## Incorporation of $^{1}H/^{13}C/^{15}N$-{Ile, Leu, Val} into a Perdeuterated, $^{15}N$-Labeled Protein: Potential in Structure Determination of Large Proteins by NMR

William J. Metzler, Michael Wittekind, Valentina Goldfarb, Luciano Mueller, and Bennett T. Farmer II*

*Pharmaceutical Research Institute*
*Bristol-Myers Squibb, P.O. Box 4000*
*Princeton, New Jersey 08543-4000*

*Received February 15, 1996*

Proton−proton NOEs are the primary data for protein structure determination by NMR; it is therefore imperative that protons involved in these NOEs be accurately assigned. To this end, uniform $^{13}C/^{15}N$ isotopic labeling has become the mainstay for solution-state structural studies of both small and large proteins.[1] For small proteins, such labeling often leads both to a faster[2] and to a more precise determination of structure.[3] For proteins in the range of 100−230 residues, uniform $^{13}C/^{15}N$ labeling is essential to NMR structural studies. With further increases in protein size, spectral dispersion yields to sensitivity as the limiting experimental factor. Substantial gains in sensitivity have been recently reported for many triple-resonance scalar correlation experiments on proteins deuterated at non-exchangeable proton sites.[4,5] Although perdeuteration also increases sensitivity in amide-detected NOE experiments on large proteins,[6−8] in solution-state structure determination, perdeuteration severely limits both the maximum number and the nature of possible proton−proton NOEs.

In NOE experiments, random fractional deuteration (∼60%) of otherwise unlabeled proteins increases both general proton resolution and sensitivity.[6,9] Similarly, random fractional deuteration limits spin diffusion by reducing the number of available magnetization-transfer pathways.[6,9] This reduction permits a more efficacious use of longer NOE mixing times. With uniformly $^{13}C$-labeled proteins, however, random fractional deuteration will lead to a distribution in the total deuterium isotope effect[4,10] on each $^{13}C$ chemical shift. This distribution can severely broaden resonances along the $^{13}C$ dimension. In addition to the random fractional deuteration of unlabeled proteins, techniques have also been described for type-specific protonation within an otherwise deuterated protein.[11−13] In contrast to random fractional deuteration, type-specific protonation preferentially increases local rather than global proton density; type-specific protonation therefore retains more of the benefits of perdeuteration.

To circumvent the problems inherent in both the perdeuteration and the random fractional deuteration of uniformly $^{13}C/^{15}N$-labeled proteins, the original technique for type-specific protonation has been modified to provide the following: $^{13}C/^{15}N$ labeling of the protonated specific residue types; uniform $^{15}N$ labeling; and more complete deuteration of non-exchangeable sites on all other residue types. Furthermore, the modified expression medium does not require all amino-acid residue types; only the $^{1}H/^{13}C/^{15}N$-labeled specific residue types need be supplied. In this communication, we describe the methodology behind this novel approach to type-specific protonation; evaluate in a small protein the specificity and extent of protonation for one set of residue types ({Ile, Leu, Val} ≡ ILV); and demonstrate the potential benefits of such type-specific ILV protonation to the NMR structure determination of a large protein.

Previous work has shown that a coarse global fold can be determined for a perdeuterated large protein (HCA II) using solely $^{1}H_N-^{1}H_N$ NOEs with a 6 Å cutoff distance.[7] With this in mind, ILV residues have been chosen as the $^{1}H/^{13}C$-labeled specific residues for the following three reasons: (1) favorable methyl-group $T_2$ relaxation even with $^{13}C$ labeling;[14,15] (2) strong NOE interactions in the protein core; and (3) minimal scrambling in *E. coli* with other amino acids. ILV residues are predominantly found in hydrophobic pockets and cores, regions not well probed structurally by $^{1}H_N-^{1}H_N$ NOEs. For this reason, structural information derived from $^{1}H_N-^{1}H_N$ NOEs should be uniquely complemented by that derived from NOEs between ILV side chain protons.

The well-characterized murine Grb2 N-SH3/SOSE protein−peptide complex[16] has been used to demonstrate the specificity of this labeling scheme. $^{1}H/^{13}C/^{15}N$-ILV residues have been incorporated into the perdeuterated and uniformly $^{15}N$-labeled N-terminal SH3 domain of mGrb2[17] ($^{ILV}$N-SH3/SOSE). The labeling methodology consists of expressing the protein in minimal medium containing 95% $D_2O$, [$^{2}H$]-glucose, $^{15}NH_4SO_4$, and $^{1}H/^{13}C/^{15}N$-{Ile, Leu, Val} amino acids. The exact composition of the expression medium is provided in the supporting information. Glucose, instead of acetate,[18] has been used as the sole carbon source to minimize potential scrambling of the labeled ILV residues with other residue types. With [$^{1}H$]-glucose, aromatic rings retain significant levels of protonation;[19] [$^{2}H$]-glucose is therefore required.

The 2D constant-time $^{1}H-^{13}C$ HSQC on $^{ILV}$N-SH3/SOSE (Figure 1) demonstrates the high selectivity in $^{1}H/^{13}C$ incorporation: no non-ILV residues are labeled. Furthermore, the spectrum contains no discernible deuterium isotope-shifted peaks associated with either methylene or methyl correlations. This indicates that most, if not all, of the ILV side chain groups are fully protonated. In line with previous observations,[11] only 15−20% fractional protonation of ILV $C_\alpha$ carbons is observed in a 1D $^{13}C$-filtered proton spectrum (Figure 3 in supporting

* To whom correspondence should be addressed.
(1) Wittekind, M.; Mapelli, C.; Farmer, B. T., II; Suen, K. L.; Goldfarb, V.; Tsao, J.; Lavoie, T.; Barbacid, M.; Meyers, C. A.; Mueller, L. *Biochemistry* **1994**, *33*, 13531−13539. Metzler, W. J.; Farmer, B. T., II; Constantine, K. L.; Friedrichs, M. S.; Lavoie, T.; Mueller, L. *Protein Sci.* **1995**, *4*, 450−459. Tanaka, T.; Ames, J. B.; Harvey, T. S.; Stryer, L.; Ikura, M. *Nature* **1995**, *376*, 444−447.
(2) Vis, H.; Boelens, R.; Mariani, M.; Stroop, R.; Vorgias, C. E.; Wilson, K. S.; Kaptein, R. *Biochemistry* **1994**, *33*, 14858−14870.
(3) Clore, G. M.; Gronenborn, A. M. *Science* **1991**, *252 (5011)*, 1390−9.
(4) Grzesiek, S.; Anglister, J.; Ren, H.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 4369−4370.
(5) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655−11666. Venters, R. A.; Huang, C.-C.; Farmer, B. T., II; Trolard, R.; Spicer, L. D.; Fierke, C. A. *J. Biomol. NMR* **1995**, *5*, 339−344.
(6) LeMaster, D. M.; Richards, F. M. *Biochemistry* **1988**, *27*, 142−50. Torchia, D. A.; Sparks, S. W.; Bax, A. *J. Am. Chem. Soc.* **1988**, *110*, 2320−2321.
(7) Venters, R. A.; Metzler, W. J.; Spicer, L. D.; Mueller, L.; Farmer, B. T., II *J. Am. Chem. Soc.* **1995**, *117*, 9592−9593.
(8) Grzesiek, S.; Wingfield, P.; Stahl, S.; Kaufman, J. D.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 9594−9595.
(9) Torchia, D. A.; Sparks, S. W.; Bax, A. *J. Am. Chem. Soc.* **1988**, *110*, 2320−2321.
(10) Hansen, P. E. *Prog. NMR Spectrosc.* **1988**, *20*, 207−255.
(11) Crepsi, H. L.; Rosenberg, R. M.; Katz, J. J. *Science* **1968**, *161*, 795−796.
(12) Markley, J. L.; Potter, I.; Jardetzky, O. *Science* **1968**, *161*, 1249−1251.
(13) Arrowsmith, C. H.; Pachter, R.; Altman, R. B.; Iyer, S. B.; Jardetzky, O. *Biochemistry* **1990**, *29*, 6332−6341.
(14) Kay, L. E.; Bull, T. E.; Nicholson, L. K.; Griesinger, C.; Schwarbe, H.; Bax, A.; Torchia, D. A. *J. Magn. Reson.* **1992**, *100*, 538−558.
(15) Bax, A.; Max, D.; Zax, D. *J. Am. Chem. Soc.* **1992**, *114*, 6923−6925.
(16) Wittekind, M.; Mapelli, C.; Goldfarb, V.; Lee, V.; Meyers, C. A.; Mueller, L. Manuscript in preparation.
(17) Lowenstein, E. J.; Daly, R. J.; Batzer, A. G.; Li, W.; Margolis, B.; Lammers, R.; Ullrich, A.; Skolnick, E. Y.; Schlessinger, J. *Cell* **1992**, *70*, 431−442.
(18) Venters, R. A.; Calderone, T. L.; Spicer, L. D.; Fierke, C. A. *Biochemistry* **1991**, *30*, 4491−4494.
(19) LeMaster, D. M. *Prog. NMR Spectrosc.* **1994**, *25*, 371−419.

S0002-7863(96)00487-8 CCC: $12.00 © 1996 American Chemical Society

**Figure 1.** Expansion of the $\beta/\gamma/\delta$ region from the 2D gradient-enhanced, constant-time $^{1}$H−$^{13}$C HSQC spectrum on $^{\mathrm{ILV}}$N-SH3/SOSE. Only {Ile, Leu, Val} $\beta/\gamma/\delta$ correlations are observed in the $^{13}$C region upfield of 47 ppm; all such correlations are accounted for. Peaks labeled with an asterisk are unassigned but do not correspond to the known chemical shifts[16] of any $^{1}$H/$^{13}$C resonance pair in this protein. These extra peaks may therefore reflect some level of conformational heterogeneity within the protein−protein complex. Contours are plotted at a spacing of 1.6. The details pertaining to data collection are provided in the supporting information.

**Figure 2.** The superposition of the HCA II crystal structure (thin line) and the energy-minimized average HCA II structure (thick line) is presented in stereoview. The details of the simulated structure calculation for HCA II are described in the text.

information). The high level of deuteration at the ILV $C_{\alpha}$ carbons should increase sensitivity for all amide-detected experiments, including the HC(CC)(CO)NH[20] and the $^{13}$C/$^{15}$N-separated NOESY.[21] The mechanism by which the α proton is replaced by a deuteron has been postulated to involve exchange with solvent deuterons (or protons).[11] Such an exchange mechanism may allow expression of uniformly $^{15}$N-labeled protein in which only the amide and α sites are substantially protonated.

The $^{1}$H/$^{13}$C resonances in large, type-specifically ILV protonated proteins ($^{\mathrm{ILV}}$proteins) can be assigned based on correlations observed in a 4D $^{13}$C/$^{15}$N-separated NOESY (*unpublished results*). In this assignment process, the efficacy of the NOESY is enhanced by focused intraresidue spin-diffusion from the side chain protons out to their intraresidue and sequential amide protons. Although this assignment process is only minimally complicated by spectral overlap of Leu and Val methyl resonances, such overlap may prevent the unique assignment of NOEs between ILV methyl protons. To address this issue, we have also developed a protocol to express $^{\mathrm{IV}}$protein, in which Leu residues are both $^{13}$C- and $^{2}$H-labeled. A constant-time $^{1}$H−$^{13}$C HSQC spectrum on $^{\mathrm{IV}}$N-SH3/SOSE (Figure 4 in supporting information) confirms the applicability of this protocol in mitigating spectral overlap of Leu and Val methyl resonances.

(20) Montelione, G. T.; Lyons, B. A.; Emerson, S. D.; Tashiro, M. *J. Am. Chem. Soc.* **1992**, *114*, 10974−10975.

(21) Kay, L. E.; Clore, G. M.; Bax, A.; Gronenborn, A. M. *Science* **1990**, *249*, 411−414. Muhandiram, D. R.; Xu, G. Y.; Kay, L. E. *J. Biomol. NMR* **1993**, *3*, 463−470.

Analysis of a 3D $^{13}$C-edited/$^{13}$C-separated NOESY ($\tau_{m}$ = 150 ms, 1.4 mM) indicates an NOE cutoff distance of ∼5 Å between $^{13}$C-attached protons in $^{\mathrm{ILV}}$N-SH3/SOSE. Although the NOESY on $^{\mathrm{ILV}}$HCA II will experience decreased sensitivity due both to $^{13}$C labeling and to the concomitant use of 4D heteronuclear-separated spectroscopy, the increased dispersion is absolutely essential in order to accurately assign NOEs involving ILV methyl protons. In a 4D $^{13}$C/$^{13}$C-separated NOESY on $^{\mathrm{ILV}}$HCA II, the absolute level of sensitivity is expected to approach that in the 4D $^{15}$N/$^{15}$N-separated NOESY on $^{2}$H-HCA II.[7] This expectation is based on a comparison of experimental $^{15}$N and $^{1}H_{N}$ $T_2$ relaxation times for $^{2}$H-HCA II (51 and 30 ms, respectively; $\tau_{c}$ = 11.4 ns)[22] with predicted $^{13}C_{methyl}$ and $^{1}H_{methyl}$ $T_2$ relaxation times for $^{\mathrm{ILV}}$HCA II (46 and 48 ms, respectively).[23]

To demonstrate the benefit of this ILV labeling scheme to structure determination of large proteins, simulations were performed using theoretically observable aliphatic−aliphatic and aliphatic−amide NOEs in perdeuterated and uniformly $^{15}$N-labeled $^{\mathrm{ILV}}$HCA II (29 kDa) derived from the crystal structure.[24] Restraints were derived using NOE cutoff distances of 4.5, 4.0, and 3.5 Å for proton methyl−methyl, methyl−(non-methyl), and (non-methyl)−(non-methyl) interactions, respectively. In addition, cutoff distances of 4.5 and 4.0 Å were used for amide−methyl and amide−(non-methyl) interactions. To allow for inaccuracies in NOE intensities resulting from spin diffusion and differential relaxation, the upper bound of each distance restraint was set to 1.5 times the extracted distance. Intraresidue aliphatic−aliphatic NOE restraints and any knowledge of stereoassignments were excluded from the calculations. After combining these restraints (604 in total) with previously described $^{1}H_{N}$−$^{1}H_{N}$ NOE restraints[7] (596 in total), an ensemble of 20 structures was generated by distance geometry and restrained dynamic simulated annealing.[25]

The N-terminus and several loops were ill-defined in these simulations, primarily because these regions contain no ILV and few side chain $^{1}H_{N}$-containing residues, Excluding the N-terminal 16 residues in the HCA II crystal structure, the precision of the NMR ensemble is 3.1 Å, with the energy-minimized, average structure superimposing on the target crystal structure with a backbone RMSD of 2.9 Å (Figure 2). If the disordered loops are excluded in the superposition, the RMSD decreases to 2.1 Å. This represents a dramatic improvement over the ensemble refined solely against $^{1}H_{N}$−$^{1}H_{N}$ restraints, whose RMSD from the crystal structure was 6.4 Å (for residues 21−260).[7] The inclusion of torsion-angle and/or chemical-shift restraints may yield even further improvement.

**Acknowledgment.** B.T.F. acknowledges Dr. Ron Venters of Duke University for useful discussions in the initial phases of this work.

**Supporting Information Available:** Composition of $^{\mathrm{ILV}}$N-SH3 expression medium; acquisition details for the constant-time $^{1}$H−$^{13}$C HSQC spectrum in Figure 1; 1D $^{13}$C-filtered proton spectrum indicating residual $H_{\alpha}$ signal intensity in $^{\mathrm{ILV}}$N-SH3/SOSE; 2D constant-time $^{1}$H−$^{13}$C HSQC spectrum of $^{\mathrm{IV}}$N-SH3/SOSE (5 pages). See any current masthead page for ordering and Internet access instructions.

JA9604875

(22) Venters, R. A.; Farmer, B. T., II; Fierke, C. A.; Spicer, L. D. Submitted for publication.

(23) Model studies[14] were used to predict $T_2(^{13}C_{methyl})$ for $^{\mathrm{ILV}}$HCA II; experimental $^{1}H_{methyl}$ line widths for a 38.5 kDa ILV-labeled flavoprotein ($\tau_{c}$ ∼ 18−19 ns) (*unpublished results*) were used to predict $T_2(^{1}H_{methyl})$ for $^{\mathrm{ILV}}$HCA II.

(24) Liljas, A.; Kannan, K. K.; Bergsten, P.-C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carlbom, U.; Jarup, L.; Lovgren, S.; Petef, M. *Nat. New Biol.* **1972**, *235*, 131−137.

(25) Nilges, M.; Gronenborn, A. M.; Brunger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38.