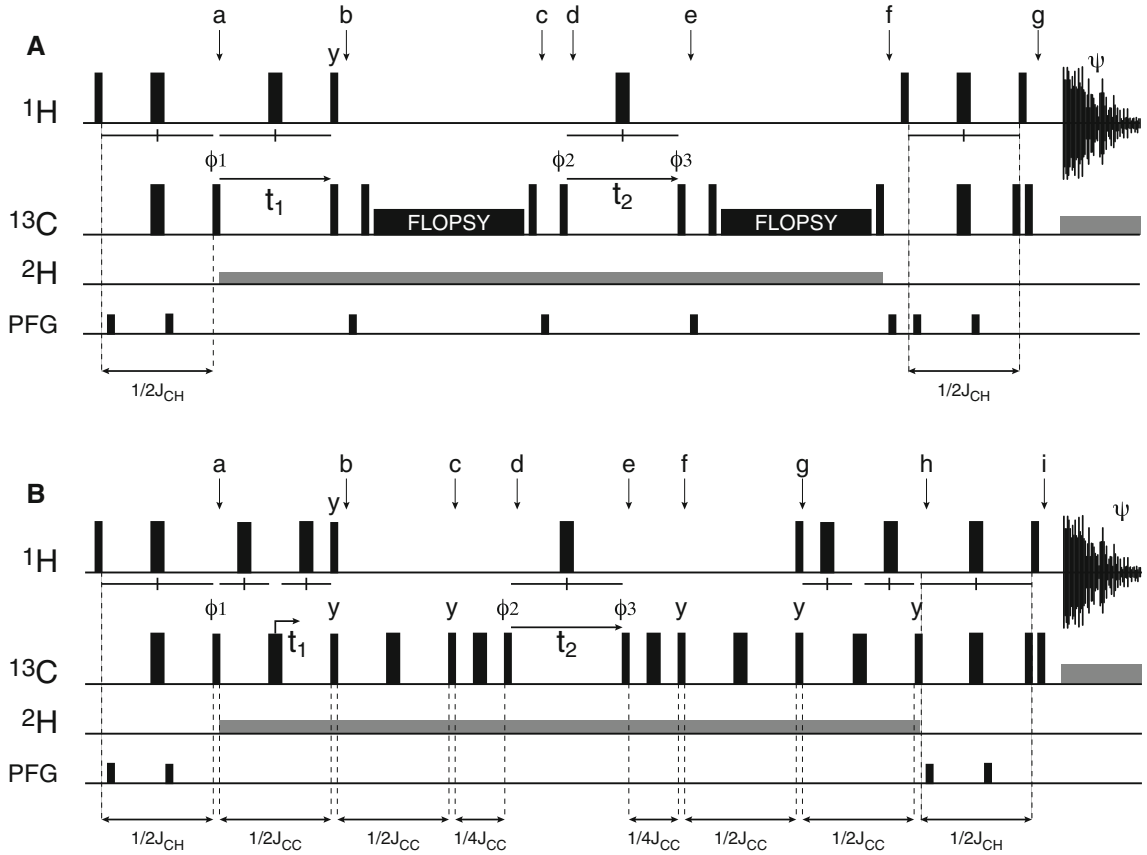

Fig. 2 **a** (H)C-TOCSY-C-TOCSY-(C)H (or "HC–CH TOCSY") pulse sequence correlating the ^{13}C of the aliphatic side-chains Ala, Ile, Leu and Val. All narrow and wide rectangular pulses are applied with flip angles of 90° and 180°, respectively, along the x-axis, unless indicated otherwise. The experiments were carried out at a proton frequency of 700 MHz unless otherwise stated. The magnetization starts on ^{1}H of the CH_3 groups and the signal is detected on the same protons. The respective spectral widths were SW_1 (^{13}C CH_3) = 22 - ppm, SW_2 (^{13}C aliph) = 66 ppm and SW_3 (^{1}H) = 12 ppm and the ^{1}H and ^{13}C carriers are positionned at 4.7 and 40 ppm. Methyl carbon chemical shifts are encoded during t_1 as a multiple-quantum coherence (Tugarinov et al. 2003), and the aliphatic carbon (Cδ, Cγ, Cβ and Cα) shifts during t_2 as a single-quantum coherence with ^{1}H and ^{2}H decoupling. ^{2}H decoupling is necessary during t_2 but can optionally be applied during the two isotropic mixing periods. ^{13}C hard pulses are applied with $\omega_1/2\pi = 16$ kHz, carbon homonuclear TOCSY mixing is using FLOPSY (Kadkhodaie et al. 1991) or DIPSI-3 (Shaka et al. 1988) mixing schemes with $\omega_1/2\pi = 8$ kHz and ^{13}C decoupling during acquisition GARP ($\omega_1/2\pi = 3$ kHz). The length of the two isotropic mixing periods ($b \rightarrow c$ and $e \rightarrow f$) is either 14.6 and 19.4 ms (experiments using FLOPSY on Bruker) or 15.5 ms and 23.3 ms (experiments using DIPSI-3 on Varian-Agilent). The purging of isotropic mixing (elimination of the orthogonal components) can be performed in two ways: either by z-filter (field gradients along z at 30–40 G/cm for 1 ms) as shown in this figure or by trim pulse (relying on B_1 inhomogeneity in the transverse plane). The following phase cycling aiming at the suppression of axial peaks has been used: φ1 = x, −x, φ2 = y, y, −y, −y, φ3 = 4(y), 4(−y), rec(ψ) = x, −x, −x, x, −x, x, x, −x. Quadrature detection along t_1 and t_2 is achieved via States-TPPI of phases φ1 and φ2 respectively. **b** "Out-and-back" relayed COSY pulse sequence adapted from Tugarinov and Kay (2003a). The beginning (before a) and the end of the pulse sequence (after h) as well as the aliphatic carbon evolution (t_2) are identical to HC–CH TOCSY sequence. Methyl carbon chemical shifts are encoded as a multiple-quantum coherence between points *a* and *b* using either a constant-time or a semi-constant time scheme. The magnetization is transferred in 3– and 4–spins systems during 3 successive periods [2 periods of $1/2J_{CC}$ ($a \rightarrow b$ and $b \rightarrow c$) and one of $1/4J_{CC}$ ($c \rightarrow d$)] and back again in the inverse order [($e \rightarrow f$) ($f \rightarrow g$) ($g \rightarrow h$)]. All shaped pulses present in the original sequence of Tugarinov and Kay (2003a) are replaced by hard pulses, owing to the linearization of the side-chains (see text). This sequence can be adapted to transfer coherence over smaller spin topologies by removing pairs of transfer periods: for instance the [$b \rightarrow c$] and [$f \rightarrow g$] blocks or the [$c \rightarrow d$] and [$e \rightarrow f$] ones. The phase cycling is identical to the HC–CH TOCSY sequence

Data processing

The spectra were transformed using MddNMR 2.2, a reconstruction software using multi-dimensional decomposition (MDD) and compressed sensing (CS) written by Vladislav Orekhov and colleagues at the University of Gothenburg (Sweden). The processing scripts were generated using the qMDD graphical user interface. CS algorithms use a l_1-norm regularization (Stern et al. 2007) to minimize the artefacts due to the non-uniform acquisition and with the iterative soft thresholding (IST) algorithm (Kazimierczuk and Orekhov 2011), convergence could be achieved in fewer

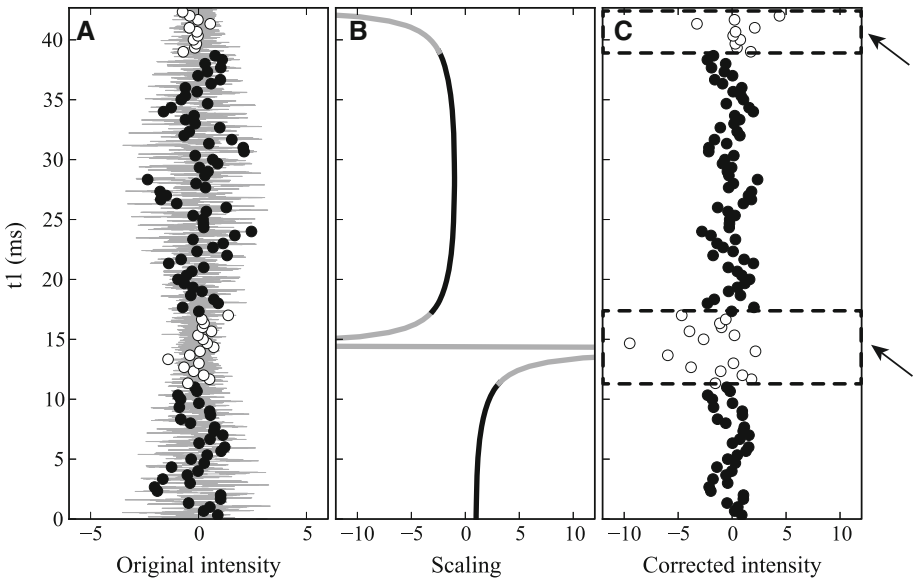

Fig. 3 J-deconvolution method prior to iterative soft thresholding (IST) (Kazimierczuk and Orekhov 2011). For the sake of clarity, the case of a one-dimensional NMR signal is discussed. **a** The analog NMR signal is depicted in light grey. The data points that have been sampled are in black and the non-sampled points in white. **b** The function 1.0/cos ($\pi \times J_{CC}t_1$) is shown but only the black portion of the curve, which corresponds to the black data points in (**a**) is actually used. Note that this function is not defined for $t_1^a = 1/(2 \times J_{CC}) = 14.3$ ms. Furthermore, it changes sign and exhibits large numerical values in the vicinity of t_1^a. **c** J-deconvolution of the experimental signal by multiplication of the data points in (**a**) by the curve in (**b**). The resulting black points are more uniform and do not display the J-modulation visible in (**a**)

than 500 iterations. NMRPipe (Delaglio et al. 1995) was then used to phase correct along all dimensions after the CS reconstruction. The processing was carried out on a MacBook Pro running under MacOS 10.8.

The J_{CC} coupling that remains active during the t_1 evolution period was removed by J-deconvolution during the processing stage (cf. Fig. 3) (Shimba et al. 2003; Scholz et al. 2007; Marion 2010). The signal evolution of a ^{13}C methyl group coupled to a single neighbor (Fig. 3a) will be modulated as cos ($\pi \times J_{CC}t_1$) with $J_{CC} \approx 35$ Hz. Only the data points in black have been measured. The white data points with much lower signal-to-noise ratios were omitted. If the time-domain signal is multiplied by the inverse function, i.e. 1/cos ($\pi \times J_{CC}t_1$) (shown in Fig. 3b) before the IST processing, the corresponding J-splitting is eliminated (cf Fig. 3c). If uniform sampling would have been used, the white data points which contain mainly experimental noise would have been scaled up more than the relevant black data points. This unstable behavior would introduce too much noise in the resulting frequency-domain spectrum.

Results and discussion

The objective of the current work is to achieve full stereospecific assignment of all AILV side-chains with a single sample in the shortest possible experimental time. The availability of linearized ^{13}C AILV side-chains will allow us to replace a residue-specific COSY-based experiments (Tugarinov and Kay 2003a; b; Hu et al. 2012) with a TOCSY-based pulse scheme. This experiment is expected to perform well for all residues without much optimization.

Isotopic labeling of Ile and Ala

As branched ^{13}C spin systems compromise the efficiency of isotropic mixing sequences in TOCSY experiments (Eaton et al. 1990), linearized ^{13}C-labeled AILV side chains have to be obtained from linearized precursors or labeled amino-acids (cf. Fig. 1). While such compounds have been already described for Ile-γ_2 (Ayala et al. 2012) Val and Leu (Tugarinov and Kay 2003a; Gans et al. 2010; Mas et al. 2013), analogs for Ile-δ_1 are not yet available. Here we use the previously developed 2-hydroxy-2-ethyl-3-keto-butanoic acid and acetolactate molecules to regio- and stereospecifically link the Ile-γ_2 (Ayala et al. 2012), the Leu-δ_1 and Val-γ_1 groups to the backbone (Gans et al. 2010). We used an enzymatic synthesis protocol (Kerfah et al. 2014), to obtain compound (3) (cf. Fig. 1, 2-hydroxy- 2-(1′,1′-[$^2H_2$], 1′,2′-[$^{13}C_2$])ethyl- 3-oxo-1,2,3-[$^{13}C_3$]- 4,4,4-[$^2H_3$] butanoic acid) that yields the Ile–δ_1 ^{13}C side-chain linearization: the δ_1-CH_3 in Ile is linked to the backbone through a linear ^{13}C spin system while the γ_2-CH_3 remains unlabeled. U-[^{13}C], 2-[^2H] Ala was obtained from NMR-Bio.

Stereospecific labeling schemes for the Leu and Val assignment

The strategy described above for Ile cannot be applied without any modification to Val and Leu. In both Val and Leu residues, the two CH_3 groups (Cγ and Cδ respectively) are topologically equivalent in the absence of symmetry-breaking isotopic labeling. This concept was introduced a long time ago by Senn et al. (1989) who have used 10 % of [$^{13}C_6$]-D-glucose and 90 % of unlabeled D-glucose: the Val $C\gamma_1$ which originates from the same C–C fragment as the Cβ will exhibit a J_{CC} coupling but the $C\gamma_2$ will not. The price paid in terms of sensitivity for a 10 % ^{13}C-labeled (and protonated) sample is not compatible with large proteins.

In the present work, we opted instead for stereo-specific labeling of Leu and Val in the presence of a deuterated background as recently proposed (Mas et al. 2013). For the key precursor of both Val and Leu aceto-lactate (also known as 2-hydroxy-2-methyl-3-ketobutanoic acid), the two different labeling patterns will be:

- 2 hydroxy- 2-[$^2H_3$]methyl-3-keto- 1,2,3,4-[$^{13}C_4$]butanoic acid [precursor (1) in Fig. 1], a *linear labeling* is obtained from the Leu-δ_1/Val-γ_1 down to the backbone $^{13}C\alpha$ and C′.
- 2-hydroxy- 2-[^{13}C]methyl-3-keto -3,4-[$^{13}C_2$]- 4,4,4-[$^2H_3$] butanoic acid [precursor (2) in Fig. 1], a linear linkage between the Leu-δ_2/Val-γ_2 and the previously assigned Leu-δ_1/Val-γ_1 is generated via the $^{13}C\beta$ (for Val) and $^{13}C\gamma$ (for Leu) (*V-shaped labeling*).

These two molecules [precursors (1) and (2)] which are topo-logically and thus spectroscopically distinguishable can be mixed into a single sample at equimolar ratio and thereby provide first the assignment of the Leu-δ_1/Val-γ_1 and then the linkage of the Leu-δ_2/Val-γ_2 to the Leu-δ_1/Val-γ_1 moiety.

Optimized MSG labeled sample for complete regio- and stereo-specific assignment of AILV

With the aim of producing a single MSG sample, for which complete and stereospecific assignment can be obtained, the following protocol was chosen. An equimolar concentration of precursors (1) and (2) were added to the culture medium to produce the two Leu isotopomers (Leu-δ_1 and Leu-δ_2) and the two Val isotopomers (Val-γ_1 and Val-γ_2). An equimolar concentration of precursors (3) and (4) in S configuration leads to the two Ile isotopomers (Ile-δ_1 and Ile-γ_2). Note that (3) is supplied as a racemic mixture at twice the same concentration. Alanine is introduced as U-[^{13}C], 2-[^{2}H] Ala available commercially.

To ensure that all labeling patterns are present in the final protein at the same incorporation level, two conditions should be fulfilled: the precursors should be added at equivalent concentration in the culture medium and secondly, their incorporation should occur with nearly identical kinetics. The in-cell conversion of the precursors into ILV residues involves a common enzyme, the ketol-acid reductoisomerase (EC 1.1.1.86). It has been reported that this enzyme has a 5- to 8-fold higher activity with Ile precursors than with Val/Leu precursors (Dumas et al. 2001). This differential activity, which could lead to a lower incorporation of labeled Val/Leu can be compensated for by adding the precursor for Val/Leu to the medium 1 h before induction and adding the precursors for Ile only 20 min before induction. The alanine was added in conjunction with the Ile precursors to avoid any isotopic scrambling to the Ile-γ_2 position via [^{13}C-]pyruvate from the labeled Ala (Kerfah et al. 2014).

The MSG thus produced contains two types of labeling for each ILV residue (cf Fig. 1) with an incorporation level greater than 45 % for each labeling pattern (and >95 % of Ala).

«Out-and-back» -TOCSY based pulse sequence

The HC–CH TOCSY pulse sequence used for assigning and correlating the $^{13}CH_3$ group with the backbone is depicted in Fig. 2a. A TOCSY pulse train is used to transfer magnetization along the side-chain from the CH_3 to the Cα carbon.

As the two labeling patterns described earlier contain protons only on the methyl groups, the coherence pathway originates and ends with these protons. An "out-and-back" design is thus required although the return transfer has a cost in terms of sensitivity in large macromolecules. For the linear labeling pattern (Ile–δ_1, Ile–γ_2, Val-γ_1 and Leu-δ_1), the carbonyl nucleus is also ^{13}C labeled (cf Fig. 1), but the magnetization is not transferred to these nuclei because of the large chemical shift difference with the other spins. In other words, the rf pulses are not strong enough to cover efficiently the full ^{13}C chemical shift range.

Isotropic mixing is applied on ^{13}C coherences that are antiphase with respect to protons. This is made possible by the out-and-back design of our pulse sequence. The performance of a ^{13}C–^{13}C isotropic mixing scheme on in-phase (Cx) or anti-phase (CxHz) coherences is identical as the protons remain unaffected. As a by-product of the isotropic mixing scheme (if analyzed as an improved train of 180° ^{13}C pulses) the effect of $^1J_{CH}$ is suppressed. In terms of sensitivity, the only cost is the faster relaxation of anti-phase coherences as compared to in-phase ones, which is marginal due to the slow T_1 relaxation of the methyl ^{1}H. Our sequence has also similarities with the C-TOCSY-CHD2 experiment (Otten et al. 2010) which starts directly from ^{13}C coherences and involves a single one-way isotropic mixing to the methyl groups. Since the overall S/N

of an NMR experiment is proportional to $\gamma_{ex} \times \gamma_{det}^{3/2}$ (the gyromagnetic ratio of the initially excited and of the detected spins), the C-TOCSY-CHD2 has a fourfold reduced sensitivity that could be partially compensated by the single period of isotropic mixing. Furthermore, as it is not possible to fully transfer an in-phase Cx to in-phase Hx in a CH_3 group, these authors have partially deuterated the methyls, selecting only the CHD_2 isotopomer for its favorable detection.

The labeling patterns depicted in Fig. 1 show that most ^{13}C are not isolated but coupled to their neighbor(s) by $^1J_{CC}$ coupling(s). Whether these splittings become visible in the spectra depends depends on how fast the protein spins relax and how far data points are sampled in the indirect dimensions. The J-coupling could be removed either by decoupling (at the acquisition stage) or by J-deconvolution (during data processing, see below). We have not implemented any decoupling in the indirect dimensions because a selective decoupling scheme (Van Melckebeke et al. 2004) able to invert all adjacent carbons for AILV is likely to perturb also some CH_3 resonances in the F_1 dimension (CH_3). Moreover, in the case of MSG, t_2^{max} was short (13 ms) and we did not notice significant changes with and without ^{15}N or $^{13}C'$ decoupling (McCoy and Mueller 1993). For smaller proteins, where longer t_2^{max} can be used, implementing decoupling is likely to improve the resolution.

Theoretical comparison with relayed COSY approaches

Let us now compare the transfer efficiency of the two pulse schemes, first in the absence of relaxation and then on fast relaxing systems. Notably, a method that achieves a larger transfer but in longer time may lose its competitive edge in large proteins with fast transverse relaxation.

In the relayed COSY method, the coherences are transferred between J-coupled spins in the absence of radio-frequency by free J-coupling evolution. In HC–CH TOCSY, the transfer occurs by isotropic mixing in the presence of a strong rf field.

The transfer of coherences from spin I to spin S using a TOCSY pulse sequence is theoretically twice as fast as in a free J-evolution. For TOCSY (Braunschweiler and Ernst 1983), the Hamiltonian is given as:

$$\mathcal{H}_i = 2\pi J_{IS}(IxSx + IySy + IzSz)$$

whereas the two first terms are missing for free J-evolution:

$$\mathcal{H}_f = 2\pi J_{IS} IzSz.$$

In a TOCSY experiment, an in-phase to in-phase coherence transfer (Ix → Sx) requires a time of $1/2J_{IS}$; in contrast, in the absence of rf, an in-phase to anti-phase coherence transfer (Ix → 2IySz) requires the same time and should be

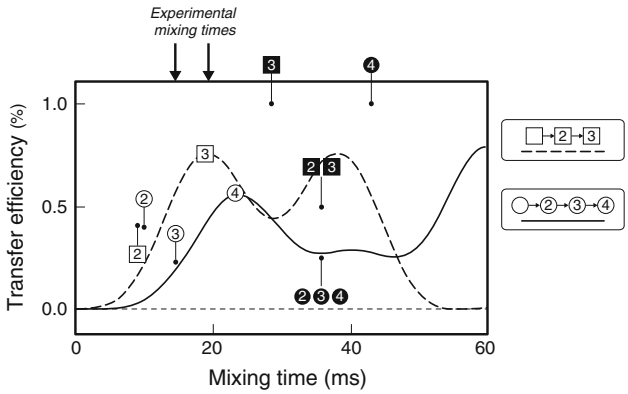

Fig. 4 Theoretical time dependence of coherence transfers in isotropic (TOCSY) mixing and in multiple relayed COSY. Linear spin systems are considered with uniform J-coupling ($^1J_{CC}$ = 35 Hz) and without any relaxation. Our ILV-labeled protein contains linear 3-spin systems (Ile-γ_2, Val-γ_1, Val-γ_2 and Leu-δ_2) identified by *squares* (□) and the *solid line* and linear 4-spin systems (Ile-δ_1, Leu-δ_1) identified by *circles* (○) and the *dotted line*. The curves, adapted from Eaton et al. (1990), show the TOCSY transfer for in-phase coherences (Ix → Sx) from one end of the spin system to the other one, i.e. from spin 1 to spin 3 or 4. Their first local maxima occur respectively at t = 18.3 ms (symbol [3] in a *square*) for 3-spin systems and at t = 23.4 ms for 4-spin systems (symbol (*4*) in a *circle*). The *two arrows* above the plot indicate the experimental value for the short (14.6 ms) and for the long mixing times (19.4 ms) (see Fig. 2 caption). On the same plot, the efficiency of other intermediate transfers has been indicated: the *open symbols* [2], (*2*) and (*3*) correspond to the first maximum of the [1] → [2] transfer in 3-spin systems (occurring at 9.5 ms) and of the (*1*) → (*2*) and (*1*) → (*3*) transfer in 4-spin systems (at 10 and 14.3 ms respectively). Note that, for longer mixing times optimized for the full side-chain transfers, these intermediate transfers are much smaller. On the same figure, the intensity of the correlation peaks in the relayed COSY methods are indicated by *black symbols*: a 100 % [1] → [3] transfer is achieved at 28 ms (2 periods of $1/2J_{CC}$) and a 100 % [1] → [4] transfer is achieved at 42 ms (3 periods of $1/2J_{CC}$). In the pulse sequence in Fig. 2b, a trade-off for both 3- and 4-spin systems is used with 2 periods of $1/2J_{CC}$ and one $1/4J_{CC}$ one (=35 ms) (Tugarinov and Kay 2003a): the cross-peaks for the [1] → [2][3] transfer only reach 50 % and that for the [1] → [2][3][4] transfer only reach 25 %

followed by an identical delay for refocusing (2IzSy → Sx) (see Fig. 4 caption for details). For larger spin-systems (3 or 4), several defocusing/refocusing periods can be merged for the COSY design.

As relaxation is a major factor for large macromolecular systems, it can be anticipated that isotropic mixing would be more efficient for coherence transfer than free J-evolution.

Let us focus on the more demanding transfer from the CH_3 of AILV to the Cα/Cβ in linear 2-, 3- and 4 spin systems with the following pattern:

$C^1H_3- C^2$, $C^1H_3- C^2- C^3$ and $C^1H_3- C^2- C^3- C^4$

To transfer the coherence from C^1 to C^3 or C^4, two options are available: a single TOCSY mixing of suitable length or a series of free J-evolution periods separated by 90° pulses. This latter approach has been used by

Tugarinov and Kay (2003a, b) because it still performs well on branched amino-acids. In Fig. 4, we compare the efficiency of the TOCSY strategy with that of the relayed COSY *in the absence of relaxation*. For 3-spin and linear 4-spin systems, 76 and 55 % of the magnetization reaches the end of the spin system within 18.3 ms and 23.4 ms respectively using TOCSY transfer.

For the relayed COSY approach, the transfer efficiency depends upon the number and length of concatened periods. By concatening 2 or 3 periods tuned to $1/2J_{CC}$, a full transfer over 3 or 4 spins respectively is obtained in 28 or 42 ms respectively (cf Fig. 4). To assign the resonances in high molecular weight proteins, it is suitable to detect more signals than that of the Ca. Thus, the pulse design shown in Fig. 2 uses two periods of $1/2J_{CC}$ length followed by one of $1/4J_{CC}$ (between points a and d) (in the reverse order between points e and h). As discussed by Tugarinov and Kay (2003a), the coherences originating from the methyl groups are spread over 2 or 4 terms for Val and (Ile, Leu) respectively:

$$4H_zC_x^{\beta}C_z^{\alpha} + 2H_zC_y^{\alpha}$$
$$8H_zC_x^{\gamma}C_y^{\beta}C_x^{\alpha} + 4H_zC_x^{\gamma}C_z^{\beta} + 2H_zC_y + 4H_zC_z^{\beta}C_x^{\alpha}$$

After 35 ms, two peaks with 50 % intensity are obtained for Val and three peaks with 25 % for Ile and Leu (the first component is a non-detectable 3-spin coherence).

Because HC–CH TOCSY and relayed COSY are "out-and-back" in nature, comparing the efficiency and duration of transfer in one direction or the other is valid. Relayed COSY performs much better than TOCSY only if one aims at transferring the coherences exclusively from the CH_3 to the Cα; however, a longer time (between 28 and 42 ms) is required. In the other case, where one wishes to obtain more than one correlation peak for assignment, the TOCSY method achieves this goal in less time.

Let us now focus on the case of large proteins, where transverse coherences are rapidly relaxing: in this case, it may be advantageous to opt for a slightly less efficient but faster technique. Investigating theoretically the effect of relaxation would require a detailed knowledge of the side-chain flexibility as well as suitable models for spin relaxation in the presence of rf field (for the TOCSY experiment).

Instead of a multiparameter theoretical study, we have experimentally evaluated the performance of the two methods (relayed COSY and HC–CH TOCSY) on two different samples: on ubiquitin, a fast tumbling small protein, and on MSG (81 KDa). To avoid any bias due to different sampling methods along t_1 (standard *vs* constant time), 2D correlation spectra [F_2(^{13}C aliph) and F_3 (^{1}H)] were recorded under similar experimental conditions (cf. supplementary figure S2).

In the case of ubiquitin, the two methods exhibit similar sensitivity: stronger intermediate correlations (Cδ → Cγ, Cδ → Cβ or Cγ → Cδ) are detected in relayed COSY but the CH_3-Cα correlations are more intense in HC–CH TOCSY (average ratio between 2.0 and 2.5). These values are in agreement with the predictions shown in Fig. 4 if the transfer efficiencies are taken to the square as a result of the out-and-back scheme. In contrast, in the case of MSG, HC–CH TOCSY is significantly more sensitive than relayed COSY. For example, if one considers only the CH_3-Cα correlations for Ile, Val and Leu, more that 100 correlations have intensities at fivefold the noise level in HC–CH TOCSY and only a few (<10) in the relayed COSY. In small proteins with favorable relaxation properties, both methods perform similarly, but in larger proteins such as MSG (its t_c is one order of magnitude larger than for ubiquitin), HC–CH TOCSY clearly outperforms the other method. The few CH_3-Cα correlations detected in the 2D COSY spectrum of MSG (cf Fig. S2) have a ^{1}H chemical shift close to the random-coil value, suggesting that they belong to more flexible parts of MSG. However, for high-resolution 3D data sets, the losses of sensitivity are larger for HC–CH TOCSY than for relayed COSY where part of the t_1-evolution ($^{13}CH_3$) in a semi-constant sampling scheme can be merged with one $1/2J_{CC}$ delay.

Transfers in 3- and 2-spin systems are less challenging than for longer side-chains because shorter mixing times are needed. The correlations from the CH_3 to the backbone Cα or within V-shaped labeling are strong and weakly affected by relaxation. Notably, the Ala correlations are easily detected despite the non-optimal mixing times.

If the side-chains are not linearized as described here, the relayed COSY remains likely superior because branched spin-systems are highly detrimental for isotropic mixing as demonstrated by Eaton et al. (1990). Alternatively, on linearized side-chains, the HC–CH TOCSY approach is superior in terms of sensitivity for large proteins and also more easily optimized by changing the mixing time length.

Non-uniform sampling and J-deconvolution

In NMR of proteins, resolution and sensitivity are generally mutually exclusive. For instance, apodization functions that enhance resolution have a negative impact on the signal-to-noise ratio. At the acquisition stage, NUS affords new ways to optimize resolution and/or sensitivity over similar or reduced measurement times. By skipping some data points, it becomes possible to sample signals at longer t_1 and t_2 values with the ultimate limit set by the relaxation properties of the sampled spins (the CH_3 carbon during t_1 and the aliphatic ^{13}C during t_2 in the present case).

The digital resolution in our experimental spectra (Figs. 6, 7) is 27 Hz/pt along F_1 (^{13}C CH_3) and 77 Hz/pt along F_2 (^{13}C aliph). An average value of $R_2 = 21.7 \pm 12.1$ s^{-1} has been

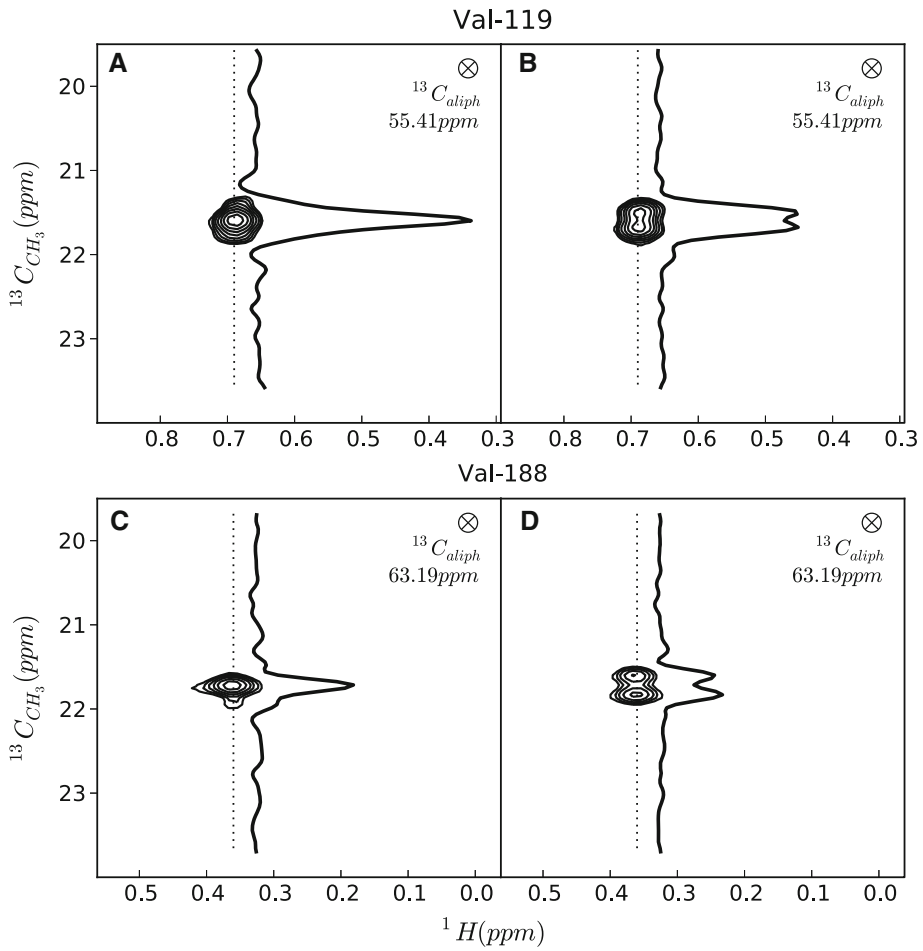

Fig. 5 J-coupling removal in the $F_1(^{13}C\ CH_3)$ dimension. During the t_1 evolution period, the $^{1}J_{CC}$ (35 Hz) coupling between the methyl ^{13}C and its adjacent carbon is active. While its suppression during acquisition by selective pulses is not feasible for all residue types simultaneously, it can be eliminated by suitable deconvolution during processing. The (F_1–F_3) planes are taken orthogonally to the strips shown in Fig. 6 at the frequency of the $^{13}C\alpha$ cross-peak. The couplings visible in *panels b* and *d* are eliminated in *panels a* and *c*. The contour levels for the 2D plots and the *vertical scale* for the 1D slices are identical in all panels. For the upper pair of spectra, the apparent line-width decreases from 57 to 33 Hz

reported for the slow component of the ^{13}C SQ coherences of protonated Ile CH_3 in MSG (Ollerenshaw et al. 2005). The transverse rate for the other carbons have not been measured experimentally but values in the range of 40–50 s^{-1} can be estimated: these carbons do not bear protons but have higher order parameters (S^2) than terminal CH_3. With only 20 % data points sampled, we are able to cover the entire range of chemical shifts for aliphatic carbons ($SW_2 = 66$ ppm) and achieve good spectral resolution.

Another source of line-broadening is the presence of scalar coupling(s) with adjacent carbons. In the labeling patterns of Fig. 1, the methyl $^{13}C'$s have one neighbor while the others have two. We have discussed above the benefit of experimentally decoupling $^{1}JC\alpha N$ and $^{1}JC\alpha C'$ in the case of slowly relaxing proteins (F_2 dimension). Improved resolution and sensitivity in the F_1 (CH_3) dimension can be achieved by removing the J-splitting by deconvolution prior to processing. As discussed in the experimental section (cf. Fig. 3), the deconvolution is only possible if no data point has been recorded for values close to $t_1^a = 1/(2 \times J_{CC})$: a large scaling factor would be necessary for these points, which already suffer from a poor signal-to-noise ratio.

The benefits of J-deconvolution are illustrated in Fig. 5: the same HC–CH TOCSY data set was processed with and without J-deconvolution and ^{13}C–^{1}H planes taken from the two resulting 3D matrices are compared. The "apparent" linewidth was quantified on a group of Ile cross-peaks (Cδ1 to Cα) using CCPNMR and shown to decrease on average from 57 to 33 Hz when deconvolution is used. This apparent width accounts for both the $^{1}J_{CC}$ and the real relaxation broadening.

Two important features of the J-deconvolution should be emphasized: (1) it cannot be applied on a data set sampled in a uniform manner (see experimental section); (2) it behaves differently than experimental decoupling because, as a modeling technique, it relies on the stringent assumption that the signal is modulated exactly as cos ($\pi \times J_{CC}t_1$). Due to cross-correlation effects (that may differ from one residue to another due to local flexibility), the amplitudes of the two doublet lines along the F_1 dimension are not always identical. As the signal is thus the superimposition of a smaller antiphase doublet on a larger in-phase component, it may result into incomplete deconvolution with remaining side-bands (Marion 2010).

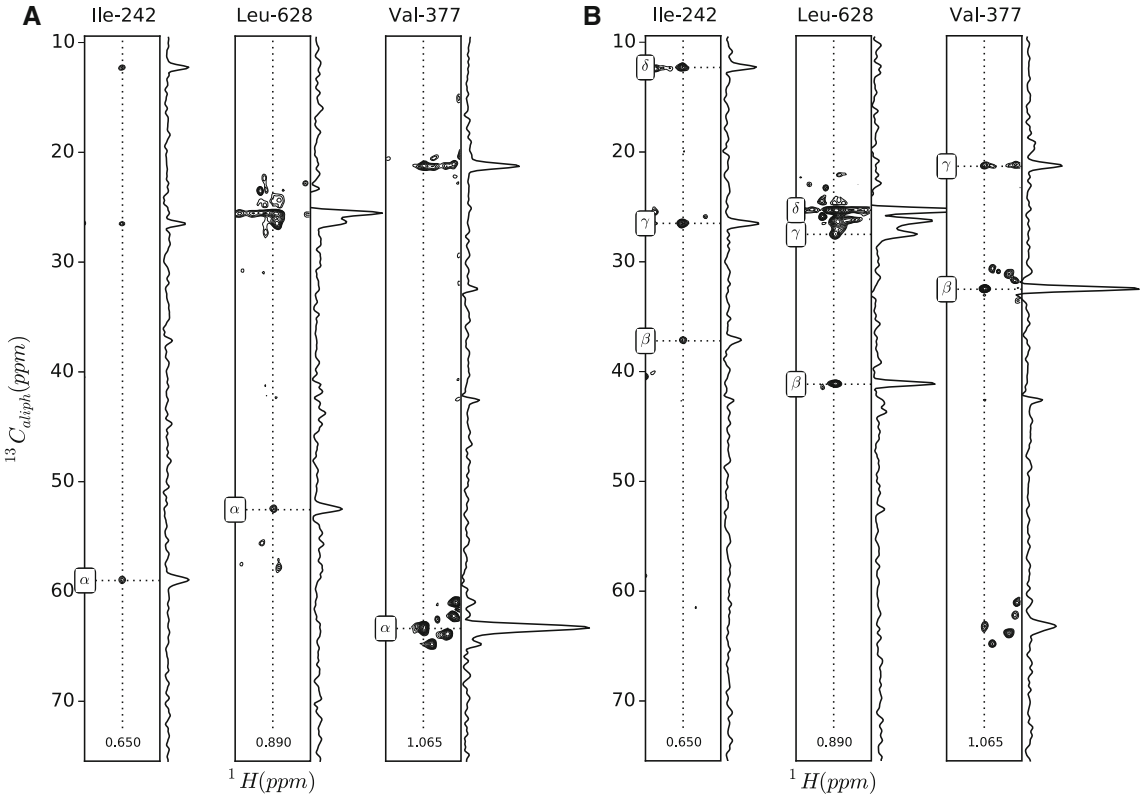

Fig. 6 Strips of the HC–CH TOCSY experiment recorded on a Bruker Avance 700 spectrometer on a MSG at 400 mM concentration (3 mm Shigemi tube, 19.4 ms FLOPSY mixing time, overall experimental time 65 h). The axes of the 3D data sets are F_1(^{13}C CH_3), F_2(^{13}C aliph) and F_3 (^{1}H) as defined in the pulse sequence shown in Fig. 2a. For each residue, the strip is taken at the ^{13}C chemical shift (F_1) of the methyl carbons (respectively Ile-δ_1, Leu-δ_1 and Val-γ_1). The proton chemical shift of the CH_3 group is indicated at the bottom of the strip. In *panel a*, a 19.4 ms mixing time is used to optimize the transfer from the Ile-δ_1, Leu-δ_1 or Val-γ_1 to the Cα resonance. In *panel b*, a 14.6 ms mixing time leads to transfers to intermediate nuclei (Cγ or Cβ). The cross-peaks for Val residues are generally stronger than for Ile and Leu due to the shorter side-chain. All contour plots are drawn at the same level and an estimate of the S/N ratio of the data is provided by the 1D slices drawn at the same *vertical scale*

In summary, an improved resolution can be obtained by combining NUS and J-deconvolution optimized for the relaxation properties of the studied protein.

AILV methyl group assignment

Our strategy for stereospecific assignment of branched amino-acids combines the production of ^{13}C linearized side chains and the recording of optimized correlation (TOCSY) spectra. In a first step, Ile-δ_1, Ile-γ_2, Leu-δ_1 and Val-γ_1 methyl groups can be assigned by virtue of their linkage to the peptide backbone in a linear ^{13}C spin system. Then, for each identified Leu-δ_1/Val-γ_1 group, the Leu-δ_2/Val-γ_2 group can be assigned since they are linked through a ^{13}Cγ or ^{13}Cβ in Leu and Val, respectively.

HC–CH TOCSY experiments were recorded on an MSG sample labeled as in Fig. 1. According to the previously estimated evolution of the coherences (Eaton et al. 1990) (cf Fig. 4), two values of the mixing time were selected: the longer one (19.4 ms) aims at transferring the coherences over 4 spins to detect the Cα of Ile and Leu and the shorter one (14.6 ms) is tuned for transfers over 3 spins for Val and V-shaped isotopomers.

Figure 6 illustrates the first step of the side-chain assignment: ^{13}C correlation strips of Ile-242, Leu-628 and Val-377 side-chains to the backbone are depicted for a mixing time of 19.4 ms (inset A) and 14.6 ms (inset B). These strips are taken at the frequency of the Ile-δ1, Leu-δ1 and Val-g1 carbon in F_1. For the shorter mixing time, cross-peaks to Cγ or Cβ are visible while the Cα become apparent at longer mixing time. The theoretical dependence of the TOCSY transfers reported in Fig. 4 agrees with the notion that the detection of the Cα is more demanding than for other carbons because of the longer mixing time required, which is relevant in the case of large proteins with fast relaxing signals. Note that the coherence transfer is more efficient in 3-spin systems (Val) than in 4-spin ones (Ile, Val).

The assignment process involves matching the Cα and Cβ frequencies in this side-chain experiment (with enhanced ^{13}C resolution due to non-uniform sampling) and in the

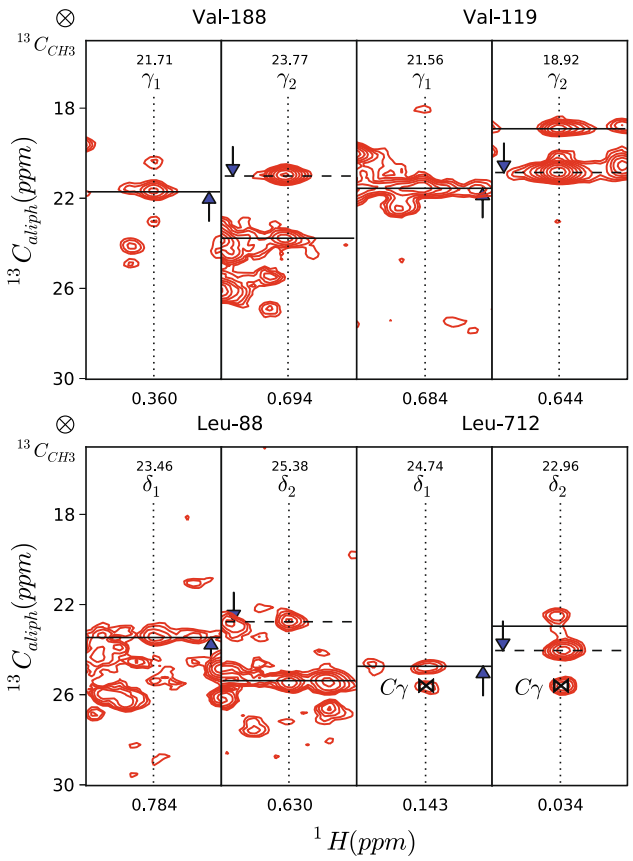

Fig. 7 Strips of the HC–CH TOCSY experiment recorded on a Varian–Agilent spectrometer on MSG at 1 mM concentration (5 mm Shigemi tube, 23.3 ms DIPSI-3 mixing time, overall experimental time 70 h). The axes of the 3D data sets are F_1(^{13}C CH3), F_2(^{13}C aliph) and F_3 (^{1}H) as defined in the pulse sequence shown in Fig. 2. For each residue, two strips taken at the ^{13}C chemical shift (F_1) of the methyl carbons (respectively Val-γ_1 and Val-γ_2, Leu-δ_1 and Leu-δ_2) are shown. The corresponding frequency is indicated at the top of the strip and the diagonal peak ($F_1 = F_2$) marked with a continuous line. One methyl group (Val-γ_1 and Leu-δ_1) correlates with the backbone (Cα) (outside of the depicted area but visible in figure S3). The second one (Val-γ_2 and Leu-δ_2) provides the correlation with the first one (*dashed line*) that is shifted by 0.7 ppm due to the deuterium isotopic shift ($CH_3 \rightarrow CD_3$) (this difference is indicated by *blue arrows*)

backbone experiments (such as HNCA, CBCANH…). In the case of crowded spectra (as in large proteins), the benefits of enhanced ^{13}C resolution due to NUS are lost if the backbone experiments are not recorded with adequate resolution.

Figure 7 illustrates the second step, where the V-shaped isotopomers (Leu-δ_2, Val-γ_2) are employed to stereospecifically assign the second methyl groups in these residues. The strip taken at the Leu-δ_2/Val-γ_2 $^{13}CH_3$ frequency exhibits a cross-peak at the Leu-δ_1/Val-γ_2 $^{13}C^2H_3$ frequency as well as with the spins in-between. Matching this Leu-δ_1/Val-γ_2 $^{13}C^2H_3$ frequency with the diagonal $^{13}CH_3$ peak in the linearly-labeled isotopomers is made difficult by isotopic shifts. In the V-shaped labeling isotopomers, the Leu-δ_1/Val-γ_2 ^{13}C bears 3 deuterium nuclei while it is protonated in the linearly-labeled isotopomers. This 0.7–0.9 ppm isotopic shift (Rosen et al. 1996) is indicated by blue arrows in Fig. 7 for Val and Leu residues. Currently, most software packages used for NMR analysis are not able to deal with this inconvenient H–D isotopic effect. Along F_1 (methyl ^{13}C), all resonances correspond to protonated carbons in contrast to the F_2 dimension (aliphatic ^{13}C), where all peaks (except the diagonal) come from deuterated carbons. Shifting the entire spectrum along F_2 by the amount of the isotopic shift is thus not a solution, but a basic script can easily perform the correction for the protonated $^{13}CH_3$ directly on the peak list.

To evaluate the performance of our approach, we have quantified the number of CH_3-Cα correlations that could be detected in two TOCSY experiments (experimental time <6 days) using a sample that contains only 30 nmol of labeled MSG (80 mL at 400 mM) as described in Fig. 6. Using such low amount of labeled material 62, 89, 53, 63 % of these correlations for Ile-δ_1, Ile-γ_2, Leu-δ_1 and Val-γ_1 are found to be fivefold larger than the noise level. The percentage of detected correlations for Leu-δ_1 and Val-γ_1 increases to 85 and 91 % using a 1-mM sample in a regular 5 mm Shigemi tube (300 nmol). The completeness of the Tugarinov–Kay assignment (Tugarinov and Kay 2003a) is higher than ours but relies on 7 different experiments using a 1 mM sample each optimized for a given type of side-chain with an overall experimental time longer than 22 days.

The amount of detected correlations can be influenced by 3 phenomena: the efficiency of the precursor incorporation, the local density of the CH_3 groups and the side-chain flexibility. In our labeling protocol, the Leu- and Val-precursors for (cf Fig. 1a) are added one hour before induction of expression and thus slight variations in the cell growth rates may have impacted the relative incorporation of Leu/Val versus Ile/Ala in the two sample preparations used for the current study. In the 400 mM sample, the ^{13}C labeling of Ala introduces 73 additional CH_3 groups in MSG that could also compromise the benefit of the HMQC (methyl-TROSY) (Tugarinov et al. 2003), since the TROSY effect is maximum for high levels of deuteration and it can be locally reduced by the presence of other CH_3 groups. Finally, inspection of a ^{1}H–^{13}C methyl correlation spectrum of any protein shows a wide range of intensities due to local environment (intermethyl distances) and flexibility (order parameter S^2).

If the investigated protein exhibits limited solubility, the strategy described above could be improved by producing two different samples, one labeled on Leu-δ_1, Val-γ_1 and Ile-δ_1 and the second on Leu-δ_2, Val-γ_2 and Ile-δ_2. Provided that the experimental conditions can be kept identical (pH, temperature…), the interpre-tation of the less crowded spectra would be made easier but the spectrometer time would be twice as long.

The Ala residues could also be assigned completely using the same sample at no additional expense. The same strategy could also be expanded to Thr (which shares the same spin-system topology as Val-γ_1 and Ile-γ_2), once suitable ^{13}C labeled precursors become available. The CH_3-TROSY effect in large proteins (see above) is deflated by the proton density and preparing more than one sample with alternate labeling becomes a potentially attractive strategy to enhance the sensitivity.

For a reliable assignment in a high molecular weight protein, two spectra are thus necessary, one with a shorter mixing tuned for the Cβ and one with a longer mixing time tuned for the Cα. These assignments are possibly complemented by relayed COSY experiments (Tugarinov and Kay 2003a, b; Tugarinov et al. 2013), which may be more sensitive for a number of specific residues.

Conclusion

In this paper, a new strategy combining advanced isotopic labeling and NMR spectrocopic tools has been proposed to achieve the full and stereospecific assignment of AILV side-chains in larger proteins. A unique ^{13}C–^{13}C TOCSY experiment with enhanced resolution afforded by an optimized NUS schedule has been employed. This experiment, which performs well only on linear spin topologies, requires a protein sample where all side-chain spin systems have been suitably (isotopically) linearized. In addition, we have shown that different labeling patterns could be designed to discriminate all CH_3 ILV groups. First, in the easy instance of Ile, where Ile-δ_1 and Ile-γ_2 are non-equivalent spin-systems, but also for Val and Leu, where the *pro-R* and *pro-S* methyl groups are anchored to non equivalent spin-systems that could be discriminated spectroscopically. Finally, the implementation of the out-and-back HC–CH TOCSY approach is also applicable to other aliphatic residues (such as Thr) that are shorter and inherently linear, once their precursors become available.

Acknowledgments We would like to thank Dr. S.J. Remington for providing the MSG plasmid, Dr. D. Chipman for providing the ALSII plasmid, Pr. Vladislav Yu. Orekhov for providing MDDNMR software, Mrs I. Ayala, Drs. P. Gans and M. Plevin for stimulating discussions and Drs Laura Lerner and Michael Caffrey for manuscript editing. This work used the high-field NMR and the isotopic labeling facilities at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 Grant Agreement No. 260887.

References

Ayala I, Hamelin O, Amero C et al (2012) An optimized isotopic labeling strategy of isoleucine-γ_2 methyl groups for solution NMR studies of high molecular weight proteins. Chem Commun 48:1434. doi:10.1039/c1cc12932e

Bax A, Clore GM, Gronenborn AM (1990) ^1H–^1H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^1H and ^{13}C spectra of ^{13}C-enriched proteins. J Magn Reson 88:425–431. doi:10.1016/0022-2364(90)90202-K

Braunschweiler L, Ernst RR (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J Magn Reson 53:521–528. doi:10.1016/0022-2364(83)90226-3

Chipman D, Barak Z, Schloss JV (1998) Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases. Biochim Biophys Acta 1385:401–419. doi:10.1016/S0167-4838(98)00083-1

Davis DG, Bax A (1985) Assignment of complex proton NMR spectra via two-dimensional homonuclear Hartmann–Hahn spectroscopy. J Am Chem Soc 107:2820–2821. doi:10.1021/ja00295a052

Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. doi:10.1007/BF00197809

Dumas R, Biou V, Halgand F et al (2001) Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase. Acc Chem Res 34:399–408. doi:10.1021/ar000082w

Eaton HL, Fesik SW, Glaser SJ, Drobny GP (1990) Time dependence of ^{13}C–^{13}C magnetization transfer in isotropic mixing experiments involving amino acid spin systems. J Magn Reson 90:452–463. doi:10.1016/0022-2364(90)90050-J

Eich G, Bodenhausen G, Ernst RR (1982) Exploring nuclear spin systems by relayed magnetization transfer. J Am Chem Soc 104:3731–3732. doi:10.1021/ja00377a036

Gans P, Hamelin O, Sounier R et al (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem Int Ed 49:1958–1962. doi:10.1002/anie.200905660

Gardner KH, Kay LE (1997) Production and incorporation of ^{15}N, ^{13}C, ^2H (^1H-δ_1 methyl) isoleucine into proteins for multidimensional NMR studies. J Am Chem Soc 119:7599–7600. doi:10.1021/ja9706514

Godoy-Ruiz R, Guo C, Tugarinov V (2010) Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc 132:18340–18350. doi:10.1021/ja1083656

Goto NK, Gardner KH, Mueller GA et al (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ_1) methyl-protonated ^{15}N-, ^{13}C-, ^2H-labeled proteins. J Biomol NMR 13:369–374. doi:10.1023/A:1008393201236

Güntert P, Braun W, Billeter M, Wüthrich K (2001) Automated stereospecific proton NMR assignments and their impact on the precision of protein structure determinations in solution. J Am Chem Soc 111:3997–4004. doi:10.1021/ja00193a036

Hill CM, Pang SS, Duggleby RG (1997) Purification of *Escherichia coli* acetohydroxyacid synthase isoenzyme II and reconstitution of active enzyme from its individual pure subunits. Biochem J 327:891–898. doi:10.1042/bj3270891

Hu W, Namanja AT, Wong S, Chen Y (2012) Selective editing of Val and Leu methyl groups in high molecular weight protein NMR. J Biomol NMR 53:113–124. doi:10.1007/s10858-012-9629-2

Hyberts SG, Märki W, Wagner G (1987) Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c. Eur J Biochem 164:625–635. doi:10.1111/j.1432-1033.1987.tb11173.x

Kadkhodaie M, Rivas O, Tan M et al (1991) Broadband homonuclear cross polarization using flip-flop spectroscopy. J Magn Reson 91:437–443. doi:10.1016/0022-2364(91)90210-K

Kainosho M, Torizawa T, Iwashita Y et al (2006) Optimal isotope labeling for NMR protein structure determinations. Nature 440:52–57. doi:10.1038/nature04525

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Ed Engl 50:5556–5559. doi:10.1002/anie.201100370

Kerfah R, Plevin MJ, Pessey O et al (2014) Scrambling free combinatorial labeling of alanine-β, isoleucine-δ1, leucine-proS and valine-proS methyl groups for the detection of long range NOEs. J Biomol NMR 61:73–82. doi:10.1007/s10858-014-9887-2

Marion D (2010) Combining methods for speeding up multidimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins. J Magn Reson 206:81–87. doi:10.1016/j.jmr.2010.06.007

Mas G, Crublet E, Hamelin O et al (2013) Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. J Biomol NMR 57:251–262. doi:10.1007/s10858-013-9785-z

McCoy MA, Mueller L (1993) Selective decoupling. J Magn Reson Ser A 101:122–130. doi:10.1006/jmra.1993.1021

Neri D, Szyperski T, Otting G et al (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry 28:7510–7516. doi:10.1021/bi00445a003

Ollerenshaw JE, Tugarinov V, Skrynnikov NR, Kay LE (2005) Comparison of ^{13}CH3, ^{13}CH$_2$D, and ^{13}CHD$_2$ methyl labeling strategies in proteins. J Biomol NMR 33:25–41. doi:10.1007/s10858-005-2614-2

Otten R, Chu B, Krewulak KD et al (2010) Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. J Am Chem Soc 132:2952–2960. doi:10.1021/ja907706a

Plevin MJ, Hamelin O, Boisbouvier J, Gans P (2011) A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins. J Biomol NMR 49:61–67. doi:10.1007/s10858-010-9463-3

Pristovšek P, Franzoni L (2006) Stereospecific assignments of protein NMR resonances based on the tertiary structure and 2D/3D NOE data. J Comput Chem 27:791–797. doi:10.1002/jcc.20389

Rosen MK, Gardner KH, Willis RC et al (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol 263:627–636. doi:10.1006/jmbi.1996.0603

Scholz I, Jehle S, Schmieder P et al (2007) J-deconvolution using maximum entropy reconstruction applied to ^{13}C–^{13}C solid-state cross-polarization magic-angle-spinning NMR of proteins. J Am Chem Soc 129:6682–6683. doi:10.1021/ja070849g

Senn H, Werner B, Messerle BA et al (1989) Stereospecific assignment of the methyl ^{1}H NMR lines of valine and leucine in polypeptides by nonrandom ^{13}C labeling. FEBS Lett 249:113–118. doi:10.1016/0014-5793(89)80027-4

Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. J Magn Reson 77:274–293. doi:10.1016/0022-2364(88)90178-3

Shimba N, Stern AS, Craik CS et al (2003) Elimination of ^{13}Cα splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. J Am Chem Soc 125:2382–2383. doi:10.1021/ja027973e

Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l(1)-norm reconstruction. J Magn Reson 188:295–300. doi:10.1016/j.jmr.2007.07.008

Tang C, Iwahara J, Clore GM (2005) Accurate determination of leucine and valine side-chain conformations using U-[^{15}N/^{13}C/^{2}H]/[^{1}H-(methine/methyl)-Leu/Val] isotope labeling, NOE pattern recognition, and methine Cγ–Hγ/Cβ–Hβ residual dipolar couplings: application to the 34-kDa Enzyme IIA Chitobiose. J Biomol NMR 33:105–121. doi:10.1007/s10858-005-1206-5

Tugarinov V, Kay LE (2003a) 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. doi:10.1021/ja030345s

Tugarinov V, Kay LE (2003b) Side chain assignments of Ile-δ1 methyl groups in high molecular weight proteins: an application to a 46 ns tumbling molecule. J Am Chem Soc 125:5701–5706. doi:10.1021/ja021452

Tugarinov V, Kay LE (2004) Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G. J Am Chem Soc 126:9827–9836. doi:10.1021/ja048738u

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^{1}H–^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125:10420–10428. doi:10.1021/ja030153x

Tugarinov V, Choy W-Y, Orekhov VY, Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci USA 102:622–627. doi:10.1073/pnas.0407792102

Tugarinov V, Venditti V, Clore GM (2013) A NMR experiment for simultaneous correlations of valine and leucine/isoleucine methyls with carbonyl chemical shifts in proteins. J Biomol NMR 58:1–8. doi:10.1007/s10858-013-9803-1

Van Melckebeke H, Simorre J-P, Brutscher B (2004) Suppression of artifacts induced by homonuclear decoupling in amino-acid-type edited methyl ^{1}H–^{13}C correlation experiments. J Magn Reson 170:199–205. doi:10.1016/j.jmr.2004.06.016

Vuister GW, Yamazaki T, Torchia DA, Bax A (1993) Measurement of two- and three-bond ^{13}C–^{1}H J couplings to the C delta carbons of leucine residues in staphylococcal nuclease. J Biomol NMR 3:297–306. doi:10.1007/BF00212516

Wüthrich K, Wider G, Wagner G, Braun W (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. J Mol Biol 155:311–319. doi:10.1016/0022-2836(82)90007-9