Journal of Biomolecular NMR, **19:** 267–272, 2001.
KLUWER/ESCOM

Selective 'unlabeling' of amino acids in fractionally ^{13}C labeled proteins: An approach for stereospecific NMR assignments of CH_3 groups in Val and Leu residues

H.S. Atreya & K.V.R. Chary*
Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai-400005, India

Received 7 November 2000; Accepted 3 January 2001

Key words: amino acid biosynthesis, amino acid unlabeling, coupling constant, diastereotopic methyl groups, HSQC, isotope labeling, multidimensional NMR, stereospecific assignments

Abstract

A novel methodology for stereospecific NMR assignments of methyl (CH_3) groups of Val and Leu residues in fractionally ^{13}C-labeled proteins is presented. The approach is based on selective 'unlabeling' of specific amino acids in proteins while fractionally ^{13}C-labeling the rest. A 2D [^{13}C-^{1}H] HSQC spectrum recorded on such a sample is devoid of peaks belonging to the 'unlabeled' amino acid residues. Such spectral simplification aids in unambiguous stereospecific assignment of diastereotopic CH_3 groups in Val and Leu residues in large proteins. This methodology has been demonstrated on a 15 kDa calcium binding protein from *Entamoeba histolytica* (*Eh*-CaBP).

The development of new experimental techniques and methodologies in the field of NMR spectroscopy, in recent years, has facilitated in obtaining high resolution 3D structures of proteins in solution (Wüthrich, 1986; Bax and Grzesiek, 1993; Wagner, 1997; Sattler et al., 1999). On one hand, these techniques and methodologies have been aimed at enhancing the sensitivity and spectral resolution for large proteins ($M_r > 20$ kDa), while on the other, the focus has been on devising experiments to derive different types of structural restraints such as distance and torsion angle constraints, chemical shifts, dipolar couplings etcetera, which in turn are used in elucidating their 3D structures (Clore et al., 1998). One such class of restraints is derived from stereospecific assignments of diastereotopic methyl (CH_3) groups of Val and Leu residues in proteins. These assignments provide the most important information concerning the orientation of isopropyl groups in Val and Leu residues about the C^{α}-C^{β} and C^{β}-C^{γ} bonds, respectively, and as a result have significant influence on the precision of derived 3D structures (Billeter et al., 1989; Güntert et al., 1989; Zwahlen et al., 1998). This is best illustrated by the structural characterization of differences between X-ray-derived and NMR-derived high resolution structures for α-amylase tendamistat based on stereospecific assignments of these CH_3 groups (Billeter et al., 1989).

To date, the most widely used approach for stereospecific assignment of CH_3 groups in Val and Leu residues relies on fractional ^{13}C labeling (Neri et al., 1989). The basis of this approach lies in the fact that, in living organisms, the biosynthesis of both Val and Leu residues from glucose is stereoselective (Hill et al., 1971; Sylvester and Stevens, 1979). This implies that *pro-R* and *pro-S* CH_3 groups of the isopropyl moiety in these residues originate from two different pyruvate molecules. One pyruvate molecule contributes a two-carbon fragment (containing *pro-R* CH_3 and an adjacent >CH- group) to the isopropyl group, while the other contributes a *pro-S* CH_3 group. Thus, when the host micro-organism, overexpressing the protein of interest, is grown on minimal media containing 10% [$^{13}C_6$]-glucose and

*To whom correspondence should be addressed. E-mail: chary@tifr.res.in

90% [$^{12}C_6$]-glucose as the sole source of carbon, the probability that both diastereotopic CH_3 groups have originated from [$^{13}C_6$]-pyruvate molecules (which is derived from [$^{13}C_6$]-glucose) is only 1%, while the chances of either of these CH_3 groups originating from a [$^{13}C_6$]-pyruvate molecule is about 10%. Stereospecific distinction between these diastereotopic CH_3 groups has been achieved by recording a ^{13}C-^{13}C coupled 2D [^{13}C-^{1}H] HSQC spectrum on a fractionally ^{13}C-labeled protein, which yields a doublet for the ^{13}C resonance of the *pro-R* CH_3 group (γ^1 in Val and δ^1 in Leu) and a singlet for the ^{13}C resonance of the *pro-S* CH_3 group (γ^2 in Val and δ^2 in Leu) (Neri et al., 1989).

Other alternative strategies proposed to achieve such a stereospecific distinction between the CH_3 groups in proteins include both chemical as well as biochemical methods. In one method, block labeled (fractionally labeled) Val and Leu amino acids are incorporated into proteins (Tate et al., 1995). This approach suffers from the fact that isolation of fractionally labeled amino acids from microbial fermentation is non-trivial and fraught with low yields. Chemical methods include incorporation of chemically synthesized stereoselectively deuterated leucine (Ostler et al., 1993). Alternatively, new pulse sequences have been proposed to obtain stereospecific assignments of CH_3 groups in ^{13}C-fractionally labeled proteins. In one method, Leu ($^3J(H_\gamma,C_\delta)$) magnitudes were measured and utilized to distinguish its *pro-R* CH_3 group from that of *pro-S* (Sattler et al., 1992). In this method, stereospecific assignments of β-methylene groups are needed as a prerequisite along with Leu ($^3J(H_\beta,C_\delta)$ and $^3J(H_\beta,H_\gamma)$) magnitudes, which cannot be obtained easily for large sized proteins. In another method, a 3D CT-(H)CCH-COSY pulse sequence has been proposed to resolve the overlapping ^{1}H-^{13}C cross peaks arising from Val and Leu residues in a 2D-HSQC spectrum, by spreading them in the third dimension (Hu and Zuiderweg, 1996). In yet another attempt, a 2D spin-echo difference experiment has been proposed for stereospecific assignments of Val CH_3 groups (Vuister et al., 1993). This method is based on the measurement of $^3J(^{13}C^\gamma$-$^{15}N)$ or $^3J(^{13}C^\gamma$-$^{13}C')$ coupling constants. However, this experiment does not throw light on the stereospecific assignment of Leu CH_3 groups.

In spite of these existing methodologies, aimed at obtaining unambiguous stereospecific assignments of diastereotopic CH_3 groups, severe spectral overlap in the case of large sized proteins ($M_r > 20$ kDa) poses a formidable problem. This is evident from Figure 1A, which indicates the extent of overlap in spectral regions containing ^{13}C-^{1}H cross peaks of CH_3 groups in Val and Leu residues (shaded) with those of other amino acid residues that have cross peaks in their vicinity. This characterization is based on a statistical analysis of complete chemical shift data of proteins available in the BioMagResBank (BMRB) (Seavey et al., 1991). A total number of about 12 000 ^{13}C and 48 000 ^{1}H chemical shifts has been used in this analysis. The rectangular boxes in the figure depict regions containing 95% of the chemical shifts of individual resonances that lie within three times the standard deviation (i.e $\pm 3\sigma$ from the mean). Chemical shift values outside this range were treated as unusual and were not used in the analysis. Thus, it is seen from the figure that cross peaks arising from Ala(C^β-H^β), Lys(C^γ-H^γ), Ile($C^{\gamma1}$-$H^{\gamma1}$ and $C^{\gamma2}$-$H^{\gamma2}$), Met(C^ε-H^ε) and Thr(C^γ-H^γ) correlations have significant overlap with those of Val(C^γ-H^γ) or/and Leu(C^δ-H^δ) correlations, hampering the stereospecific assignments of the latter. Further, about 20% of the Val(C^γ-H^γ) cross peaks are seen to overlap with Leu(C^δ-H^δ) correlations, which cannot be resolved using the techniques mentioned above. In order to address the severity of such overlap particularly in large proteins, we have statistically measured the percentage of Val and Leu methyl resonances which overlap with those of Ala(C^β-H^β), Lys(C^γ-H^γ), Ile($C^{\gamma1}$-$H^{\gamma1}$ and $C^{\gamma2}$-$H^{\gamma2}$), Met(C^ε-H^ε) and Thr(C^γ-H^γ) resonances (see Figure 1B). For this purpose, 40 proteins with individual molecular weights in excess of 15 kDa were chosen from the BMRB (BMRB accession numbers are available in supplementary Table 1). Only those proteins for which complete side chain ^{13}C and ^{1}H assignments are available were used in this analysis. Cross peaks having both ^{13}C chemical shifts within 0.2 ppm and ^{1}H chemical shifts within 0.02 ppm were treated as overlapping. As is evident in Figure 1B, the extent of overlap of Val and Leu methyl cross peaks with those belonging to other residues is 25–40% in many cases.

Such severe spectral overlap can be overcome in the methodology presented in this communication. In this approach, specific amino acids in a given protein are selectively 'unlabeled' while simultaneously allowing the fractional ^{13}C labeling of other residues. Thus, the absence of cross peaks belonging to selectively unlabeled amino acids in various heteronuclear correlation spectra results in an enhanced spectral simplification, leading to stereospecific assignments of diastereotopic methyl groups of Val and Leu residues in a straightforward manner. This methodology is

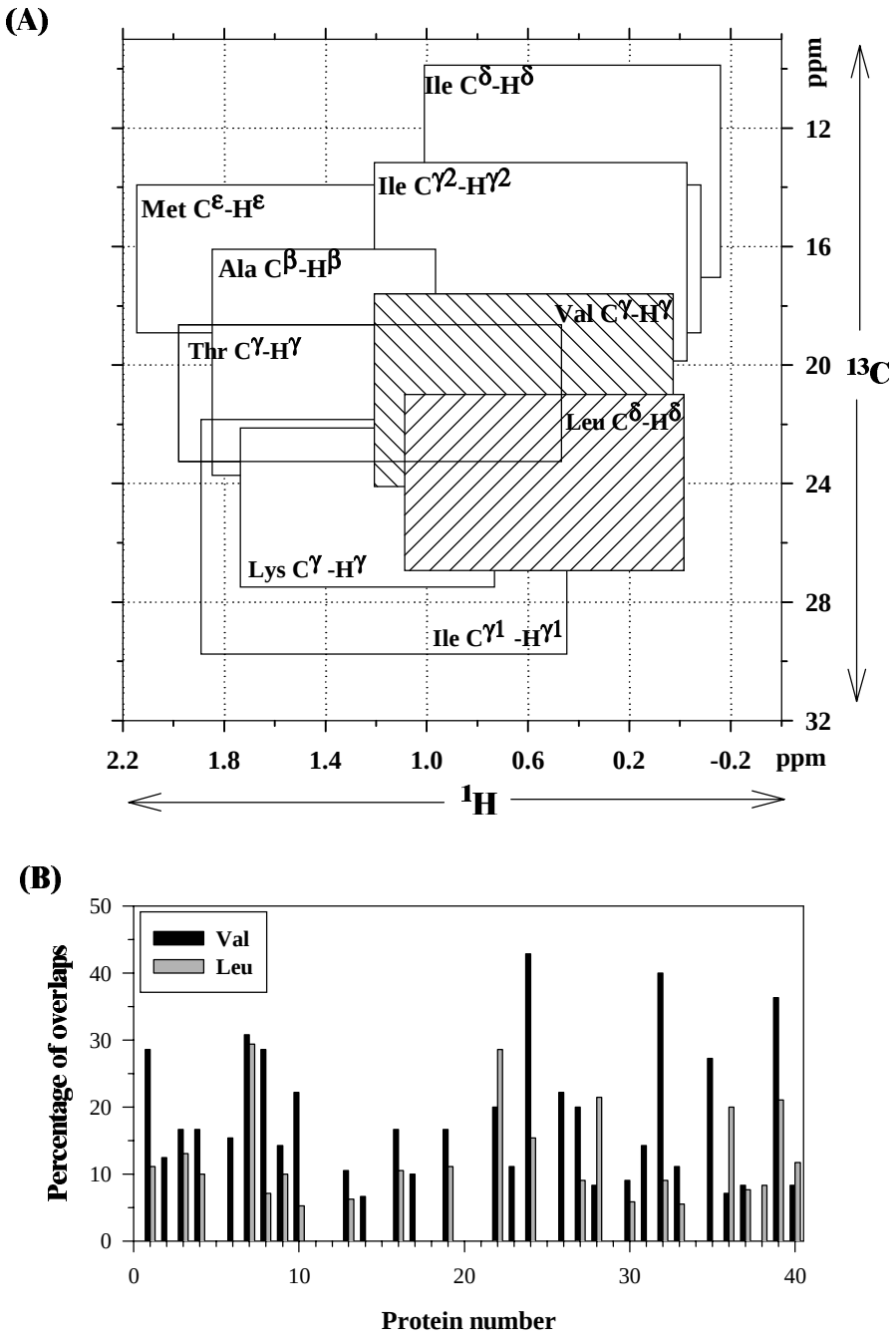

Figure 1. (A) A selected region of the 2D [^{13}C-^{1}H] HSQC spectrum containing ^{13}C-^{1}H cross peaks of CH_3 groups in Val and Leu residues (shaded) with other ^{13}C-^{1}H correlations that are in their vicinity (see text). The rectangular boxes depict regions containing 95% of ^{13}C-^{1}H cross peaks belonging to various amino acid residues that show up in this selected region. (B) Percentage of Val ($^{13}C^{\gamma}$-H^{γ}) and Leu ($^{13}C^{\delta}$-H^{δ}) cross peaks that overlap with those of other amino acid residues shown in (A), for 40 proteins chosen from the BMRB. Proteins ($M_r \geq 15$ kDa) for which complete side-chain ^{13}C and ^{1}H chemical shifts are available were chosen for the analysis.

demonstrated on a 15 kDa (134 residues) calcium-binding protein from the protozoan *Entamoeba histolytica* (*Eh*-CaBP).

In the present study, simultaneous unlabeling of selected amino acid residues and fractional ^{13}C labeling of the remaining amino acids was achieved by growing the host micro-organism (*E. coli* in the present study) on a minimal medium containing a mixture of roughly 10% [$^{13}C_6$]-glucose and 90% [$^{12}C_6$]-glucose, along with desired specific amino acid(s), to be unlabeled, in unlabeled form.

Eh-CaBP was expressed in *Escherichia coli* BL21(DE3) strain containing the PET-3c expression system. The minimal media used to overexpress *Eh*-CaBP contained M9 salts (Sambrook et al., 1989) supplemented with 0.250 g/l of $MgSO_4.2H_2O$, 0.015 g/l of $CaCl_2$, 0.4 g/l [$^{13}C_6$]-glucose, 3.6 g/l [$^{12}C_6$]-glucose and 1.0 g/l each of desired unlabeled amino acid(s) (stock solution of 1.0 g in 50 ml of H_2O was prepared and filter sterilized). Mid-log phase cells were treated with isopropylthio-β-galactoside for 4 h. *Eh*-CaBP gets expressed to an extent of ∼30% of the total cell proteins. The purification protocol involves the heat treatment of induced recombinant *E. coli* cell lysate followed by ion exchange chromatography. During heat treatment, the majority of *E. coli* proteins (except *Eh*-CaBP) get coagulated and are removed by centrifugation. *Eh*-CaBP, being a heat stable protein, remains in the supernatant and is passed through a DE-52 anion exchange column matrix. Binding of *Eh*-CaBP to the column was achieved in the presence of EGTA and the protein was specifically eluted by 10 mM $CaCl_2$ in 50 mM Tris-HCl buffer (pH 7.5) at 25 °C. The protein yield was expectedly higher than in the case of uniform ^{13}C or/and ^{15}N labeling. This is attributed to the additional supply of amino acids, which provide the micro-organisms with an abundantly enriched source of nutrients. Hence, samples with unlabeled amino acids were prepared using half the [$^{13}C_6$]-glucose required for the control (no unlabeling) sample. This keeps the total amount of [$^{13}C_6$]-glucose required (and hence the cost of the whole process) the same as when no unlabeling is carried out. In the present case, three different protein samples were prepared as discussed below. These were: the control sample (no unlabeling), a sample with Lys, Leu, Ile and Thr unlabeled while fractionally ^{13}C-labeling the rest, and another one with Lys, Ile and Thr unlabeled, while fractionally ^{13}C-labeling the rest.

NMR experiments were carried out on a Varian Unity$^+$ 600 MHz NMR spectrometer equipped with a pulsed field gradient unit and a triple resonance probe with actively shielded Z-gradients, operating at a ^1H frequency of 600.051 MHz. NMR measurements were performed with a sample of 0.6 ml of 2.0 mM *Eh*-CaBP in 30 mM $CaCl_2$ and 50 mM deuterated Tris buffer, at an optimized pH (6.0) and temperature (35 °C), in 99.9% ^2H_2O. The experiments recorded include sensitivity enhanced [^{13}C-^1H] HSQC (Kay et al., 1992) with both constant time and non-constant time options.

Figure 2A shows a selected region of the ^{13}C-^{13}C coupled 2D [^{13}C-^1H] HSQC spectrum recorded on a 2 mM sample of fractionally ^{13}C-labeled *Eh*-CaBP. As evident from this figure, the region suffers from an extensive spectral overlap and only a few Val(C^γ-H^γ) and Leu(C^δ-H^δ) cross peaks are amenable for stereospecific assignments. This prompted us to simplify the region by amino acid selective unlabeling. Based on Figure 1, three amino acids, namely, Ile, Lys and Thr, whose ^{13}C-^1H cross peaks overlap extensively with those of Val(C^γ-H^γ) and Leu(C^δ-H^δ) were chosen for selective unlabeling; two different protein samples were prepared as described above. In the first sample (hereafter referred to as Sample 1), Ile, Leu, Lys and Thr were selectively unlabeled and the remaining amino acids were fractionally ^{13}C-labeled. This resulted in the suppression of all cross peaks arising from unlabeled residues, and observation of only Ala(C^β-H^β) and Val(C^γ-H^γ) correlations (Figure 2B). Though Ala(C^β-H^β) cross peaks partially overlap with Val(C^γ-H^γ) cross peaks, our attempts to unlabel Ala failed, as it cross-metabolized to Val. In the second sample (hereafter referred to as Sample 2), Ile, Lys and Thr residues were selectively unlabeled, resulting in the observation of only Val(C^γ-H^γ) and Leu(C^δ-H^δ) correlations (Figure 2C).

The knowledge of Val ($^1H^\gamma$) and Leu ($^1H^\delta$) chemical shifts reported earlier (BMRB accession no: 4271) and Val ($^{13}C^\gamma$) and Leu ($^{13}C^\delta$) chemical shifts obtained by recording a 3D HCCH TOCSY (Bax et al., 1990) spectrum on a uniformly ^{13}C-labeled *Eh*-CaBP was used to assign the various (^{13}C-^1H) cross peaks in the above mentioned (^{13}C-^1H) HSQC spectra. As is evident in Figure 2B, all the Val ($^{13}C^\gamma$-$^1H^\gamma$) cross peaks could be unambiguously resolved. Even in a situation where two ^1H (H^γ or H^δ) resonances are 0.01 ppm apart, we could assign the respective ^{13}C-^1H cross peaks unambiguously. For example, $^1H^\gamma$ protons of V11 and V97, which resonate at 1.09 and 1.10 ppm, respectively, show two distinct cross peaks at $\omega_1 = 20.95$ and 20.85 ppm, respectively (see

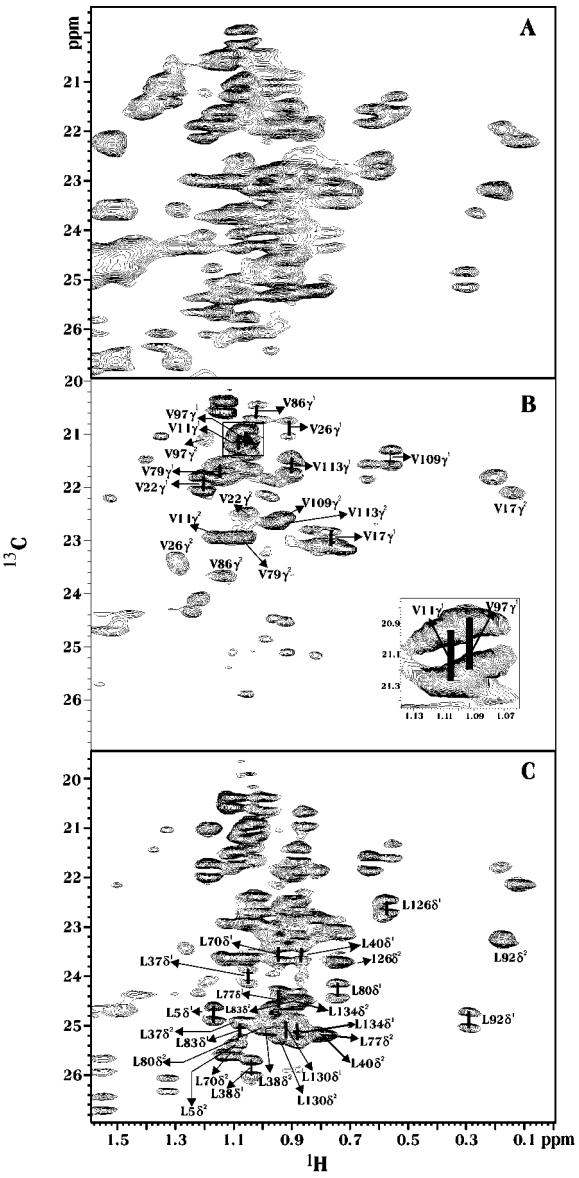

Figure 2. A selected region of the ^{13}C-^{13}C coupled and sensitivity enhanced 2D [^{13}C-^{1}H] HSQC spectrum recorded on (A) ^{13}C-fractionally labeled, (B) Ile, Lys, Leu and Thr unlabeled and fractionally ^{13}C labeling of the rest and (C) Ile, Lys and Thr unlabeled and fractionally ^{13}C labeling of the rest, *Eh*-CaBP samples. Experimental parameters were as follows: recycle delay 1 s, 32 scans/t_1 increment. The numbers of time domain data points were 512 and 2048 along t_1 and t_2, respectively. The ^{1}H carrier frequency was kept at the water resonance (4.68 ppm) and the ^{13}C carrier frequency was at the center of the aliphatic carbon region (35 ppm). The data were multiplied with a sine-bell window function shifted by $\pi/4$ along both t_1 and t_2 axes and zero filled to 2048 along the t_1 axis, prior to 2D-FT. The digital resolution along the ω_1 and ω_2 axes corresponds to 4.87 and 3.90 Hz/pt, respectively. Stereospecific assignments (according to IUPAC nomenclature) for Val(C^{γ}-H^{γ}) cross peaks are indicated in (B), while those of Leu(C^{δ}-H^{δ}) cross peaks are indicated in (C). The inset in (B) shows an expanded region containing $C^{\gamma 1}$-$H^{\gamma 1}$ cross peaks for V97 and V11.

inset in Figure 2B), aiding the assignments of their respective ^{13}C-^{1}H cross peaks. This was further substantiated by the assignment of cross peaks belonging to their individual diastereotopic partners with the expected singlet or doublet peak structure along the ω_1 axis.

Likewise, with the prior knowledge of $^{13}C^{\delta}$ and $^{1}H^{\delta}$ chemical shifts, $^{13}C^{\delta}$-$^{1}H^{\delta}$ cross peaks belonging to Leu CH_3 groups (shown in Figure 2C) were assigned. Further, the near degeneracy of cross peaks belonging to Leu methyl groups was resolved with the supporting evidence we could derive from the assignments of cross peaks belonging to their diastereotopic partners, as discussed above. For example, L40(δ^2) and L77(δ^2) accidently have degenerate $^{13}C^{\delta}$ and $^{1}H^{\delta}$ chemical shifts and as a result show a single cross peak in the (^{13}C-^{1}H) HSQC spectrum(see Figure 2C). However, their diastereotopic partners, L40(δ^1) and L77(δ^1), are well resolved and appear as distinct doublets in the spectrum. This helped us in the stereospecific assignment of $^{13}C^{\delta}$-$^{1}H^{\delta}$ cross peaks belonging to these residues. These assignments were further substantiated by recording a [^{13}C-^{1}H] CT-HSQC spectrum on Sample 2 (Supplementary Figure 1). In this experiment, when the constant time delay is set to 28.6 ms (corresponding to $1/^{1}J_{CC}$ [$^{1}J_{CC} \sim 35$ Hz]), Val(C^{γ}-H^{γ}) and Leu(C^{δ}-H^{δ}) cross peaks arising from the *pro-R* CH_3 groups are opposite in sign relative to those arising from *pro-S* CH_3 groups, thus enabling their distinction (Kay et al., 1992).

Owing to the fact that Val(C^{γ}-H^{γ}) cross peaks partially overlap with Leu(C^{δ}-H^{δ}) cross peaks, we tried to unlabel Val in addition to the unlabeling of Ile, Lys and Thr residues. Such an attempt was not successful, as Val partially cross-metabolized into Leu, thus tending to unlabel the latter too. However, with the aid of two different protein preparations outlined above, stereospecific assignments for all the CH_3 groups of Val and Leu residues were obtained unambiguously (available as Supplementary Table 2). It is important to note that such distinction of Val and Leu cross peaks achieved in a selective fashion as described here, cannot be obtained in any of the methodologies proposed earlier. In the case of higher molecular weight proteins ($M_r > 20$ kDa), the methodology outlined above can be further combined with random fractional deuteration (Gardner and Kay, 1998), which results in an enhanced spectral sensitivity and resolution.

In conclusion, we have presented a new methodology for obtaining stereospecific assignments of diastereotopic methyl groups in Val and Leu residues

in large-molecular-weight proteins. In this approach, selective amino acids in a ^{13}C-fractionally labeled protein are 'unlabeled'. This leads to the absence of their cross peaks in a 2D [^{13}C-^{1}H] HSQC spectrum, resulting in an enhanced spectral simplification and subsequently to unambiguous stereospecific assignments. Efforts are made to simplify ^{13}C/^{15}N edited 3D NOESY and 3D TOCSY spectra using such unlabeling strategies to derive more structural restraints, leading to high precision protein structures.

Acknowledgements

The facilities provided by the National Facility for High Field NMR, supported by Department of Science and Technology (DST), Department of Biotechnology (DBT), Council of Scientific and Industrial Research (CSIR), and Tata Institute of Fundamental Research, Mumbai, are gratefully acknowledged. We would like to thank Prof. Alok Bhattacharya (JNU, New Delhi) for providing the *Eh*-CaBP clone. We dedicate this paper to Prof. Girjesh Govil on his 60th birthday.

References

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Magn. Reson.*, **88**, 425–431.

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.

Billeter, M., Kline, A.D., Braun, W., Huber, R. and Wüthrich, K. (1989) *J. Mol. Biol.*, **206**, 677–687.

Clore, G.M. and Gronenborn, A.M. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 5891–5898.

Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.

Hill, R.K., Yan, S. and Arfin, S.M. (1971) *J. Am. Chem. Soc.*, **95**, 7857–7859.

Hu, W. and Zuiderweg, E.R. (1996) *J. Magn. Reson.*, **B113**, 70–75.

Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.

Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry*, **28**, 7510–7516.

Ostler, G., Spteriou, A., Moody, C.M., Khan, J.A., Birdsall, B., Carr, M.D., Young, D.W. and Feeney, J. (1993) *FEBS Lett.*, **318**, 177–180.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Vols 1–3, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbour, NY.

Sattler, M., Schwalbe, H. and Griesinger, C. (1992) *J. Am. Chem. Soc.*, **114**, 1126–1127.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) *Prog. NMR Spectrosc.* **34**, 93–158.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.M. (1991) *J. Biomol. NMR*, **1**, 217–236.

Sylvester, S.R. and Stevens, C.M. (1979) *Biochemistry*, **18**, 4529–4531.

Tate, S., Ushioda, T., Utsunomiya-Tate, M., Shibuya, K. and Ohyama, M. (1995) *Biochemistry*, **34**, 14637–14638.

Vuister, G.W., Wang, A. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 5334–5335.

Wagner, G. (1997) *Nat. Struct. Biol.*, **4**, 841–844.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.

Zwahlen, C., Vincent, S.J.F., Gardner, K.H. and Kay, L.E. (1998) *J. Am. Chem. Soc.*, **120**, 4825–4831.