J|A|C|S
ARTICLES
Published on Web 10/16/2003

Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods

Vitali Tugarinov and Lewis E. Kay*

Contribution from the Protein Engineering Network Centres of Excellence and the Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received June 4, 2003; Revised Manuscript Received July 31, 2003; E-mail: kay@pound.med.utoronto.ca

Abstract: New NMR experiments are presented for the assignment of methyl ^{13}C and ^{1}H chemical shifts from Ile, Leu, and Val residues in high molecular weight proteins. The first class of pulse schemes transfers magnetization from the methyl group to the backbone amide spins for detection, while the second more sensitive class uses an "out-and-back" transfer scheme in which side-chain carbons or backbone carbonyls are correlated with methyl ^{13}C and ^{1}H spins. Both groups of experiments benefit from a new isotopic labeling scheme for protonation of Leu and Val methyl groups in large deuterated proteins. The approach makes use of α-ketoisovalerate that is ^{13}C-labeled and protonated in one of its methyl groups ($^{13}CH_3$), while the other methyl is $^{12}CD_3$. The use of this biosynthetic precursor leads to production of Leu and Val residues that are $^{13}CH_3$-labeled at only a single methyl position. Although this labeling pattern effectively reduces by 2-fold the concentration of Leu and Val methyls in NMR samples, it ensures linearity of Val and Leu side-chain ^{13}C spin-systems, leading to higher sensitivity and, for certain classes of experiments, substantial simplification of NMR spectra. Very near complete assignments of the 276 Ile (δ1 only), Leu, and Val methyl groups in the single-chain 723-residue enzyme malate synthase G (MSG, molecular tumbling time 37 ± 2 ns at 37 °C) have been obtained using the proposed isotopic labeling strategy in combination with the new NMR experiments.

Introduction

In the decade since the initial development of triple resonance multidimensional NMR methodology, the size limitations associated with the study of biomolecules by NMR spectroscopy have significantly decreased. A major advance in this regard has been the use of highly deuterated, ^{15}N-,^{13}C-labeled samples, leading to attenuation of the relaxation rates of NMR active nuclei and concomitant sensitivity and resolution gains in spectra.[1,2] A second, important contribution, building on the labeling scheme mentioned above, has been the development of TROSY-based spectroscopy.[3] Applications include the near complete backbone assignment of a 110 kDa homooctameric protein, 7,8-dihydroneopterin aldolase,[4] detailed NMR studies of the 723-residue monomeric enzyme malate synthase G from *E. coli* (81.4 kDa),[5,6] studies of membrane proteins such as OmpX,[7,8] OmpA,[9] and PagP,[10] and the assignment of a dimeric construct of the human tumor suppressor protein p53 (67 kDa, 279 residues).[11] New technologies have recently emerged, holding promise for the study of supramolecular structures as well. For example, Wüthrich and co-workers have published a CRIPT-based $^{1}H-^{15}N$ correlation spectrum of ^{15}N-labeled GroES bound to GroEL, a complex of 900 kDa,[12] and our laboratory has shown that high-quality $^{1}H-^{13}C$ TROSY correlation spectra of Ile methyls in the protease ClpP, a 305 kDa 14-mer, can be recorded (5 °C) in very modest measuring times.[13]

The great majority of the methodology that has been developed to date has focused on correlation spectroscopy involving backbone protein spins. Building on the success of such developments, our laboratory has now turned to the assignment of side chains. Our focus, however, has been to restrict analysis to the methyls of Ile, Leu, and Val for a number

(1) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357−406.
(2) Farmer, B. T.; Venters, R. A. In *NMR of perdeuterated large proteins*; Krishna, N. R., Berliner, L. J., Eds.; Kluwer Academic/Plenum Publishers: New York, 1998; Vol. 16, pp 75−120.
(3) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(4) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.* **2000**, *122*, 7543−7548.
(5) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10025−10035.
(6) Tugarinov, V.; Kay, L. E. *J. Mol. Biol.* **2003**, *327*, 1121−1133.
(7) Fernandez, C.; Hilty, C.; Bonjour, S.; Adeishvili, K.; Pervushin, K.; Wüthrich, K. *FEBS Lett.* **2001**, *504*, 173−8.
(8) Fernandez, C.; Adeishvili, K.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2358−63.
(9) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 334−8.
(10) Hwang, P. M.; Choy, W.-Y.; Lo, E. I.; Chen, L.; Forman-Kay, J. D.; Raetz, C. R. H.; Prive, G. G.; Bishop, R. E.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13560−13565.
(11) Mulder, F. A. A.; Ayed, A.; Yang, D.; Arrowsmith, C. H.; Kay, L. E. *J. Biomol. NMR* **2000**, *18*, 173−6.
(12) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. *Nature* **2002**, *418*, 207−221.
(13) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 10420−10428.

10.1021/ja030345s CCC: $25.00

of reasons. First, methyl groups give rise to intense correlations and have favorable relaxation properties so that even for applications involving high molecular weight systems methyl spectra are of good quality.[14] Second, methyls are most often localized to hydrophobic cores of proteins so that methyl distances measured from NOESY experiments provide valuable restraints for structure determination.[15−17] Third, methyls are excellent reporters of dynamics in proteins.[18,19] Fourth, robust approaches for the production of Ile (δ1), Leu, Val-methyl protonated, highly deuterated ^{15}N,^{13}C-labeled proteins are available.[20] This labeling pattern preserves many of the important features of perdeuteration with respect to relaxation benefits, while maintaining a critical number of protons at important side-chain positions for further structural and dynamics studies. Finally, it seems quite clear that, for high molecular weight proteins, strategies based on the complete assignment of side chains will fail due to issues involving poor sensitivity and resolution.

In a recent publication, we have discussed the difficulties associated with the assignment of Ile, Leu, and Val methyls in large proteins where both sensitivity and resolution issues are critical.[21] In particular, the nonlinearity associated with these branched residues compromises the efficiency of the transfer of magnetization from the methyl at the periphery of the side chain to the backbone $^{13}C^{\alpha}$ using the popular TOCSY-based schemes which are employed so effectively in applications involving small proteins.[22−27] This is due to the fact that at the branch point the magnetization can go in two separate directions; only one leads to productive transfer to the backbone. We showed further that, in the case of Ile residues, the use of COSY-type magnetization transfers results in substantial sensitivity gains because it is possible to direct the magnetization along the desired trajectory to the backbone.[21] Sensitivity gains on the order of 50% were realized in an application to maltose binding protein at 5 °C (correlation time of 46 ns). The assignment of the methyls of Leu and Val remains an issue, however, because it is not possible to "steer" the magnetization from $^{13}CH_3$ to $^{13}C^{\alpha}$ using either COSY or TOCSY approaches for these residues in proteins with uniform ^{13}C labeling.

We present new labeling methods that linearize spin systems of Leu and Val by replacing one of the $^{13}CH_3$ methyls by a $^{12}CD_3$ group and show that, for the 723 residue enzyme malate synthase G (MSG, molecular tumbling time of 37 ns at 37 °C),[5] good-quality COSY-based methyl−HN correlation spectra can be obtained for these residues in addition to Ile. New correlation experiments which relay magnetization in an out-and-back manner from methyl spins to aliphatic carbons (HMCM[CG]-CBCA) and from methyls to carbonyls and back (Val-HMCM-(CBCA)CO and Ile,Leu-HMCM(CGCBCA)CO) are presented and applied to MSG. Using data from both classes of experiments, along with the HN, ^{15}N, ^{13}CO, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ assignments obtained previously[5] with 4D TROSY-based NMR spectroscopy,[28] we have been able to obtain near complete assignments of the 276 Ile (δ1 only), Leu, and Val methyls in the protein.

Materials and Methods

Sample Preparation. A methyl protonated {I(δ1 only), L($^{13}CH_3$, $^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of MSG was obtained by protein overexpression from a culture of *E. coli* BL21(DE3)pLysS cells transformed with the plasmid pMSG-B encoding all of the residues of MSG with (i) the substitution of Ser 2 with an alanine, (ii) the addition of an N-terminal methionine, and (iii) the addition of an eight-residue C-terminal hexa-histidine tag (−LE−HHHHHH). The protein was expressed in 2 L of D_2O M9 medium as described in detail previously[5] using 2 g/L of U-[^{13}C,^{2}H]-glucose (CIL, Andover, MA) as the main carbon source and 1 g/L of $^{15}NH_4Cl$ (CIL, Andover, MA) as the nitrogen source. One hour prior to induction, 120 mg of 2-keto-3,3-d_2-1,2,3,4-^{13}C-butyrate and 200 mg of 2-keto-3-methyl-d_3-3-d_1-1,2,3,4-^{13}C-butyrate (^{13}C-labeled α-ketoisovalerate deuterated at the β-position and with one of the two methyl groups − $^{12}CD_3$; see Figure 1) were added to the growth medium. The growth was continued for 7 h at 37 °C. The sodium salt of 2-keto-1,2,3,4-^{13}C-butyric acid was obtained from Isotec (Miamisburg, OH), and the 3,3-[^{1}H] positions were quantitatively exchanged to ^{2}H by incubation of a 2.7 mM solution of α-ketobutyrate in 99.9% D_2O at 45 °C, pH 10.5 (uncorrected) for 20 h, following the procedure of Gardner et al.[29] The sodium salt of 2-keto-3-methyl-d_3-1,2,3,4-^{13}C-butyrate was custom synthesized at Isotec (Miamisburg, OH), and the 3-[^{1}H] position was exchanged to ^{2}H by incubation of a 25 mM solution in 99.9% D_2O at 45 °C, pH 12.4 (uncorrected) for 3 h, as described by Goto et al.[20] Both reactions were followed to completion by ^{1}H NMR. After initial purification on a nickel affinity column (Ni-NTA Agarose, Qiagen), the protein was fully denatured and refolded in vitro to protonate all amide positions. The details of the in vitro refolding protocol and subsequent purification steps were described previously.[5] The NMR sample was 0.9 mM in protein, 92% H_2O/8% D_2O, 25 mM sodium phosphate, pH 7.1, 20 mM $MgCl_2$, 0.05% NaN_3, 0.1 mg/mL Pefabloc, and 5 mM DTT, 37 °C. Protonation of non-ILV methyl groups was not observed, and isotopomers other than the CH_3 variety were not detected. A methyl protonated {L($^{13}CH_3$,$^{13}CH_3$),V($^{13}CH_3$,$^{13}CH_3$)} U-[^{15}N,^{2}H] sample of MSG in D_2O was also prepared (0.8 mM) as described above, except that 2-keto-3-methyl-^{13}C-butyric-3-d_1-4-^{13}C acid (Isotec, OH) was used as a precursor (i.e., α-ketoisovalerate with both methyls of the $^{13}CH_3$ type, all other carbons ^{12}C) and [^{12}C,^{2}H]-glucose was used as the main carbon source. All of the above compounds are commercially available from Isotec.

NMR Spectroscopy. NMR experiments were performed on an 800 MHz four-channel Varian Inova spectrometer equipped with a pulsed-field gradient conventional room-temperature triple resonance probe. 2D ^{1}H−^{13}C constant time (CT) HMQC experiments were acquired with the constant time period adjusted to 28 ms ($1/J_{CC}$) and (116, 768) complex points in the (^{13}C, ^{1}H) dimensions with corresponding acquisition times of (27.2, 64 ms) and a net experimental time of 50 min (relaxation delay of 1.5 s). 3D Ile,Leu-(HM)CM(CGCBCA)NH and 3D Val-(HM)CM(CBCA)NH experiments providing correlations

(14) Gardner, K. H.; Rosen, M. K.; Kay, L. E. *Biochemistry* **1997**, *36*, 1389−1401.
(15) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627−636.
(16) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212.
(17) Metzler, W. J.; Wittekind, M.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *J. Am. Chem. Soc.* **1996**, *118*, 6800−6801.
(18) Nicholson, L. K.; Kay, L. E.; Baldisseri, D. M.; Arango, J.; Young, P. E.; Bax, A.; Torchia, D. A. *Biochemistry* **1992**, *31*, 5253−5263.
(19) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8*, 932−935.
(20) Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 369−374.
(21) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 5701−5706.
(22) Montelione, G. T.; Lyons, B. A.; Emerson, S. D.; Tashiro, M. *J. Am. Chem. Soc.* **1992**, *114*, 10974−10975.
(23) Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. *J. Biomol. NMR* **1993**, *3*, 225−231.
(24) Grzesiek, S.; Anglister, J.; Bax, A. *J. Magn. Reson., Ser. B* **1993**, *101*, 114−119.
(25) Lyons, B. A.; Montelione, G. T. *J. Magn. Reson., Ser. B* **1993**, *101*, 206−209.
(26) Gardner, K. H.; Konrat, R.; Rosen, M. K.; Kay, L. E. *J. Biomol. NMR* **1996**, *8*, 351−356.
(27) Lin, Y.; Wagner, G. *J. Biomol. NMR* **1999**, *15*, 227−239.
(28) Yang, D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2571−2575.
(29) Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 7599−7600.

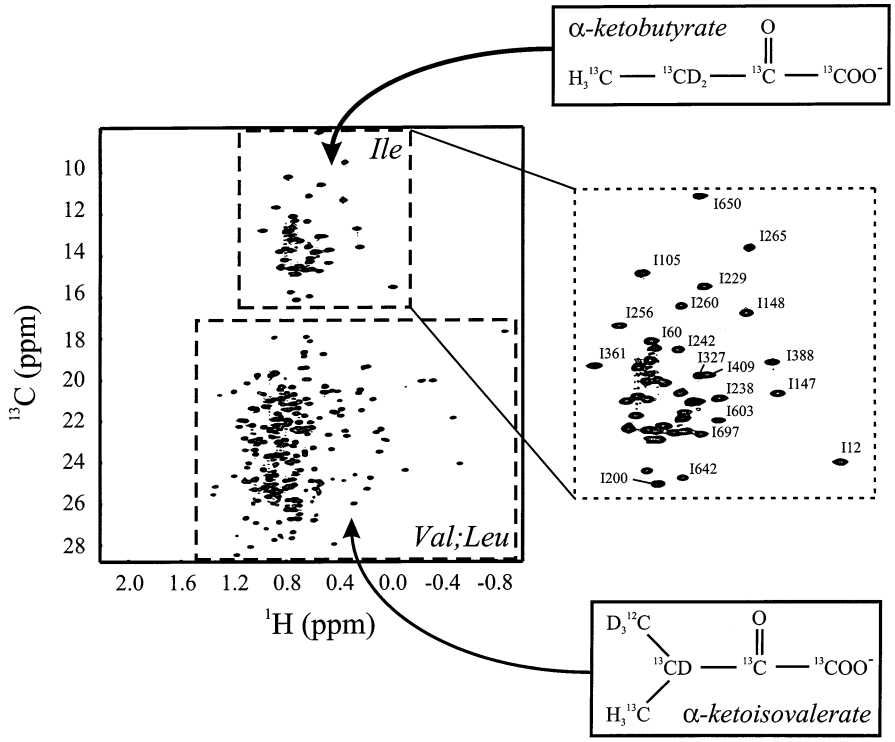

Figure 1. ^{1}H−^{13}C CT-HMQC spectrum of the methyl region of a methyl protonated {I(δ1 only), L(^{13}CH$_3$,^{12}CD$_3$),V(^{13}CH$_3$,^{12}CD$_3$)} U-[^{15}N,^{13}C,^{2}H] sample of MSG recorded at 800 MHz, 37 °C, with a 28 ms constant time period. The biosynthetic precursors used for production of methyl-protonated Ile (δ1 only), Leu, and Val with the methyl labeling pattern described in the text are shown. Assignments for selected Ile residues are indicated.

of the form [$\Omega_{Cm}(i)$, $\Omega_N(i)$, $\Omega_{HN}(i)$] (and in some cases [$\Omega_{Cm}(i)$, $\Omega_N(i+1)$, $\Omega_{HN}(i+1)$]) were recorded with (56, 40, 768) and (28, 42, 768) complex points in the (^{13}C, ^{15}N, ^{1}HN) dimensions, respectively, with corresponding acquisition times of (13.3, 17.2, 64 ms) and (13.4, 18.1, 64 ms). A relaxation delay of 1.35 s was used in both experiments along with 48 and 32 scans/FID, giving rise to net acquisition times of 117 and 93 h, respectively. The 3D Ile,Leu-(HM)CM(CGCBCA)NH pulse scheme was adapted from the Ile-(HM)CM(CGCBCA)NH described earlier (and referred to previously as Ile-(H)C(CA)NH-COSY)[21] by (i) omission of the Ile δ1-selective RE-BURP pulse,[30] (ii) adjustment of the β-selective RE-BURP pulse, and (iii) modification of phase-cycling as indicated in the legend to Figure 2a. The 3D Val-(HM)CM(CBCA)NH experiment was obtained by a trivial modification to the Ile,Leu sequence (see legend to Figure 2). 3D Ile,Leu-HM-(CMCGCBCA)NH and 3D Val-HM(CMCBCA)NH experiments providing correlations of the form [$\Omega_{Hm}(i)$, $\Omega_N(i)$, $\Omega_{HN}(i)$] (and in some cases [$\Omega_{Hm}(i)$, $\Omega_N(i+1)$, $\Omega_{HN}(i+1)$]) were both recorded with (18, 40, 768) complex points in the (^{1}H, ^{15}N, ^{1}HN) dimensions with corresponding acquisition times of (14.2, 17.2, 64 ms). Relaxation delays of 1.35(1.30) s were used along with 96(80) scans/FID, giving rise to net acquisition times of 113(95) h for 3D Ile,Leu-HM(CMCGCBCA)-NH (3D Val-HM(CMCBCA)NH). Selection of the ^{1}H−^{15}N TROSY component in all HN-detected NMR experiments was achieved passively by relying exclusively on the efficient relaxation of the anti-TROSY component during the transfer delays in the pulse schemes.[28,31]

The HMCM[CG]CBCA experiment, providing correlations of the form [$\Omega_{Caliph}(i)$, $\Omega_{Cm}(i)$, $\Omega_{Hm}(i)$], was recorded with (60, 50, 768) complex points in the ($^{13}C_{aliph}$, $^{13}C_m$, $^{1}H_m$) dimensions with corresponding acquisition times of (5.13, 13.6, 64 ms). A relaxation delay of 1.0 s was used along with 16 scans/FID, giving rise to a net acquisition time of 59 h. The Ile,Leu-HMCM(CGCBCA)CO and Val-HMCM-(CBCA)CO experiments (correlations of the form [$\Omega_{CO}(i)$, $\Omega_{Cm}(i)$, $\Omega_{Hm}(i)$]) were both acquired with (16, 50, 768) complex points in the (^{13}CO, $^{13}C_m$, $^{1}H_m$) dimensions with corresponding acquisition times of (7.7, 13.6, 64 ms). A relaxation delay of 1.2 s along with 48 and 16 scans/FID resulted in net acquisition times of 58 and 21 h, for Ile,-Leu-HMCM(CGCBCA)CO and Val-HMCM(CBCA)CO, respectively.

A [^{13}C-F_1, ^{13}C-F_2]-edited NOESY (30 ms mixing time) was recorded on a methyl protonated {L(^{13}CH$_3$,^{13}CH$_3$),V(^{13}CH$_3$,^{13}CH$_3$)} U-[^{15}N,^{2}H] sample of MSG in D_2O to confirm the assignments obtained via through-bond experiments and, in a number of cases where assignments for only 1 of the 2 methyls of Leu were available, to obtain assignments of the second methyl. Correlations of the form [$\Omega_{Cm}(i)$, $\Omega_{Cm}(j)$, $\Omega_{Hm}(j)$] were obtained, where methyl protons i and j are proximal, with (70, 70, 768) complex points and acquisition times of (30, 30, 64 ms). A relaxation delay of 1.3 s was employed, along with 8 scans/FID for a net acquisition time of 63 h.

All NMR spectra were processed using the suite of programs provided in the NMRPipe/NMRDraw software package.[32] Briefly, the ^{15}N time domains of all of the HN-detected spectra were doubled using mirror image linear prediction,[33] before apodization with a squared cosine window function, and subsequent Fourier transformation. The methyl carbon ($^{13}C_m$) and proton ($^{1}H_m$) indirect-detected time domains of all of the spectra were doubled using mirror-image linear prediction and apodized with cosine window functions. The nonconstant time aliphatic ($^{13}C_{aliph}$) and carbonyl (^{13}CO) dimensions of HMCM[CG]-CBCA, Val-HMCM(CBCA)CO, and Ile,Leu-HMCM(CGCBCA)CO data sets as well as both methyl carbon dimensions of the NOESY data set were doubled using forward−backward linear prediction.[34] Linear prediction in a given dimension was performed only after all of the other spectral dimensions were transformed as described by Kay et al.[35] The transformed data sets were reduced to include only the

(30) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93−141.
(31) Yang, D.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 3−10.
(32) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(33) Zhu, G.; Bax, A. *J. Magn. Reson.* **1990**, *90*, 405−410.
(34) Zhu, G.; Bax, A. *J. Magn. Reson.* **1992**, *98*, 192−199.
(35) Kay, L. E.; Ikura, M.; Zhu, G.; Bax, A. *J. Magn. Reson.* **1991**, *91*, 422−428.

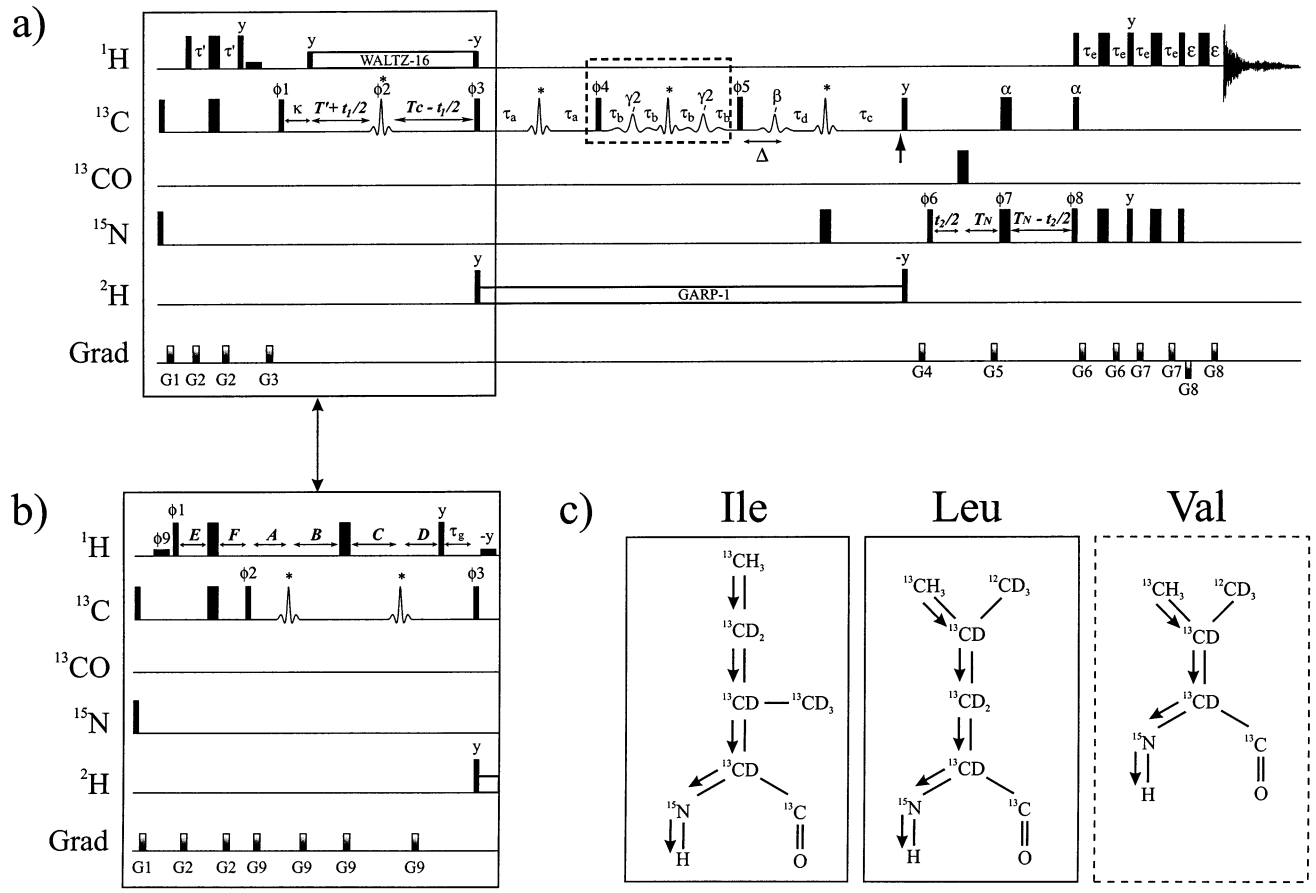

Figure 2. Methyl−HN correlation experiments for methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] protein samples. Omission of the pulse scheme element enclosed in the dashed rectangle in (a) provides correlations for Val (Val-(HM)CM(CBCA)NH), while the complete scheme is employed for Ile and Leu (Ile,Leu-(HM)CM(CGCBCA)NH). All narrow (wide) rectangular pulses are applied with flip angles of 90° (180°) along the *x*-axis unless indicated otherwise. The ^{1}H and ^{15}N carriers are positioned at 4.7 (water) and 119 ppm, respectively. All proton pulses are applied with a field strength of 40 kHz, with the exception of the ^{1}H WALTZ-16[42] decoupling field and flanking pulses which use a 7.0 kHz field and the water flip-back pulses ($\sim$1.2 ms rectangular pulses). All $^{15}N(^{2}H)$ pulses employ a 6.9(1.9) kHz field, with ^{2}H decoupling achieved using a 0.6 kHz GARP-1[43] field. The ^{13}C carrier is set to 20 ppm until prior to ^{13}C 90° pulses of phase ϕ1 (a) and ϕ2 (b) when it is moved to 40 ppm and then switched to 58 ppm prior to the ^{15}N 90° pulse of phase ϕ6. Rectangular 90° and 180° ^{13}C pulses applied prior to gradient g4 use the highest available power, while ^{13}C (labeled with "α") and ^{13}CO 90° (180°) pulses are applied thereafter with a field of $\Delta/\sqrt{15}$ ($\Delta/\sqrt{3}$), where Δ is the difference (Hz) between $^{13}C^{\alpha}$ (58 ppm) and ^{13}CO (176 ppm) shifts.[44] RE-BURP[30] shaped pulses indicated by "β" (1.3 ms, bandwidth of $\pm$1.5 kHz, centered at 38.5 ppm for Ile,Leu; 1.88 ms, bandwidth of $\pm$980 Hz, centered at 31.5 ppm for Val) and "γ2" (3 ms, bandwidth of $\pm$600 Hz, centered at 16 ppm) invert β carbons and γ2 carbons (Ile), respectively. RE-BURP pulses labeled with asterisks are centered at 38 ppm (300 μs). The phases of these pulses are carefully adjusted to maximize sensitivity. The vertical arrow at the end of delay τ_c indicates the position of application of the "β" Bloch-Siegert shift compensation pulse.[44] If the experiment is performed at a field other than 800 MHz, the duration of all selective pulses should be scaled by (800/y), where y is the spectrometer field. (a) Ile,Leu-(HM)CM(CGCBCA)NH. The phase cycling employed is: $\phi1 = y$; $\phi2 = 2(x)$, $2(y)$; $\phi3 = 4(y)$, $4(-y)$; $\phi4 = 8(y)$, $8(-y)$; $\phi5 = y$, $-y$; $\phi6 = 8(y)$, $8(-y)$; $\phi7 = 2(x)$, $2(-x)$; $\phi8 = x$; rec = $2(x,x,-x,-x)$, $2(-x,-x,x,x)$. Quadrature detection in F_1 is achieved with States-TPPI[45] of phase ϕ1, while quadrature in F_2 employs the enhanced sensitivity pulsed field gradient method,[46,47] where for each value of t_2 separate data sets are recorded with (g5, ϕ8) and (−g5, $\phi8 + 180°$). For each successive t_2 value, ϕ6 and the phase of the receiver are incremented by 180°. The delays used are $\tau' = 1.8$ ms, $\kappa = 1.52$ ms, $T_C = 7$ ms, $T' = T_C - \kappa$, $\tau_a = 5$ ms, $\tau_b = 2.5$ ms (the total duration between pulses of phase ϕ4 and ϕ5 is $4\tau_b$), $\tau_c = 10$ ms, $\tau_d = 7.3$ ms, $\Delta = 2.7$ ms, $T_N = 12.4$ ms, $\tau_e = 2.25$ ms, $\epsilon = 250$ μs. The durations and strengths of the pulsed field gradients are g1 = (1.0 ms, 4 G/cm), g2 = (0.2 ms, 8 G/cm), g3 = (1.0 ms, 4 G/cm), g4 = (0.5 ms, 10 G/cm), g5 = (1.25 ms, 30 G/cm), g6 = (0.2 ms, 5 G/cm), g7 = (0.25 ms, 3.5 G/cm), g8 = (62.5 μs, 28.7 G/cm). All gradients are applied along the *z*-axis. (b) The initial part of the pulse scheme for assignment of methyl protons: Ile,Leu-HM(CMCGCBCA)NH and Val-HM(CMCBCA)NH (omit the portion in dashed rectangle). The delays and evolution periods are $A = T_C/2 - t_1/4$; $B = (T_C + \tau_g)/2 - t_1/4$; $C = (T_C + \tau_g)/2 + t_1/4$; $D = T_C/2 + t_1/4$; $E = \tau' + t_1'/2$; $F = \tau' - t_1'/2$, where $\tau' = 1.8$ ms; $T_C = 5.7$ ms; $\tau_g = 1.52$ ms. The value of $t_1' = 0$ for $A \geq 0$; when $A = 0$, the value of t_1 is set to $2T_C$ and subsequently t_1' is incremented by 1 dwell time for each successive t_1 point. Only when $F = 0$ does the duration of the pulse scheme increase. The phase cycling employed is $\phi1 = x$, $-x$; $\phi2 = 2(y)$, $2(-y)$; $\phi3 = 4(y)$, $4(-y)$; $\phi4 = 8(y)$, $8(-y)$; $\phi5 = y$, $-y$; $\phi6 = 8(y)$, $8(-y)$; $\phi7 = 2(x)$, $2(-x)$; $\phi8 = x$; $\phi9 = -x$, x; rec = $2(x,-x,-x,x)$, $2(-x,x,x,-x)$. Quadrature detection in F_1 is achieved with States-TPPI of phases ϕ1 and ϕ9. The durations and strengths of the pulsed field gradients are g2 = (50 μs, 8 G/cm), g9 = (50 μs, 18 G/cm). All of the other pulses, delays, gradient strengths, and durations are as in scheme (a). (c) Schematic diagrams of the intraresidue magnetization transfer steps employed for Ile (δ1 protonated), Leu and Val (with only one methyl group of the $^{13}CH_3$ variety). The pulse sequence code is available from the authors upon request.

regions of interest and analyzed using the NMRView program[36] in conjunction with tcl/tk scripts written in-house.

Results and Discussion

Several years ago, our laboratory introduced a labeling scheme for the production of Ile (δ1), Leu, Val-methyl protonated, highly deuterated ^{15}N,^{13}C-labeled proteins in which protonation is achieved through the addition of a pair of precursors to the growth medium, [3-^{2}H],^{13}C α-ketoisovalerate (methyl protonation of Val and Leu) and [3,3-^{2}H],^{13}C α-ketobutyrate (δ1 methyl protonation of Ile).[20,29] Proteins prepared in this manner retain a high level of deuteration, facilitating

(36) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603−614.

assignments of both backbone nuclei and side-chain methyl positions. As the size of proteins studied by NMR has increased, so too have the difficulties with assignment of the side-chain methyls. For example, near complete assignments of the methyl ^{1}H and ^{13}C chemical shifts were obtained for a complex of maltose binding protein (MBP) and β-cyclodextrin (37 °C, 370 residues, 17 ns tumbling time)[37] as well as the membrane protein OmpX dissolved in micelles (30 °C, 148 residues, 21 ns tumbling time)[38] using conventional TOCSY-based magnetization transfer experiments. In contrast, such experiments failed in applications involving MBP at 5 °C (correlation time of 46 ns).[21] In this case, the large effective size of the system and the inherent losses in the magnetization transfer between methyl and alpha carbons due to the branched structure of Ile, Leu, and Val residues are extremely detrimental. For example, in the case of an Ile residue, the competing pathway $^{13}C^{methyl} \rightarrow {}^{13}C^{\gamma 2}$ decreases the net transfer of magnetization to the backbone ($^{13}C^{methyl} \rightarrow {}^{13}C^{\alpha}$) by a factor of close to 2.5. An alternative approach for Ile residues has been developed on the basis of a directed COSY-transfer of the signal up the side chain that exploits the favorable distribution of ^{13}C chemical shifts at the branch point so that the "dead end" pathway to $^{13}C^{\gamma 2}$ can be blocked.[21] Significant improvements (50%) in sensitivity relative to TOCSY schemes have been demonstrated in applications involving MBP at 5 °C. This same strategy is, unfortunately, not possible for Leu and Val, and we have, therefore, modified our labeling approach slightly to address this problem.

Our new method effectively linearizes the spin systems of Leu and Val by introducing a $^{12}CD_3$ group at one of the two methyl positions, while the second is ^{13}C-labeled and protonated. In this approach, α-ketoisovalerate with one methyl $^{13}CH_3$ and one $^{12}CD_3$ is used as a precursor, along with α-ketobutyrate as before. Figure 1 shows an 800 MHz ^{1}H−^{13}C CT-HMQC spectrum of a methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$), V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of MSG prepared using the biosynthetic precursors mentioned above and described in Materials and Methods. The 723-residue MSG has a high content of Ile, Leu, and Val residues: 44 Ile ($^{13}C^{\delta 1}$ only protonated), 46 Val, and 70 Leu (276 methyl groups in total). Most of these methyls are located in the hydrophobic core of the molecule and represent potentially important sources of structural and dynamical information.[39] The assignment of methyl groups in such a large system is, however, challenging not only due to the issue of sensitivity but also because of the large number of such moieties and, therefore, the need for high resolution.

It is important to emphasize at the outset that $^{13}CH_3$ groups are present at only one of two methyl positions, for Leu and Val, effectively reducing their concentration 2-fold. Indeed, intensities of 136 Leu and Val cross-peaks in the CT-HMQC spectrum are on average 2.3-fold lower than intensities of 39 Ile δ1 methyl cross-peaks that could be accurately quantified. Nevertheless, as shown below, the proposed isotopic labeling method leads to substantial gains in sensitivity and in spectral simplicity relative to approaches in which both methyls are of the $^{13}CH_3$ type. Of note, because the α-ketoisovalerate that is used as a precursor for Leu and Val in this study is a racemic mixture of [*R*-$^{13}CH_3$, *S*-$^{12}CD_3$] and [*S*-$^{13}CH_3$, *R*-$^{12}CD_3$] enantiomers, the $^{13}CH_3$ and $^{12}CD_3$ groups are incorporated nonstereospecifically into the *R* and *S* methyl positions of Leu and Val.

Assignments of Ile, Leu, and Val Methyl Groups in MSG from HN-Detected Experiments. The 3D (HM)CM(CGCBCA)NH pulse scheme designed earlier for the assignment of Ile δ1 methyl groups in methyl protonated, highly deuterated large proteins[21] has been modified as described in Materials and Methods to provide correlations of the form [$\Omega_{Cm}(i)$, $\Omega_N(i)$, $\Omega_{HN}(i)$] (and in some cases [$\Omega_{Cm}(i)$, $\Omega_N(i+1)$, $\Omega_{HN}(i+1)$]) for both Ile and Leu residues. Figure 2a illustrates this Ile,Leu-(HM)CM(CGCBCA)NH sequence. A similar scheme for recording correlations of Val can be obtained by omission of the element enclosed in the dashed rectangle in the figure (Val-(HM)CM(CBCA)NH). The magnetization transfer steps in the experiment have been described in detail previously for the case of Ile[21] and will not be repeated here. We have, however, included a diagram of the intraresidue magnetization flow for Ile, Leu, and Val residues in Figure 2c. Importantly, no evolution of magnetization occurs due to passive *J*-couplings between $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ (Val) or $^{13}C^{\gamma}$ and $^{13}C^{\delta}$ (Leu) due to the absence of ^{13}C enrichment in the second methyl group of these residues, while *J*-evolution from the $^{13}C^{\beta}-{}^{13}C^{\gamma 2}$ coupling in Ile is refocused. The "linearization" of the magnetization transfer pathways in all three cases provides a gain of a factor of 2.6 in sensitivity (neglecting relaxation) relative to the situation for fully ^{13}C-labeled side chains (Leu, Val) or where selective $^{13}C^{\gamma 2}$ pulses (Ile) are not employed. The 2-fold decrease in concentration of Leu,Val methyl groups is, therefore, more than balanced by the gain in sensitivity from the labeling scheme employed.

Figure 2b shows the modification of the pulse sequence required for assignment of methyl protons (Ile,Leu-HM(CM-CGCBCA)NH and Val-HM(CMCBCA)NH experiments). The experiment is designed to record ^{1}H evolution in a (semi-)constant time manner so as to minimize sensitivity losses. For example, for ^{1}H evolution times (t_1) less than approximately $A + B + C + D$ in Figure 2b, durations E and F are set to $\sim\frac{1}{(4J_{HC})}$, where J_{HC} is the ^{1}H−^{13}C coupling constant, and evolution proceeds in a constant time mode. For larger t_1 values, simultaneous movement of the first ^{1}H and ^{13}C 180° pulses during the initial ^{1}H−^{13}C transfer period ensures that evolution can proceed for an additional $\sim\frac{1}{(2J_{HC})}$, without increasing the length of the experiment. Only if longer acquisition times than approximately $A + B + C + D + E + F$ are required does the length of the sequence increase (for details, see legend to Figure 2); in our applications, we have used $t_{1,max}$ values that are within this duration. In addition to maximizing sensitivity, the constant time nature of the t_1 evolution period in the HM(CM[CG]CBCA)NH sequences optimizes resolution in this dimension, an important consideration in the application to MSG.

The main strength of the HN-detected approach is the excellent dispersion of chemical shifts in the HN and ^{15}N dimensions of the (HM)CM([CG]CBCA)NH and HM(CM[CG]-CBCA)NH data sets. Even for a protein as large as MSG, the assignment of observed resonances is unambiguous when HN and ^{15}N chemical shifts of Ile, Leu, and Val residues are

(37) Gardner, K. H.; Zhang, X.; Gehring, K.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 11738−11748.
(38) Hilty, C.; Fernandez, C.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2002**, *23*, 289−301.
(39) Howard, B. R.; Endrizzi, J. A.; Remington, S. J. *Biochemistry* **2000**, *39*, 3156−68.

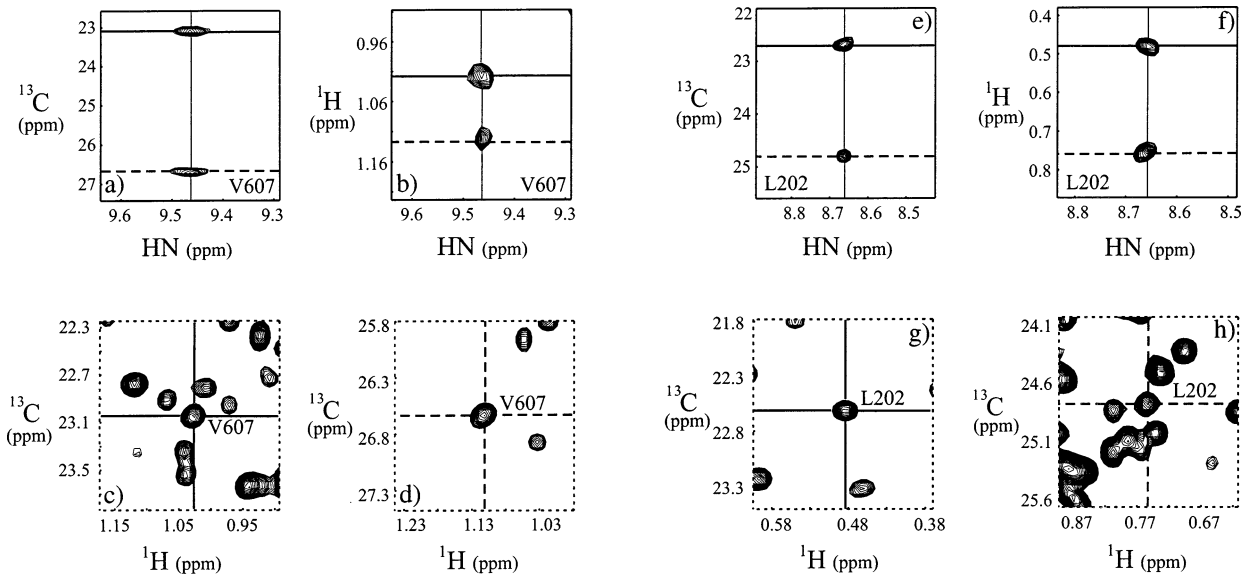

Figure 3. Selected regions from $^{13}C^{m}-HN$ planes of HN-detected 3D data sets recorded on a methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of MSG at 800 MHz, 37 °C. Correlations establishing the $^{13}C_m(^{1}H_m)$ chemical shifts of V607 are shown in (a) and (b), respectively, and are linked together using a CT-HMQC spectrum, as in (c) and (d). The $^{13}C_m$ and $^{1}H_m$ chemical shifts of cross-peaks from the 3D spectra in panels (a,b) indicated with solid and dashed lines correspond to the shifts of cross-peaks in CT-HMQC spectra (c,d) that are also indicated with solid and dashed lines, respectively, illustrating how the proton−carbon correlation information missing from the 3D (HM)CM([CG]CBCA)NH and HM(CM[CG]CBCA)NH data sets can be retrieved from the 2D $^{1}H-^{13}C$ correlation maps. Panels e−h illustrate the procedure for assigning methyl shifts for L202.

available. There are, unfortunately, several disadvantages as well. First, there is, inherently, no proton−carbon correlation information for CH_3 groups in HN-detected spectra.[37,40] This is an important consideration in linking assigned methyl proton and carbon resonances to a specific methyl group for Leu and Val (residues with pairs of methyls; recall that only one of the two Ile methyls (δ1) is observed using our labeling scheme). In principle, it is possible to design 4D experiments which circumvent this problem; however, sensitivity and resolution considerations preclude such experiments in the case of MSG. Figure 3a−h illustrates selected regions of data sets from MSG obtained using HN-detected experiments. The $^{1}H-^{13}C$ correlations for each methyl in a Leu or Val residue must be established with the help of a $^{1}H-^{13}C$ correlation map. Although this is straightforward for most residues in MSG, as illustrated in Figure 3, the substantial overlap in the Leu−Val region of the CT-HMQC spectrum makes the establishment of unambiguous proton−carbon correlations difficult in some cases. A second and more serious disadvantage of the HN-detected approach is the low sensitivity of some of the methyl−HN correlations in applications involving large proteins. Even after prolonged acquisition times (see Materials and Methods), only 50% of Leu methyls could be reliably assigned from the HN-detected data sets. The higher intrinsic sensitivity of the data for Val and the 2-fold higher concentration of Ile δ1 $^{13}CH_3$ methyl groups facilitated the assignment of 80 Val (87%) and 40 Ile (97%) methyl groups for which the amides had been previously assigned. Clearly, an alternative more sensitive and robust approach is required for complete methyl assignments in large proteins.

Assignments of Ile, Leu, and Val Methyls in MSG from Out-and-Back Methyl-Detected Experiments. One of the main reasons for the low sensitivity of the HN detected experiments is that it is not possible to transfer polarization from all three methyl protons up the side chain. The periods denoted by κ and τ_g in Figure 2a and b, respectively, are set to optimize transfer, but the optimum efficiency is only 38%, effectively throwing away polarization from 2 of the 3 protons. An alternative strategy is one in which the total polarization from the methyl protons is utilized, and this can be achieved most readily in out-and-back type experiments in which magnetization both originates from and is transferred back to methyl protons. A detailed analysis shows that such experiments are significantly better in terms of both sensitivity and resolution than corresponding TOCSY-based schemes in which magnetization originating on aliphatic carbons is transferred to methyl groups for detection. Figure 4a,b shows the methyl-detected out-and-back pulse schemes that have been developed for the assignment of methyl groups in proteins, with the magnetization flow illustrated for each of the three side chains in Figure 4c. The flow between points *a* and *b* in Figure 4a can be summarized by

$$2I_zC_y^{m} \xrightarrow{2T_C,90_{\phi 2}} 4I_zC_z^{m}C_x^{\gamma} \xrightarrow{2T_C,90_{\phi 3}} 4I_zC_z^{\gamma}C_x^{\beta} \xrightarrow{4\tau_b,90_{\phi 4}} 8I_zC_x^{\gamma}C_y^{\beta}C_x^{\alpha} - 4I_zC_x^{\gamma}C_z^{\beta} + 2I_zC_y^{\beta} + 4I_zC_z^{\beta}C_x^{\alpha} \quad (1)$$

for Ile ($\gamma = \gamma$1) and Leu residues where I_a and $C_a{}^b$ are the $a = \{x,y,z\}$ components of methyl proton and carbon ($b = \{\alpha,\beta,\gamma\}$) magnetization, respectively. For simplicity, factors describing the efficacy of magnetization transfer during each of the delays have been omitted. As in the HN-detected set of experiments, magnetization is relayed up the side chain using COSY-type transfers, and it is, therefore, possible to refocus the $^{13}C^{\beta}-^{13}C^{\gamma 2}$ *J*-coupling in Ile side chains, as before. For the case of Val, the transfer between *a* and *b* is given by

$$2I_zC_y^{m} \xrightarrow{2T_C,90_{\phi 2}} 4I_zC_z^{m}C_x^{\beta} \xrightarrow{2T_C,90_{\phi 3}} 4I_zC_z^{\beta}C_x^{\alpha} \xrightarrow{4\tau_b,90_{\phi 4}} -4I_zC_x^{\beta}C_z^{\alpha} + 2I_zC_y^{\alpha} \quad (2)$$

(40) Uhrin, D.; Uhrinova, S.; Leadbeater, C.; Nairn, J.; Price, N. C.; Barlow, P. N. *J. Magn. Reson.* **2000**, *142*, 288−293.

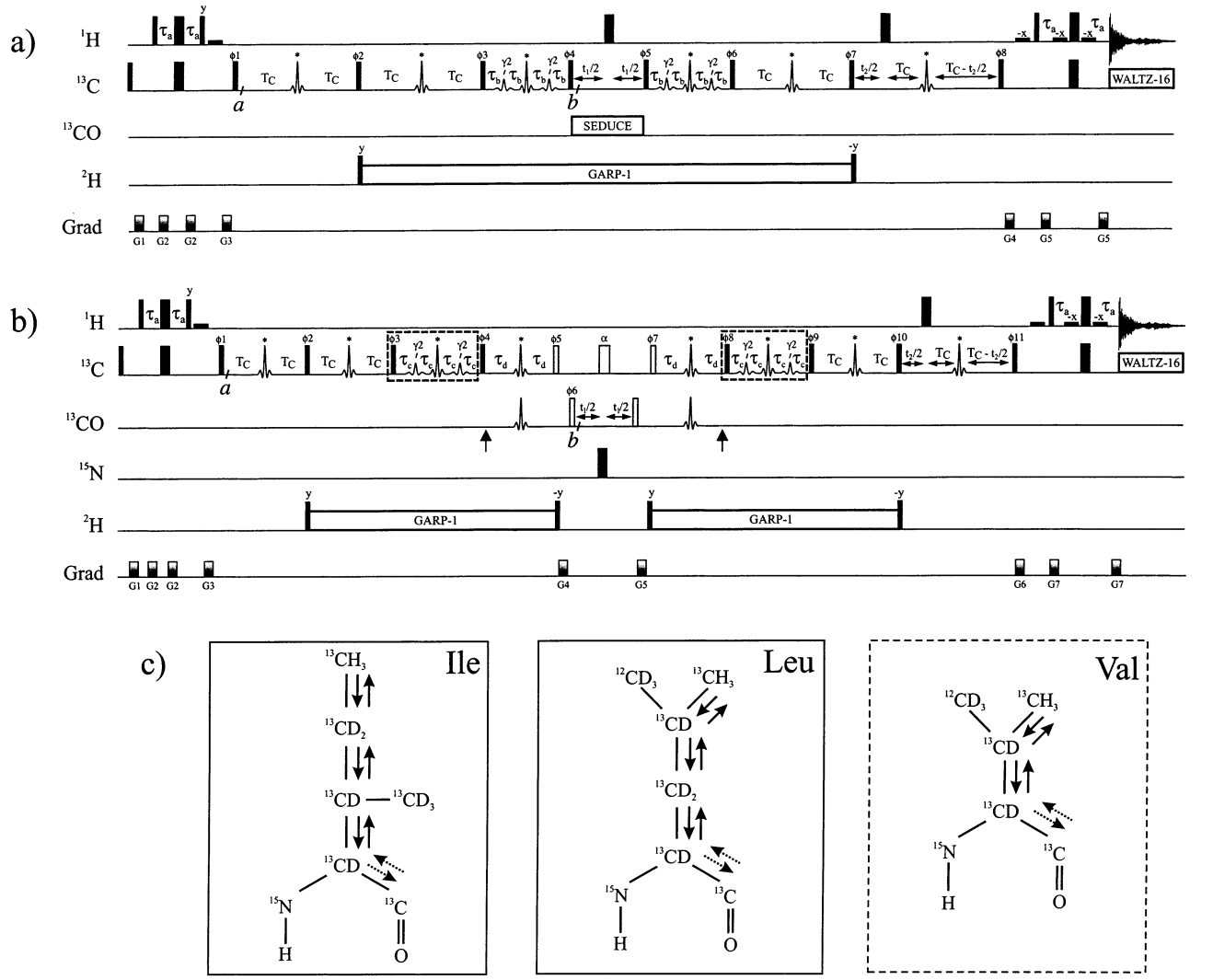

Figure 4. Pulse schemes of methyl-detected "out-and-back" experiments for assignment of methyl groups in methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] protein samples. Many of the details of the pulse schemes are similar to those described in the legend to Figure 2 and will be omitted here. (a) HMCM[CG]CBCA pulse scheme. The ^{13}C carrier is placed at 19.5 ppm (the center of the methyl region), switched to 40 ppm before the ^{13}C pulse with phase $\phi1$, and then returned to 19.5 ppm after the pulse with phase $\phi8$. SEDUCE decoupling[48] is implemented with 280 μs seduce-shaped pulses,[49] with 136 ppm cosine-modulation of the waveform. ^{13}C decoupling during acquisition is achieved using a 2.5 kHz WALTZ-16[42] field. Delays are $\tau_a = 1.8$ ms; $T_C = 7$ ms; $\tau_b = 1.75$ ms (note that $4\tau_b$ is the total duration between pulses of phase $\phi3$[$\phi5$] and $\phi4$[$\phi6$]). The phase cycling employed is $\phi1 = x, -x$; $\phi2 = \phi3 = \phi4 = 2(y), 2(-y)$; $\phi5 = \phi6 = \phi7 = 4(y), 4(-y)$; $\phi8 = x$; rec = $x, -x$. Quadrature in F_1 is achieved by incrementing the phases $\phi1$, $\phi2$, $\phi3$, and $\phi4$ by 90°, with 180° added to the phase of $\phi1$ and the receiver for each successive complex t_1 point,[45] while quadrature in F_2 is achieved by States-TPPI of phase $\phi8$. The durations and strengths of the pulsed field gradients are g1 = (1.0 ms, 5 G/cm), g2 = (0.3 ms, 10 G/cm), g3 = (1.0 ms, 12 G/cm), g4 = (1.4 ms, 12 G/cm), g5 = (0.5 ms, 5 G/cm). The ^1H 180° pulse in the center of the t_1 period is optional. (b) Ile,Leu-HMCM-(CGCBCA)CO pulse scheme. The Val-sequence is obtained by removing the elements enclosed in rectangles (dashed lines). The ^{13}C carrier is placed at 19.5 ppm (the center of the methyl region), switched to 40 ppm before the ^{13}C pulse with phase $\phi1$, to 176 ppm (the center of the ^{13}CO chemical shift region) before the pulse with phase $\phi6$, and then returned to 40 ppm prior to the pulse with phase $\phi7$ and 19.5 ppm after the pulse with phase $\phi11$. 90° (180°) ^{13}C pulses shown with open rectangles are applied with a field strength of $\Delta/\sqrt{15}$ ($\Delta/\sqrt{3}$), where Δ is the difference (in Hz) between $^{13}C^{\alpha}$ (58 ppm) and ^{13}CO (176 ppm) chemical shifts.[44] The pulses with phases $\phi5$ and $\phi7$ and the pulse marked "α" are applied at 58 ppm by phase modulation of the carrier.[50,51] The vertical arrows at the start and end of the $2\tau_d$ delays indicate the positions of application of the ^{13}CO Bloch-Siegert shift compensation pulses.[44] The ^{13}CO shaped pulses are 300 μs RE-BURP pulses applied at 176 ppm by phase-modulation of the carrier. It is critical that the phases of the high power ^{13}C RE-BURP pulses (marked with asterisks) be adjusted to maximize sensitivity. Typically, we find that the first three and last three RE-BURP pulses have phases close to 0°, while the pulse in the first (second) $2\tau_d$ period has phase 15° (8°) for the power levels that we employ (800 MHz). Delays are $\tau_a = 1.8$ ms; $T_C = 7$ ms; $\tau_c = 3.5$ ms ($4\tau_c$ is the total duration between pulses $\phi3/\phi4$ and $\phi8/\phi9$); $\tau_d = 4.75$ ms. The phase-cycle employed is $\phi1 = x, -x$; $\phi2 = \phi3 = \phi4 = 2(y), 2(-y)$; $\phi5 = y$; $\phi6 = 4(x), 4(-x)$; $\phi7 = 8(y), 8(-y)$, $\phi8 = \phi9 = \phi10 = 4(y), 4(-y)$; $\phi11 = x$; rec = $(x,-x,x,-x)$, $2(-x,x,-x,x)$, $(x,-x,x,-x)$. Quadrature in F_1, F_2 is achieved by States-TPPI of phases $\phi6$ and $\phi11$, respectively. Durations and strengths of the pulsed-field gradients are g1 = (1.0 ms, 5 G/cm), g2 = (0.3 ms, 10 G/cm), g3 = (1.0 ms, 12 G/cm), g4 = (0.8 ms, 15 G/cm), g5 = (0.6 ms, 8 G/cm), g6 = (1.4 ms, 12 G/cm), g7 = (0.5 ms, 5 G/cm). (c) Schematic diagrams of the magnetization transfer steps along the carbon skeletons of Ile ($\delta1$ protonated), Leu, and Val (with a single methyl of the $^{13}CH_3$ variety) side chains. The final transfer from $^{13}C^{\alpha}$ to ^{13}CO shown with dashed lines occurs only in pulse scheme (b). In the case where Ile, Leu, and Val methyl protonated samples with full ^{13}C-labeling are used, the phases $\phi3$ and $\phi6$ should be incremented by $\pi/2$ in (a).

Magnetization is subsequently allowed to evolve during the t_1 period and then transferred back to the methyl carbon by reversal of the above pathway. It can be shown that the four terms in scheme (1) give rise to intraresidue correlations (Ile, Leu) of the form $-1/4[-\Omega_{C\alpha} - \Omega_{C\beta} - \Omega_{C\gamma},\Omega_{Cm},\Omega_{Hm}]$, $-1/4[\Omega_{C\alpha} + \Omega_{C\beta} - \Omega_{C\gamma},\Omega_{Cm},\Omega_{Hm}]$, $-1/4[\Omega_{C\alpha} - \Omega_{C\beta} + \Omega_{C\gamma},\Omega_{Cm},\Omega_{Hm}]$, $-1/4[-\Omega_{C\alpha} + \Omega_{C\beta} + \Omega_{C\gamma},\Omega_{Cm},\Omega_{Hm}]$, $1[\Omega_{C\gamma},\Omega_{Cm},\Omega_{Hm}]$, $-1[\Omega_{C\beta},\Omega_{Cm},\Omega_{Hm}]$, $1[\Omega_{C\alpha},\Omega_{Cm},\Omega_{Hm}]$, where the coefficients

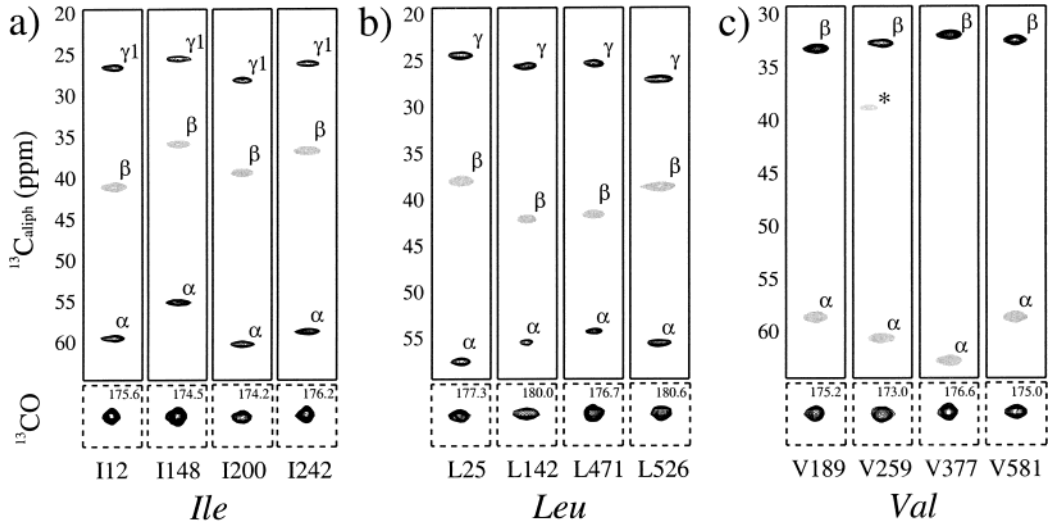

Figure 5. Selected $^{13}C_{aliph}-^{1}H_m$ and $^{13}CO-^{1}H_m$ strips from 3D HMCM[CG]CBCA (upper row, solid frame) and 3D HMCM([CG]CBCA)CO spectra (lower row, dashed-lined frame), respectively, for (a) Ile, (b) Leu, and (c) Val residues of methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] MSG at 800 MHz, 37 °C. Negative correlations ($^{13}C^{\beta}$ for Ile and Leu, $^{13}C^{\alpha}$ for Val) are shown in gray. Relevant ^{13}CO chemical shifts are indicated in the upper right corner of the HMCM([CG]CBCA)CO panels. The peak marked with an asterisk belongs to another spin-system and is more intense on an adjacent slice. Strips are labeled according to the identity of the methyl-containing residue.

preceding each of the square brackets indicate the relative intensities of the correlations ($4\tau_b$ is set to $\frac{1}{(4J_{CC})}$, where J_{CC} is the one bond aliphatic $^{13}C-^{13}C$ scalar coupling constant). The first four correlations above derive from the triple-spin term $8I_zC_x^{\gamma}C_y^{\beta}C_x^{\alpha}$, while the remaining three originate from $4I_zC_x^{\gamma}C_z^{\beta}$, $2I_zC_y^{\beta}$, and $4I_zC_z^{\beta}C_x^{\alpha}$. In the case of Val, coherences $4I_zC_x^{\beta}C_z^{\alpha}$, $2I_zC_y^{\alpha}$ (Scheme (2)) lead to correlations of the form $1[\Omega_{C\beta},\Omega_{Cm},\Omega_{Hm}]$, $-1[\Omega_{C\alpha},\Omega_{Cm},\Omega_{Hm}]$.

The magnetization flow from points *a* to *b* in the HMCM([CG]CBCA)CO experiments of Figure 4b can be summarized as

$$C_y^{m} \xrightarrow{2T_C,90_{\phi 2}} 2C_z^{m}C_x^{\gamma} \xrightarrow{2T_C,90_{\phi 3}} 2C_z^{\gamma}C_x^{\beta} \xrightarrow{4\tau_c,90_{\phi 4}} 2C_z^{\beta}C_x^{\alpha} \xrightarrow{2\tau_d}$$
$$2C_z^{\beta}C_x^{\alpha} + C_y^{\alpha} + 4C_z^{\beta}C_y^{\alpha}C_z' - 2C_x^{\alpha}C_z' \xrightarrow{90_{\phi 5},90_{\phi 6}} 2C_z^{\alpha}C_y' \quad (3)$$

and

$$C_y^{m} \xrightarrow{2T_C,90_{\phi 2}} 2C_z^{m}C_x^{\beta} \xrightarrow{2T_C,90_{\phi 4}} 2C_z^{\beta}C_x^{\alpha} \xrightarrow{2\tau_d}$$
$$2C_z^{\beta}C_x^{\alpha} + C_y^{\alpha} + 4C_z^{\beta}C_y^{\alpha}C_z' - 2C_x^{\alpha}C_z' \xrightarrow{90_{\phi 5},90_{\phi 6}} 2C_z^{\alpha}C_y'$$

for Ile,Leu-HMCM(CGCBCA)CO and Val-HMCM(CBCA)CO pulse schemes, respectively, where the I_z term in each of the operators has been omitted. Carbonyl magnetization (final terms listed in each of the schemes) evolves during the subsequent t_1 period and is returned to the originating methyl by reversal of the pathway, so that correlations of the form [$\Omega_{CO}(i)$, $\Omega_{Cm}(i)$, $\Omega_{Hm}(i)$] are obtained.

Figure 5 shows $^{1}H^{m}-^{13}C_{aliph}$ and $^{1}H^{m}-^{13}CO$ strips from HMCM[CG]CBCA and HMCM([CG]CBCA)CO data sets for selected Ile, Leu, and Val residues of MSG. Four frequencies ($^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{\gamma}$, and ^{13}CO) can be used for identification of methyl groups belonging to the same Leu residue, while $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO shifts are matched to obtain methyl pairs in Val. The sequence-specific assignments of Ile, Leu, and Val methyls have been made by matching three ^{13}C frequencies − $^{13}C^{\alpha}$, $^{13}C^{\beta}$ (from the HMCM[CG]CBCA data set), and ^{13}CO (from either Ile,Leu-HMCM(CGCBCA)CO or Val-HMCM(CBCA)CO) − to those available from earlier NMR studies of MSG,[5] taking into account the two-bond ($^{2}\Delta_{H\rightarrow D}$) isotope shift of ~0.3 ppm[1,41] at the Val $^{13}C^{\beta}$ position resulting from the different labeling approaches used in each study. One of the main advantages of the out-and-back experiments, in addition to their high sensitivity, is that they correlate proton−carbon methyl shifts directly, unlike the HN-detected data sets (see above). Note also that $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ cross-peaks of Val (Figure 5c) have opposite signs from those of Ile and Leu in the HMCM[CG]CBCA data set (see above), making it straightforward to distinguish between correlations belonging to different amino acids.

It is of interest to compare (at least theoretically) the gains associated with the labeling strategy introduced here relative to our previous approach where both methyls in Leu and Val are of the $^{13}CH_3$ variety. In the case of the HN-detected experiments, the 2-fold decrease in methyl concentration is compensated by the gain of ~2.6-fold in sensitivity due to "linearization" of the side chain using the present labeling scheme (see above) so that modest sensitivity increases of approximately 30% are expected with methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] samples. In the case of HMCM[CG]CBCA spectra, sensitivities of both Leu and Val correlations are calculated to be essentially independent of the methyl labeling scheme employed. (Note that phases $\phi 3$ and $\phi 6$ in Figure 4a must be incremented by $\pi/2$ in the case where Leu and Val methyls are both ^{13}C-labeled

(41) Venters, R. A.; Farmer, B. T.; Fierke, C. A.; Spicer, L. D. *J. Mol. Biol.* **1996**, *264*, 1101−1116.
(42) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *J. Magn. Reson.* **1983**, *52*, 335−338.
(43) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *64*, 547.
(44) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1990**, *89*, 496−514.
(45) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.
(46) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.
(47) Schleucher, J.; Sattler, M.; Griesinger, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1489−1491.
(48) McCoy, M. A.; Mueller, L. *J. Magn. Reson.* **1992**, *98*, 674−679.
(49) McCoy, M. A.; Mueller, L. *J. Am. Chem. Soc.* **1992**, *114*, 2108−2112.
(50) Boyd, J.; Soffe, N. *J. Magn. Reson.* **1989**, *85*, 406−413.
(51) Patt, S. L. *J. Magn. Reson.* **1992**, *96*, 94−102.

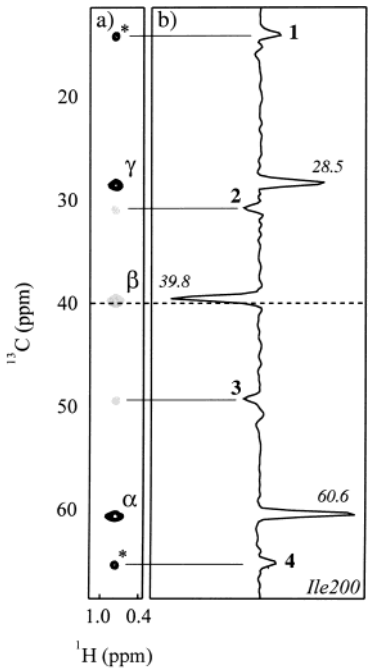

Figure 6. Triple-spin peaks in the HMCM[CG]CBCA spectrum recorded on methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] MSG. (a) A $^{13}C_{aliph}$−$^{1}H_m$ strip at the $^{13}C_m$ (δ1) chemical shift of Ile 200 of MSG. Negative peaks are shown in gray. (b) Trace along the ^{13}C dimension. The chemical shifts of α, β, and γ (γ1) nuclei are indicated on top of the corresponding peaks. The position of the carrier (at 40 ppm) is shown with a dashed horizontal line. The positions of triple-spin peaks labeled 1−4 relative to the position of the ^{13}C carrier are calculated using the following linear combinations: $(\Omega_\alpha + \Omega_\beta - \Omega_{\gamma 1})$ for peak 1; $-(\Omega_\alpha + \Omega_\beta + \Omega_{\gamma 1})$ for peak 2; $(\Omega_\alpha - \Omega_\beta + \Omega_{\gamma 1})$ for peak 3; and $-(\Omega_\alpha - \Omega_\beta - \Omega_{\gamma 1})$ for peak 4, where Ω_α, Ω_β, and $\Omega_{\gamma 1}$ are offsets of α, β, and γ1 resonances from the carbon carrier, respectively. Triple-spin cross-peaks marked with asterisks are folded in the ^{13}C dimension. Note that Ω_α, Ω_β, and $\Omega_{\gamma 1}$ are equal to 20.6, −0.2, and −11.5 ppm, respectively.

relative to sequences used to record data sets on side chains that are "linearized".) Despite the fact that sensitivity gains are not expected in the HMCM[CG]CBCA experiment, there are still considerable benefits associated with the new labeling scheme. These benefits arise from the simplicity of the resulting spectra. For example, in the case of Val residues, correlations at F_1 frequencies of $\Omega_{C\alpha}$ and $\Omega_{C\beta}$ are obtained in HMCM[CG]-CBCA data sets, while for Ile and Leu, in addition to correlations at $\Omega_{C\alpha}$, $\Omega_{C\beta}$, and $\Omega_{C\gamma}$, there are four correlations that result from the triple spin term, $8I_zC_x^\gamma C_y^\beta C_x^\alpha$ (see above). This is illustrated with experimental data in Figure 6 for Ile200. The four "multispin resonances" are 4 times lower in intensity than the cross-peaks of interest (at $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C^{\gamma 1}$ chemical shifts, Figure 6b), can be easily identified, and do not interfere with analysis. In contrast, in the case where both methyls in Leu and Val are ^{13}C-labeled (ϕ3 and ϕ6 are incremented by $\pi/2$ relative to the case where only a single methyl is labeled), it can be shown that for magnetization originating on methyl 1 (m1), multiple spin coherences of the type $8I_zC_y^{m2}C_x^\beta C_y^\alpha$ are generated for Val and of the type $8I_zC_y^{m2}C_x^\gamma C_y^\beta$, $8I_zC_y^{m2}C_y^\beta C_x^\alpha$, $16I_zC_y^{m2}C_x^\gamma C_z^\beta C_x^\alpha$, and $16I_zC_z^{m2}C_x^\gamma C_y^\beta C_x^\alpha$ for Leu. Moreover, a pair of multispin terms, $8I_zC_x^\delta C_y^{\gamma 1}C_y^\beta$ and $16I_zC_x^\delta C_y^{\gamma 1}C_z^\beta C_x^\alpha$, is generated for Ile. Therefore, 4 triple spin peaks would be observed for each Val methyl, 16 peaks for each Leu methyl, and 8 for each Ile δ1 methyl, significantly increasing the complexity of NMR spectra. Finally, straightforward calculations which neglect pulse imperfections and off-resonance effects show that single $^{13}CH_3$ labeling of Val and Leu side chains increases the sensitivities of both HMCM(CGCBCA)CO (for Leu) and HMCM(CBCA)CO (Val) data sets by at least a factor of 1.5 relative to spectra recorded on samples where both methyls of Val and Leu are ^{13}C-labeled. The increase derives from the transfer functions that are involved during the $2\tau_d$ delay in the sequence of Figure 4b. In the case of single $^{13}CH_3$ labeling, the transfer function is given by $\{\sin(\pi J_{CC}2\tau_d)\ \sin(\pi J_{C\alpha CO}2\tau_d)\}^2$, while if both methyls are of the $^{13}CH_3$ variety a transfer function of the form $2\{\cos(\pi J_{CC}2\tau_d)\ \sin(\pi J_{CC}2\tau_d)\ \sin(\pi J_{C\alpha CO}2\tau_d)\}^2$ is appropriate (neglecting relaxation). Note that the factor of 2 in the second expression above takes into account that magnetization from both methyls is transferred to the carbonyl. In addition, we have not been able to develop a pulse scheme which is simultaneously optimal for both Leu and Ile residues in the case of complete ^{13}C labeling; any sequence which creates pure transverse ^{13}CO magnetization for one of the two residues generates multispin coherences for the other. Thus, separate data sets for Leu and Ile would be necessary in the case of uniformly ^{13}C-labeled samples, further compromising the effective sensitivity. It is quite clear that the $^{13}CH_3$,$^{12}CD_3$ labeling strategy is much preferred.

The relative sensitivities of each of the experiments, on a residue-type basis, are indicated in Figure 7. To account for the differences in net acquisition times between the experiments, each is divided by the factor $\sqrt{f}$, where f is the net acquisition time of the experiment in question divided by the net acquisition time for the Val-HMCM(CBCA)CO. The average normalized signal-to-noise (S/N) is indicated in each panel, along with the percentage of the total residues that were quantified (in brackets). Residues with low S/N values (less than 4) that could not be accurately quantified or with correlations that were overlapped were not included, so that a very conservative estimate of the fraction of residues that could be quantified is obtained. In the case of the HMCM[CG]CBCA data set, only those residues with well-resolved C^β and C^α correlations were counted, with the S/N defined as the average over both correlations. It is clear that the methyl out-and-back spectra are significantly more sensitive than the corresponding HN-detected correlation maps. The HMCM[CG]CBCA experiment (Figure 7a−c) is 5−8 times more sensitive than the corresponding (HM)CM([CG]CBCA)-NH experiment (Figure 7g−i), depending on the residue, while the sensitivity of the HMCM([CG]CBCA)CO (Figure 7d−f) is 4−5 times higher than that of (HM)CM([CG]CBCA)NH. Despite the higher sensitivity of the out-and-back experiments relative to the HN-detected schemes, the fraction of residues that could be unequivocally identified in each of the data sets is similar for each type of residue − the higher S/N in the HMCM[CG]CBCA, for example, is offset to some extent by the increased dispersion of the correlations in the HN-observed experiments. It is worth emphasizing that three methyl out-and-back data sets were recorded in under 6 days of measuring time: a single HMCM[CG]CBCA correlation map, as well as a pair of data sets linking methyl and carbonyl spins (Val-HMCM(CBCA)CO and Ile,Leu-HMCM(CGCBCA)CO). In the case of the HN-detected experiments, a pair of data sets were obtained correlating methyl ^{13}C and backbone amide shifts (Ile,-Leu-(HM)CM(CGCBCA)NH and Val-(HM)CM(CBCA)NH) and a second set correlating methyl proton shifts (Ile,Leu-HM-

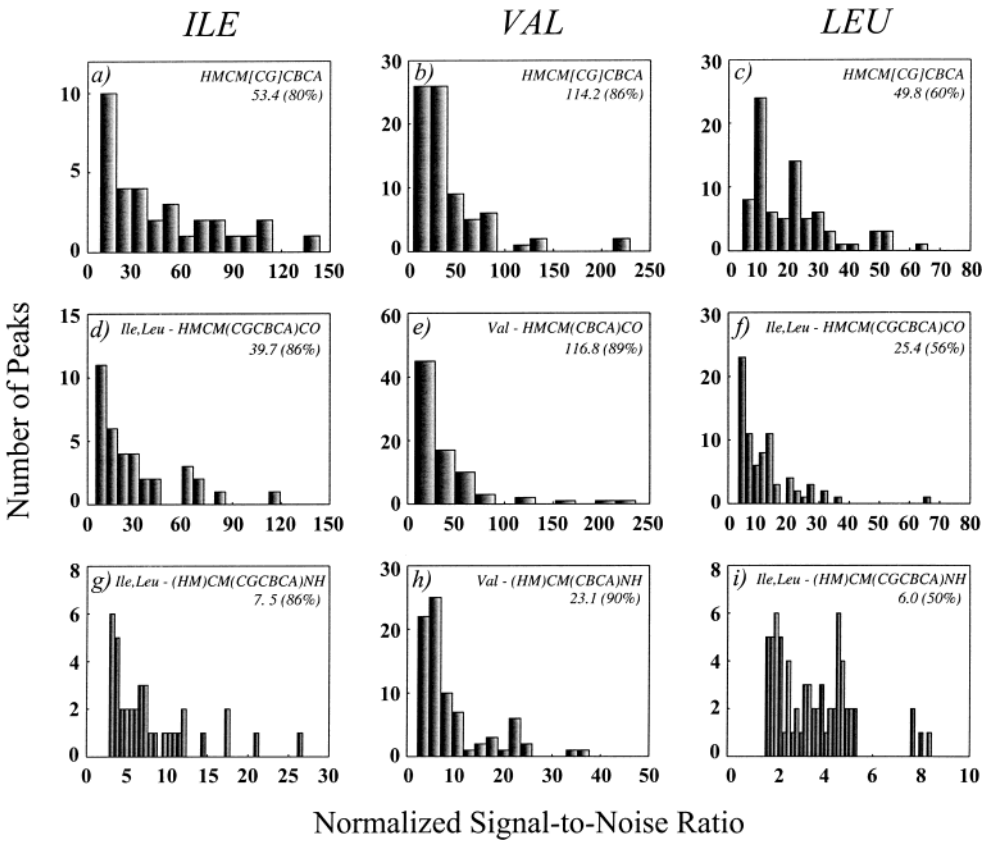

Figure 7. Histograms of normalized signal-to-noise ratios (S/N) for spectra used to assign the methyl groups of Ile (δ1 only), Leu, and Val residues in MSG. The signal-to-noise ratios of each data set were scaled by dividing the S/N by $\sqrt{f}$, where f = net acquisition time of the experiment in question/net acquisition time for the Val-HMCM(CBCA)CO data set. Several residues with extremely high S/N ratios were excluded from each histogram for better visualization. Average signal-to-noise ratios of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ correlations observed in the HMCM[CG]CBCA experiment are shown for Ile (a), Val (b), and Leu (c). S/N ratios of methyl-carbonyl correlations are indicated for Ile from the Ile,Leu-HMCM(CGCBCA)CO data set (d), Val from Val-HMCM(CBCA)-CO (e), and Leu from Ile,Leu-HMCM(CGCBCA)CO (f). S/N ratios of Ile correlations in Ile,Leu-(HM)CM(CGCBCA)NH (g), Val correlations in Val-(HM)CM(CBCA)NH (h), and Leu correlations in Ile,Leu-(HM)CM(CGCBCA)NH (i) were calculated for intraresidue peaks only. Also shown below experiment names in each panel are weighted averages of S/N ratios and, in parentheses, the percent of the total number of expected cross-peaks whose S/N ratios could be reliably quantified (S/N > 4 and only well resolved correlations; note in the case of the HMCM[CG]CBCA spectrum only those residues for which both C^{α} and C^{β} correlations could be quantified were included in the analysis).

(CMCGCBCA)NH and Val-HM(CMCBCA)NH) for a total of 17 days of machine time (conventional room-temperature probe).

Using only the methyl-detected experiments, we were able to assign 42 Ile δ1 methyls (out of 44, 95%), 86 methyls of Val (out of 92, 93%), and 90 methyls of Leu (out of 140, 64%) in MSG. Near complete assignment of Leu residues was precluded by the fact that only ~60% of Leu resonances (i) had S/N values greater than 4 in the Ile,Leu-HMCM(CGCBCA)-CO experiment or (ii) were sufficiently well resolved so that unique correlations could be identified unequivocally in this data set, and (iii) by the significant degeneracy in Leu $^{13}C^{\alpha}/^{13}C^{\beta}$ chemical shift pairs. Strong scalar coupling between $^{13}C^{\delta}$ and $^{13}C^{\gamma}$ spins in some of the Leu residues may have prevented their observation in many of the spectra recorded. However, the combination of HN-based and methyl-detected data sets led to near complete assignments of MSG methyl groups: 42 Ile δ1 methyls (95%), 90 Val methyls (98%), and 120 Leu methyls (86%). (Note that tentative assignments were not included in the percentages listed above.) Although both classes of experiments were necessary in the case of MSG, for smaller proteins it is very likely that the out-and-back data sets alone will suffice and total recording times on the order of 2−3 days should be more than sufficient. As a final check of our assignments, we have recorded a 3D [^{13}C-F_1, ^{13}C-F_2]-edited NOESY spectrum of a {L($^{13}CH_3$,$^{13}CH_3$),V($^{13}CH_3$,$^{13}CH_3$)} U-[^{15}N,^{2}H] MSG sample (mixing time of 30 ms) to obtain intraresidue methyl NOE correlations. In addition to confirming our assignments, we have used this data set to assign Leu methyls in cases where assignment of only 1 of the 2 methyls was available via the through-bond experiments described above. In this manner, an additional 7 Leu methyls were assigned, bringing the total number to 127 (91%).

In summary, near complete assignments of Ile (δ1 only), Leu, and Val methyl groups in MSG have been made using new labeling strategies in concert with new NMR experiments. These include COSY-based methyl−HN magnetization transfer schemes along with sequences which correlate methyl ^{13}C and ^{1}H spins with aliphatic carbon chemical shifts or backbone carbonyl spins in an out-and-back manner. The experiments significantly increase the size of proteins that can be studied by NMR and open up the possibility of using side-chain methyl groups to probe molecular structure and dynamics in high molecular weight systems.

Acknowledgment. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) and funding

from AstraZeneca and Syngenta. We thank Dr. C. T. Tan (Isotec, OH) for the custom synthesis of α-ketoisovalerate used as a biosynthetic precursor in this work, Prof. S. J. Remington (University of Oregon) for the gift of the plasmid encoding MSG, Prof. Cheryl Arrowsmith (Ontario Cancer Institute) for kindly allowing us to carry out the protein expression and purification in her laboratory, and Dr. Oscar Millet (University of Toronto) for many stimulating discussions. V.T. is a recipient of a Human Frontiers Science Program Postdoctoral Fellowship. L.E.K. holds a Canada Research Chair in Biochemistry.

Supporting Information Available: One table listing the ^{1}H and ^{13}C chemical shifts of methyl groups of Ile ($\delta 1$ only), Leu, and Val of malate synthase G at pH 7.1, 37 °C (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA030345S