Cambridge Isotope Laboratories, Inc.
isotope.com

Eurisotop
A CAMBRIDGE ISOTOPE LABORATORIES COMPANY

APPLICATION NOTE 16

Specific Isotopic Labeling of Methyl Groups Has Extended the Molecular Weight Limits for NMR Studies of Protein Structure and Dynamics

John B. Jordan[1] and Richard W. Kriwacki[1,2]
1 St. Jude Children's Research Hospital, Department of Structural Biology
2 Department of Molecular Sciences, University of Tennessee Health Sciences Center

Nuclear magnetic resonance (NMR) spectroscopy has emerged as the preeminent tool in solution studies of protein structure,[1-3] dynamics,[1, 4-10] and intermolecular interactions.[11-14] A key limitation, however, was that only relatively small proteins could be studied. Recent advances in isotope labeling,[15-17] the development of TROSY-based methods,[18] and the advent of cryogenic probes[19] have extended the size limit associated with protein NMR. One widely used isotopic-labeling scheme is perdeuteration, which has been shown to mitigate rapid heteronuclear (*e.g.* ^{13}C and ^{15}N) relaxation.[20] A significant drawback to this labeling scheme, however, is that it drastically reduces the number of ^{1}H-^{1}H NOEs that can be measured and subsequently used for protein structure determination. To allow the measurement of structurally informative NOEs in perdeuterated proteins, Kay and co-workers demonstrated the use of specific α-keto acid precursors to selectively protonate methyl groups of Ile, Leu, and Val residues[15,16,21,22] (referred to here as ILV). The selective incorporation of ^{1}H and ^{13}C in methyl groups of these residues allowed the measurement of structurally informative NOEs with high sensitivity and resolution1. ILV residues are abundant (~21% of all residues)[23,24] and are often found within the hydrophobic cores of globular proteins. These factors facilitate the assignment and measurement of numerous long-range amide-methyl and methyl-methyl NOEs,[16,22,25] which allow global folds of relatively large proteins to be determined. The utility of specific methyl protonation was demonstrated by Tugarinov, *et al*., through near complete assignment of ILV methyl resonances[26] and global fold determination of the 723-residue enzyme, malate synthase G (MSG),[23] and has served to extend the size limitations associated with protein NMR. The following is a brief summary of the ILV methyl group labeling schemes that are available and how these are utilized with current-day protein NMR methods. The reader is referred to the scientific literature for a more complete discussion of these topics.

A Variety of Labeling Schemes Are Possible Using α-Keto Acid Biosynthetic Precursors

Specific methyl protonation with uniform sidechain ^{13}C labeling. Several different ILV labeling patterns can be achieved bio-synthetically using different isotope-labeled α-keto acid precursors in combination with now standard uniform isotope-labeling procedures (Figure 1). For example, precursors are available that, in combination

(continued)

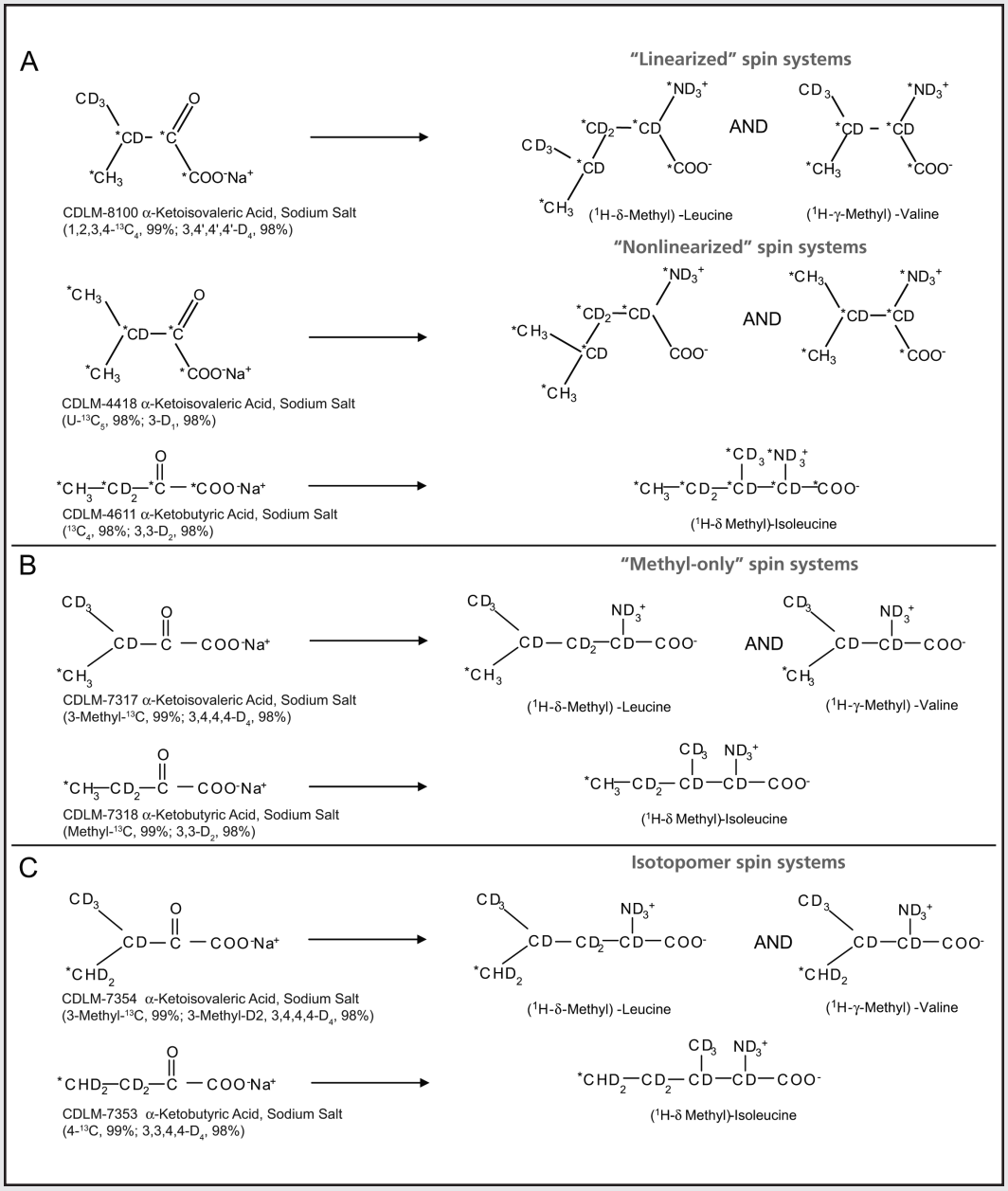

Figure 1. Differently isotope-labeled α-ketoacid metabolic precursors of Ile, Leu and Val sidechains allow a variety of labeling patterns. Precursors used to produce ILV methyl-labeled proteins for (A) NMR experiments used to assign backbone and side chain resonances and for measuring long-range NOEs, (B) for NMR experiments used to measure long-range NOEs and for mapping ligand binding, and (C) for NMR relaxation experiments to study ILV methyl group dynamics.

with the use of ^{15}N ammonium chloride, U-^{13}C/^{2}H-labeled glucose and $^{2}H_2O$, allow ^{13}C and ^{2}H labeling at non-methyl ILV sidechain sites and ^{13}C and ^{1}H labeling at one or both of the methyl groups of Val and Leu residues (in the form of α-ketoisovaleric acid giving $^{13}CH_3$/$^{13}CH_3$ labeling patterns) and at the δ1 methyl group of Ile (using α-ketobutyric acid) (Figure 1A). This labeling scheme is achieved by adding the sodium salts of α-ketobutyric and α-ketoisovaleric acids (60 mg/L and 100 mg/L, respectively) to otherwise standard $^{2}H_2O$-based minimal media about one hour prior to induction of protein expression with IPTG.[15,22,27] These precursors may be used together or separately without scrambling between the Ile and the Val/Leu residues.

This labeling scheme results in a continuous network of ^{13}C-^{13}C bonds between the polypeptide backbone and $^{13}C/^{1}H_3$ methyl groups of ILV residues and is compatible with NMR experiments that transfer magnetization from the methyl groups to other sidechain ^{13}C sites, to sites in the backbone and back,[26] or from methyl groups to sidechain ^{13}C sites.[27,28] This labeling pattern and these experiments are used to assign ILV methyl resonances to specific amino acids in the primary sequence when sequential backbone resonance assignments have previously been established. Protein samples labeled in this manner can also be used to record 3D and 4D ^{13}C and/or ^{15}N-edited NOESY data, from which amide-methyl and methyl-methyl distance restraints can be derived and structures determined.[23,27]

The utility of selective methyl protonation has been demonstrated by Lewis Kay's group at the University of Toronto. Kay's group developed ^{1}H-detected NMR experiments for use in making methyl group assignments in proteins with methyl-protonated ILV spin systems.[26] These 1H methyl-detected "out and back" experiments

include the HMCM(CG)CBCA, Ile, Leu-HMCM(CGCBCA)CO and Val-HMCM(CBCA)CO[26] experiments. These experiments are highly sensitive and were acquired using a room-temperature probe at 800 MHz in 59, 58 and 21 hours,[26] respectively, with a 0.9 mM protein sample. The use of these experiments enabled the ^{1}H methyl (1Hme) and ^{13}C methyl (^{13}Cme) resonances of nearly 78% of the ILV methyl groups in a 82 kDa protein (MSG) to be assigned.[26] Subsequently, the assignment of 3D and 4D NOESY experiments yielded distance restraint information that, when combined with other types of structural restraints, allowed determination of a well-defined global fold for this 723-residue protein.[23]

More recently, a 3D ^{13}C-detected CH_3-TOCSY experiment was developed that allows correlation of ^{1}Hme and ^{13}Cme chemical shifts with $^{13}C_{aliphatic}$ chemical shifts in ILV-protonated spin systems. (Figure 2). [27] Jordan, et al., used this experiment to obtain ILV methyl group chemical shift assignments for a 14 kDa protein domain. These assignments were subsequently used to assign 3D ^{15}N- and ^{13}CNOESY-HSQC spectra recorded using the same ILV-^{13}C/^{1}H-labeled protein sample. The ^{13}C-detected CH3-TOCSY experiment required the use of a cryogenic probe that can directly detect both ^{1}H and ^{13}C NMR signals and was acquired in ~16 hours with a 1 mM protein sample. The use of ^{13}C detection enabled very high resolution data to be acquired in the directly detected carbon dimension, facilitating the sequence-specific assignment ILV ^{13}C/^{1}H methyl resonances.[27] The limited distance information obtained from 3D NOESY spectra for the ILV-^{13}C/^{1}H labeled sample was sufficient to obtain a well-defined global fold of a 14 kDa protein.[27] In addition, this group also demonstrated that the CH_3-TOCSY experiment used with ILV-^{13}C/^{1}H labeling was effective in obtaining similar assignments for a 10 kDa protein in the context of a 42 kDa binary protein complex.

Selective ^{13}C/^{1}H Labeling of ILV Methyl Groups to Create Isolated Methyl Spin Probes

The α-ketobutyrate and α-ketoisovalerate methyl group precursors are also available with ^{13}C and ^{1}H incorporated only in the methyl groups (in either one or both of these in α-ketoisovalerate), with the remainder of the carbons present as ^{12}C with ^{2}H labeling (Figure 1B).

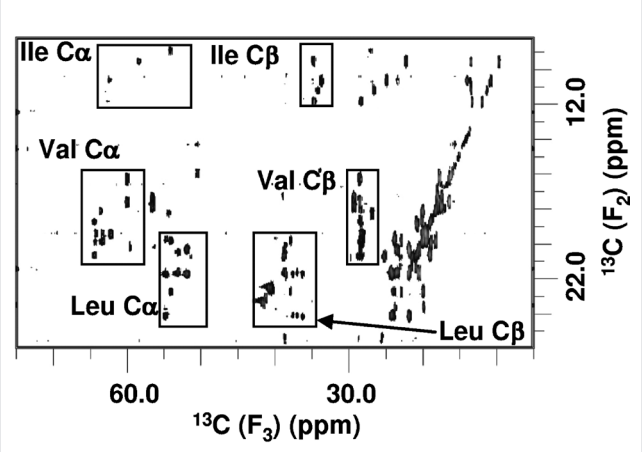

Figure 2. ^{13}C-^{13}C plane of the 3D ^{13}C-detected CH_3-TOCSY experiment acquired on a 1 mM protein sample using "nonlinearized" ILV ^{13}C/^{1}H labeled sidechains.

This approach gives rise to isolated ^{13}C/^{1}H-labeled methyl groups for ILV residues without scalar coupling to other ^{13}C nuclei and, therefore, that exhibit singlet lineshapes (when using ^{1}H decoupling) with favorable relaxation properties. For L and V residues, methyl group relaxation properties are optimal when only one of the two methyl moieties is incorporated as ^{13}C/^{1}H.[26] While this reduces by half the amount of ^{13}C and ^{1}H isotope labels incorporated into each of the two methyl groups, signal loss is compensated for by improved relaxation, resulting in narrower methyl resonances.[26] Further, the elimination of methyl-methyl relaxation between the geminal methyl groups of L and V residues allows NOE interactions between these and other ILV methyl groups to be observed over long distances (*e.g.* up to 8 Å).[23] This labeling strategy is often combined with uniform ^{15}N and ^{2}H labeling of aliphatic sites other than ILV methyl groups labeling and allows long-range amide-methyl and methyl-methyl NOEs to be observed using appropriate 3D and 4D ^{13}C- and/or ^{15}N-edited NOESY NMR experiments.[23] The observation of long-range NOEs is optimized when all aliphatic sites other than ILV methyl groups are ^{12}C and ^{2}H labeled, which is achieved through biosynthetic labeling with U-^{12}C/^{2}H-glucose as the sole carbon source and $^{2}H_2O$ as the solvent.[29] If natural-abundance glucose is used with $^{2}H_2O$ as the solvent, which is more economical, aliphatic sites (other that the specifically ^{1}H/^{13}C-labeled ILV methyl groups) will be ^{2}H-labeled to the extent of ~70%.

In addition to the applications described above (e.g. for optimal detection of long-range NOEs involving ILV methyl groups), specifically ^{1}H/^{13}C-labeled ILV methyl groups have been used as reporters of protein-ligand interactions. This approach is an attractive complement to the use of ^{1}H/^{15}N amide groups as probes of interactions, because 2D ^{1}H-^{13}C chemical shift correlation spectra can be recorded with high sensitivity and because small molecules, as well as macromolecules, often bind within hydrophobic pockets or on hydrophobic surfaces comprised of methyl groups. For proteins of known structure, this allows the binding sites for small molecules to be identified on the basis of binding-induced methyl group chemical shift perturbations. For small ^{1}H/^{13}C ILV methyl-labeled proteins, 2D ^{1}H-^{13}C HSQC spectra can be recorded quickly (~10 minutes) with relatively dilute protein solutions (~50 μM). The use of cryogenic probes reduces these requirements further. Further, Kay and coworkers have shown that, for large proteins or protein assemblies (82 kDa and 305 kDa, respectively), 2D ^{1}H/^{13}C HMQC spectra exhibit superior sensitivity and resolution due to the "methyl TROSY" effect.[29,30] For example, Hamel, et al., described the use of methyl-TROSY NMR to map the residues at the protein-protein interface of the 120 kDa CheA-CheW complex,[31] while Hajduk, et al., demonstrated that the use of selective ILV methyl labeling in high-thoughput, small-molecule screening resulted in ^{1}H-^{13}C HSQC spectra with three-five times higher signal-to-noise ratios than the corresponding ^{1}H-^{15}N HSQC specta.[32]

(continued)

Selective $^{2}H/^{13}C$ Labeling of ILV Methyl Groups for ^{2}H and ^{13}C NMR Relaxation Studies of Protein Dynamics

The topics discussed above are generally focused on using specifically labeled methyl groups to probe protein structure and to monitor protein-protein and protein-ligand interactions. It is well appreciated that methyl groups also serve as probes of side chain dynamics within proteins.[33] In particular, studies of the relaxation behavior of ^{2}H[34,35] and ^{13}C spins,[36,37] in uniformly ^{13}C and fractionally ^{2}H-labeled methyl groups have expanded our knowledge of protein side chain dynamics. In these studies, the relaxation behavior of only one type of methyl isotopomer is monitored at a time. Because deuterons and protons are randomly incorporated into aliphatic sites during bacterial biosynthesis in the presence of $^{1}H_2O$ and $^{2}H_2O$, the concentration of molecules with a particular isotopomer at a particular site (*e.g.* ILV methyl groups) is only a fraction of the total protein concentration. Consequently, the incorporation of ILV methyl group isotopomers with uniform isotopic composition (for example $^{13}CHD_2$ (note: D ≡ ^{2}H here for clarity) for ^{2}H and ^{13}C relaxation studies) would improve spectral S/N and extend the application of these experiments to a broader range of protein systems (Figure 1C). In fact, Kay and coworkers have used this approach in their studies of MSG, an 82 kDa and 723-residue protein,[6,26] as well as the 20 S proteasome with an aggregate mass of 670 kDa38.

Summary and Outlook – New Horizons Await

A discussed at the outset, NMR spectroscopy is a powerful tool in studies of protein structure, protein dynamics and the interaction of proteins with their ligands. The use of specific ILV methyl labeling was essential in extending the size limit for NMR-based global fold determination to ~80 kDa. By being both economically feasible and widely available, this isotope-labeling strategy enables such challenging protein structural studies. As this labeling strategy is more widely adopted by the NMR/structural biology community, we look forward to future applications involving even more challenging biomolecular systems. As discussed above, applications extend beyond structure determination and ligand mapping. For example, in 2005, Kay and coworkers reported the use of Ile methyl labeling of the 300 kDa protease, ClpP, and methyl TROSY methods to detect a conformational exchange process on the millisecond timescale.[39] It was argued that this exchange process is involved in the opening of pores within the midline of the ClpP barrel-like structure for product release. This example illustrates how information from X-ray crystallography, which provided the overall structure of ClpP, and NMR spectroscopy, providing insights into dynamics directly related to key steps of the catalytic cycle, can be combined to reveal how complex molecular machines perform their biological functions. With these advancements in isotopic labeling and the implementation of new NMR methods enabled by them, the tools are in hand to extend our knowledge of the relationships between biomolecular structure, dynamics and function.

References

1. Kay, L.E. **2005**. *J Magn Reson, 173,* (2), 193-207.
2. Wuthrich, K. **1986**. NMR of Proteins and Nucleic Acids. John Wiley & Sons: New York.
3. Clore, G.M.; Gronenborn, A.M. **1991**. *Progress in NMR Spectroscopy, 23,* 43-92.
4. Palmer, A.G. **2001**. *3rd Annu Rev Biophys Biomol Struct, 30,* 129-55.
5. Kay, L.E. **1998**. *Nat Struct Biol,* 5, 513-516.
6. Tugarinov, V.; Kanelis, V.; Kay, L.E. **2006**. *Nature Protocols, 1,* 749-754.
7. Kern, D.; Eisenmesser, E.Z.; Wolf-Watz, M. **2005**. *Methods Enzymol, 394,* 507-24.
8. Wand, A.J. **2001**. *Nat Struct Biol, 8,* (11), 926-31.
9. Dyson, H.J.; Wright, P.E. **2002**. *Adv Protein Chem, 62,* 311-40.
10. Dyson, H.J.; Wright, P.E. **2005**. *Methods Enzymol, 394,* 299-321.
11. Lian, L. .; Barsukov, I.L.; Sutcliffe, M.J.; Sze, K.H.; Roberts, G.C. **1994**. *Methods Enzymol, 239,* 657-700.
12. Pellecchia, M.; Sem, D.S.; Wuthrich, K. **2002**. *Nat Rev Drug Discov, 1,* (3), 211-9.
13. Betz, M.; Saxena, K.; Schwalbe, H. **2006**. *Curr Opin Chem Biol, 10,* (3), 219-25.
14. Bonvin, A.M.; Boelens, R.; Kaptein, R. **2005**. *Curr Opin Chem Biol, 9,* (5), 501-8.
15. Gardner, K.G.; Kay, L.E. Ann. **1998**. *Rev Biophys Biomol Struct , 27,* 357-406.
16. Rosen, M.K.; Gardner, K.H.; Willis, R.C.; Parris, W.E.; Pawson, T.; Kay, L.E. **1996**. *J Mol Biol, 263,* (5), 627-636.
17. Goto, N.K.; Gardner, K.H.; Mueller, G.A.; Willis, R.C.; Kay, L.E. **1999**. *J Biomol NMR, 13,* (4), 369-74.
18. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. **1997**. *Proc Natl Acad Sci, 94,* 12366-12371.
19. Kovacs, H.; Moskau, D.; Spraul, M. **2005**. *Prog NMR Spec, 46,* 131-155.
20. Venters, R. A.; Huang, C. C.; Farmer, B. T., 2nd; Trolard, R.; Spicer, L. D.; Fierke, C.A. **1995**. *J Biomol NMR, 5,* (4), 339-344.
21. Kay, L.E.; Gardner, K. H. **1997**. *Curr Op Str Biol, 7,* 722-731.
22. Goto, N.K.; Kay, L.E. **2000**. *Curr Opin Struct Biol, 10,* (5), 585-592.
23. Tugarinov, V.; Choy, W.Y.; Orekhov, V.Y.; Kay, L.E. **2005**. *Proc Natl Acad Sci USA, 102,* (3), 622-627.
24. Tugarinov, V.; Kay, L.E. **2005**. *Chembiochem, 6,* (9), 1567-1577.
25. Metzler, W.J.; Leiting, B.; Pryor, K.; Mueller, L.; Farmer, B.T., **1996**. *2nd Biochemistry, 35,* (20), 6201-6211.
26. Tugarinov, V.; Kay, L.E. **2003**. *J Am Chem Soc, 125,* (45), 13868-13878.
27. Jordan, J.B.; Kovacs, H.; Wang, Y.; Mobli, M.; Luo, R.; Anklin, C.; Hoch, J.C.; Kriwacki, R.W. **2006**. *J Am Chem Soc, 129,* (28), 9119-9128.
28. Hu, K.; Vogeli, B.; Pervushin, K. **2005**. *J Magn Reson, 174,* (2), 200-208.
29. Tugarinov, V.; Kay, L.E. **2004**. *J Biomol NMR, 28,* (2), 165-72.
30. Tugarinov, V.; Hwang, P.M.; Ollerenshaw, J.E.; Kay, L.E. **2003**. *J Am Chem Soc, 125,* (34), 10420-8.
31. Hamel, D.J.; Dahlquist, F.W. **2005**. *J Am Chem Soc, 127,* (27), 9676-7.
32. Hajduk, P.J.; Augeri, D.J.; Mack, J.; Mendoza, R.; Yang, J.G.; Betz, S.F.; Fesik, S.W. **2000**. *J Am Chem Soc, 122,* 7898-7904.
33. Ollerenshaw, J.E.; Tugarinov, V.; Skrynnikov, N.R.; Kay, L.E. **2005**. *J Biomol NMR, 33,* (1), 25-41.
34. Millet, O.; Muhandiram, D.R.; Skrynnikov, N.R.; Kay, L.E. **2002**. *J Am Chem Soc, 124,* (22), 6439-48.
35. Mulder, F.A.; Hon, B.; Mittermaier, A.; Dahlquist, F.W.; Kay, L.E. **2002**. *J Am Chem Soc, 124,* (7), 1443-51.
36. Teilum, K.; Brath, U.; Lundstrom, P.; Akke, M. **2006**, *J Am Chem Soc, 128,* (8), 2506-7.
37. Lundstrom, P.; Akke, M. **2005**. *Chembiochem, 6,* (9), 1685-92.
38. Sprangers, R.; Kay, L.E. **2007**. *Nature, 445,* (7128), 618-22.
39. Sprangers, R.; Gribun, A.; Hwang, P.M.; Houry, W.A.; Kay, L.E. **2005**. *Proc Natl Acad Sci USA, 102,* (46), 16678-83.