Isotope Labeling of Alanine Methyl Groups on a Deuterated Background for NMR Studies of High-Molecular-Weight Proteins

Chenyun Guo, Raquel Godoy-Ruiz and Vitali Tugarinov
Department of Chemistry and Biochemistry
University of Maryland, College Park, MD 20742 USA

The state-of-the-art isotope-labeling schemes commonly employed for NMR investigations of high-molecular-weight proteins utilize selective incorporation of protons and $^{13}$C isotopes into methyl groups of Ile$^{\delta}$1, Leu$^{\delta}$ and Val$^{\gamma}$ side chains in a highly deuterated environment (commonly referred to as “ILV labeling”) providing a large number of high-quality probes for NMR studies of protein structure and dynamics. Robust ILV labeling methods$^{1,2}$ and strategic location of ILV side-chains in the hydrophobic cores of protein structures$^{3,4}$ have turned ILV labeling into an indispensable tool for NMR studies of large proteins and macromolecular assemblies.$^{5}$

More than a decade ago, Gardner & Kay pioneered a technique for selective incorporation of protonated Ile$^{\delta}$1 methyl positions into protein molecules.$^{6}$ Subsequently, selective methyl labeling/protonation methods have been refined and extended to Leu$^{\delta}$ and Val$^{\gamma}$ methyl sites by the Kay group.$^{7,8}$ The introduction of methyl transverse relaxation optimized spectroscopy$^{9}$ (methyl-TROSY) has stimulated the development of improved labeling schemes.$^{10,11}$ Selective labeling of ILV methyl sites in large proteins on a deuterated background in synergy with methyl-TROSY techniques have had a significant impact on NMR studies of structure and dynamics of large protein assemblies up to ~800 kDa in molecular weight.$^{12-15}$ Nevertheless, the availability of only three ILV probes oftentimes presents a serious limitation for structural and dynamics studies of large proteins. Recent advances in methyl isotope labeling have focused on the extension of the ILV labeling methodology to methyl positions of alanines (Ala$^{\beta}$).

Methyl groups of alanines serve as an attractive extension of the ILV labeling methodology. Ala methyls are located in close proximity to the backbone; their flexibility is therefore reduced compared to other methyl-bearing side chains in proteins. Ala is one of the most frequently encountered residues in protein hydrophobic cores and at molecular interfaces.$^{16}$ Recently, several reports have focused upon extending the ILV methodology with (selective) labeling of Ala$^{\beta}$ methyl positions in large proteins. Isaacson et al. have introduced a labeling strategy for selective incorporation of $^{13}$CH$_3$ groups into alanine residues on a deuterated background.$^{17}$ Specific $^{13}$C labeling and protonation of Ala$^{\beta}$ methyls to a level of 95% with minimal background labeling (<1%) in minimal D$_2$O-based bacterial medium

(continued)
supplemented with large amounts of selectively $^{13}$C-labeled α-deuterated alanine and co-addition of three deuterated compounds: (i) $\alpha$-ketoisovalerate-$D_4$, (ii) succinate-$D_4$ and (iii) L-isoleucine-$D_{10}$ has been reported by Boisbouvier and coworkers. This labeling protocol has been closely followed for production of $[\text{U-}$H; Ala$^\beta$-$^{13}$CH$_3$] labeled malate synthase G$^{18,20}$ (MSG) – an 82-kDa enzyme containing 73 methyl groups. Alanine is the most abundant residue in MSG comprising 10.1% of the total amino-acid content. Figure 1 shows the methyl-TROSY $^1$H-$^1$C correlation map of MSG prepared using the protocol described in more detail in Appendix 1.

To maximize the number of available methyl probes, it is clearly advantageous to combine selective labeling of Ala positions with ILV methyl labeling. The use of selectively $^{13}$CH$_3$-labeled Ala$^\beta$ (2-D; 3-$^{13}$C-L-alanine) in combination with (i) selectively $^{13}$CH$_3$-labeled α-ketoisovalerate for labeling of Ile$^\beta$ positions, (ii) $^{13}$CH$_3$/$^{12}$CD$_2$-labeled α-ketoisovalerate for labeling of Val and Leu$^\delta$ sites, and (iii) deuterated succinate, achieves exactly this purpose (see Appendix 1 for the details about the used compounds). Figure 2 shows the $^1$H-$^1$C methyl-TROSY correlation map recorded on the [Ala$^\beta$-$^{13}$CH$_3$; Ile$^\beta$-$^{13}$CH$_3$; Leu,Val-$^{13}$CH$_3$/$^{12}$CD$_2$]-labeled MSG. Although the overlap of resonances within the Ala$^\beta$ group is very substantial, in MSG Ala$^\beta$ signals do not overlap with either Val$^\gamma$ or Leu$^\delta$ methyl groups.

In full agreement with previous observations, no signs of scrambling of isotope labels from alanine to other amino-acids have been detected in either sample. In fact, a comparison of signal intensities of ILV methyls in the [U-$^1$H; Ala$^\beta$-$^{13}$CH$_3$; Ile$^\beta$-$^{13}$CH$_3$; Leu,Val-$^{13}$CH$_3$/$^{12}$CD$_2$]-labeled sample with those obtained in [U-$^1$H; Ile$^\beta$-$^{13}$CH$_3$; Leu,Val-$^{13}$CH$_3$/$^{12}$CD$_2$]-labeled MSG (without Ala$^\beta$ labeling) indicated that the levels of isotope incorporation into ILV methyl positions are not compromised to any significant extent by additions of large amounts of labeled Ala to the medium. This is the direct consequence of the fact that the biosynthetic pathway leading to incorporation of labels into Ile$^\beta$ positions is “short-circuited” by addition of the suitably labeled α-ketoisovalerate to the medium, while the pathway leading to labeling of Val and Leu$^\delta$ sites is “short-circuited” by addition of the α-keto-isovalerulate precursor. It might have been expected that the addition of α-ketoisovalerate to the medium instead of deuterated isoleucine in order to ensure $^{13}$CH$_3$ labeling at Ile$^\beta$ methyl positions, would result in partial $^{13}$CH$_3$ labeling at Ile$^\beta$ methyl sites (arising from alanine-derived $^{[3-^{13}$CH$_3]$-pyruvate entering into the biosynthetic cycle of Ile$^\beta$ instead of the completely deuterated pyruvate from [U-$^1$H]-glucose). Nevertheless, Ile$^\beta$-$^1$H-$^{13}$C correlations have not been detected in the $^1$H-$^1$C methyl-TROSY spectra recorded on the [Ala$^\beta$-$^{13}$CH$_3$; Ile$^\beta$-$^{13}$CH$_3$; Leu,Val-$^{13}$CH$_3$/$^{12}$CD$_2$]-MSG indicating that during protein production the molecules of pyruvate originate predominantly from [U-$^1$H]-glucose.

Proximity of alanine side chains to the protein backbone and their high degree of order turn Ala$^\beta$ methyl groups into excellent NMR probes for a number of applications: (i) measurements of $^1$H-$^{13}$C residual dipolar couplings (RDCs) in Ala$^\beta$ methyls, (ii) characterization of fast (pico-to-nanosecond) and slow (µs-to-millisecond) dynamics at functionally important sites of enzymes, and (iii) methyl-TROSY NOE spectroscopy that can be performed on [U-$^1$H; Ala$^\beta$-$^{13}$CH$_3$; Ile$^\beta$-$^{13}$CH$_3$; Leu,Val-$^{13}$CH$_3$/$^{12}$CD$_2$]-labeled samples increasing the number of methyl probes for derivation of distance restraints compared to conventional ILV labeling methodology.
Appendix 1. Selective labeling of Alaβ positions with 13CH₃ groups (the \{U-[2H]; Alaβ-[13CH₃]\}-labeled sample) has been achieved following the protocol of Ayala et al.¹⁸ using [U-2H]-D-glucose as the main carbon source in E. coli medium and addition of (i) 800 mg of \{2-D, 3-[13C]-L-alanine; (ii) 2.5 g of succinate-D₄; (iii) 120 mg of α-ketoisovaleric acid, sodium salt (3-methyl-[13C]; 3,4,4,4-D₄) and (iv) 60 mg of α-Ketobutyric acid, sodium salt (3-methyl-[13C]; 3,3-D₂) to 1 L of D₂O-based M9 medium 1 hour prior to induction of protein overexpression with 1 mM IPTG. Selective 13CH₃ labeling of all ILV positions together with Alaβ methyls (\{U-[2H]; Alaβ-[13CH₃]; Ileδ₁-[13CH₃]; Leuδ-[13CH₃]; Leu,Val-[13CH₃/12CD₃]-labeled sample) has been achieved using the same carbon sources as above by addition of (i) 800 mg of \{2-D,3-[13C]-L-alanine; (ii) 2.5 g of succinate-D₄; (iii) 120 mg of α-ketoisovaleric acid, sodium salt (3-methyl-[13C]; 3,4,4,4-D₄) and (iv) 60 mg of α-Ketobutyric acid, sodium salt (3-methyl-[13C]; 3,3-D₂) to 1 L of the medium 1 hour prior to induction. \{Alaβ-[13CH₃]\}- and \{Alaβ-[13CH₃]; Ileδ₁-[13CH₃]; Leu,Val-[13CH₃/12CD₃]\}-labeled samples of MSG were 0.9 mM and 0.75 mM in protein concentration, respectively, in 99% D₂O and contained 25 mM sodium phosphate buffer (pH 7.1, uncorrected), 20 mM MgCl₂, 5 mM DTT and 0.05% NaN₃.

Figure 2. Left: Schematic representation of the structure of MSG with methyl carbons of Alaβ, Ileδ₁, Leuδ and Valγ positions shown with cyan-colored spheres. Center: Methyl 1H-13C HMQC correlation map recorded on a 0.75 mM \{Alaβ-[13CH₃]; Ileδ₁-[13CH₃]; Leu,Val-[13CH₃/12CD₃]\}-labeled MSG (37°C; 600 MHz). The regions of the map approximately corresponding to Ileδ₁, Val and Leu methyl positions are enclosed in dashed rectangles. The region enclosed in a solid rectangle and highlighted corresponds to Alaβ methyl correlations. Right: The Alaβ part of the map “zoomed” from the region highlighted in the center. Cross-peaks arising from Valγ and Leuδ correlations are marked with asterisks. Alaβ assignments are indicated for selected methyls.²²
### CIL Products

#### IN VIVO ILVA METHYL LABELING KIT

**CIL Catalog # CDLM-8806-KIT**

This kit contains four separate vials of the following compounds in the amounts indicated:

<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Catalog No.</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Succinate (D$_5$)</td>
<td>DLM-584</td>
<td>2.5 g</td>
</tr>
<tr>
<td>α-Ketoisovaleric acid, sodium salt (3-methyl-13C; 3,4,4,4-D$_4$)</td>
<td>CDLM-7317</td>
<td>0.12 g</td>
</tr>
<tr>
<td>α-Ketobutyric acid, sodium salt (3-methyl-13C; 3,3-D$_2$)</td>
<td>CDLM-7318</td>
<td>0.06 g</td>
</tr>
<tr>
<td>Alanine (2-D, 3,3-D$_2$)</td>
<td>CDLM-8649</td>
<td>0.8 g</td>
</tr>
</tbody>
</table>

The materials in this kit are to be used in conjunction with 1 L of deuterated minimal media. 1 L of minimal media that is suitable for use with this product typically contains 2 g of glucose-D$_5$, 1 g of ammonium salt (either 15N labeled or unlabeled, depending on preference of the user), and 11.3 g of M9 salts in D$_2$O.

#### IN VIVO ALANINE METHYL LABELING KIT

**CIL Catalog # CDLM-8805-KIT**

This kit contains four separate vials of the following compounds in the amounts indicated:

<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Catalog No.</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Succinate (U-D$_5$)</td>
<td>DLM-584</td>
<td>2.5 g</td>
</tr>
<tr>
<td>α-Keto Isovalerate (U-D$_5$)</td>
<td>DLM-4646</td>
<td>0.2 g</td>
</tr>
<tr>
<td>Isoleucine (U-D$_9$)</td>
<td>DLM-141</td>
<td>0.06 g</td>
</tr>
<tr>
<td>Alanine (2-D, 3,3-D$_2$)</td>
<td>CDLM-8649</td>
<td>0.8 g</td>
</tr>
</tbody>
</table>

The materials in this kit are to be used in conjunction with 1 L of deuterated minimal media. 1 L of minimal media that is suitable for use with this product typically contains 2 g of glucose-D$_5$, 1 g of ammonium salt (either 15N labeled or unlabeled, depending on preference of the user), and 11.3 g of M9 salts in D$_2$O.


**Don't see your desired labeling pattern? Want preweighed labeled components for your media preparation?** We can usually package our reagents to your specifications at no extra charge – please see your local sales representative for a quote.

### References