

# Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy

Vitali Tugarinov<sup>1</sup>, Voula Kanelis<sup>2</sup> & Lewis E Kay<sup>1</sup>

<sup>1</sup>Departments of Medical Genetics and Microbiology, Biochemistry and Chemistry, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, Ontario, M5S 1A8 Canada. <sup>2</sup>Program in Structural Biology and Biochemistry, Hospital for Sick Children, Toronto, Ontario, M5G 1X8 Canada. Correspondence should be addressed to L.E.K. (kay@pound.med.utoronto.ca).

Published online 13 July 2006; doi:10.1038/nprot.2006.101

**The development of isotope labeling methodology has had a significant impact on NMR studies of high-molecular-weight proteins and macromolecular complexes. Here we review some of this methodology that has been developed and used in our laboratory. In particular, experimental protocols are described for the production of highly deuterated, uniformly <sup>15</sup>N- and <sup>13</sup>C-labeled samples of large proteins, with optional incorporation of selective isotope labels into methyl groups of isoleucine, leucine and valine residues. Various types of methyl labeling schemes are assessed, and the utility of different methyl labeling strategies is highlighted for studies ranging from protein structure determination to the investigation of side-chain dynamics. In the case of malate synthase G (MSG), the time frame of the whole preparation, including the protein refolding step, is about 70 h.**

## INTRODUCTION

Recent advances in NMR spectroscopy of high-molecular-weight proteins have been strongly connected to the development of new techniques for optimal isotope labeling of protein samples<sup>1,2</sup>. For example, new experimental methods that increase the lifetimes of NMR signals of large protein molecules from the so-called transverse relaxation optimized spectroscopy (TROSY) effect<sup>3</sup> largely rely upon and benefit from new isotope labeling strategies. Of particular importance in this regard are experimental protocols that have been developed for partial or uniform deuteration of large protein molecules with selective protonation at specific sites (e.g., amide protons and/or certain methyl groups)<sup>4–10</sup>. Deuteration of a high-molecular-weight protein significantly improves the relaxation properties of the remaining subset of protons that is detected in NMR experiments<sup>11</sup>. A very recent example that illustrates nicely the utility of isotope labeling for detailed NMR structural studies of proteins in the 40- to 50-kDa molecular mass regime involves the use of stereo-array isotope labeling (SAIL) technology that achieves stereospecific protonation/deuteration of protein side chains<sup>12</sup>.

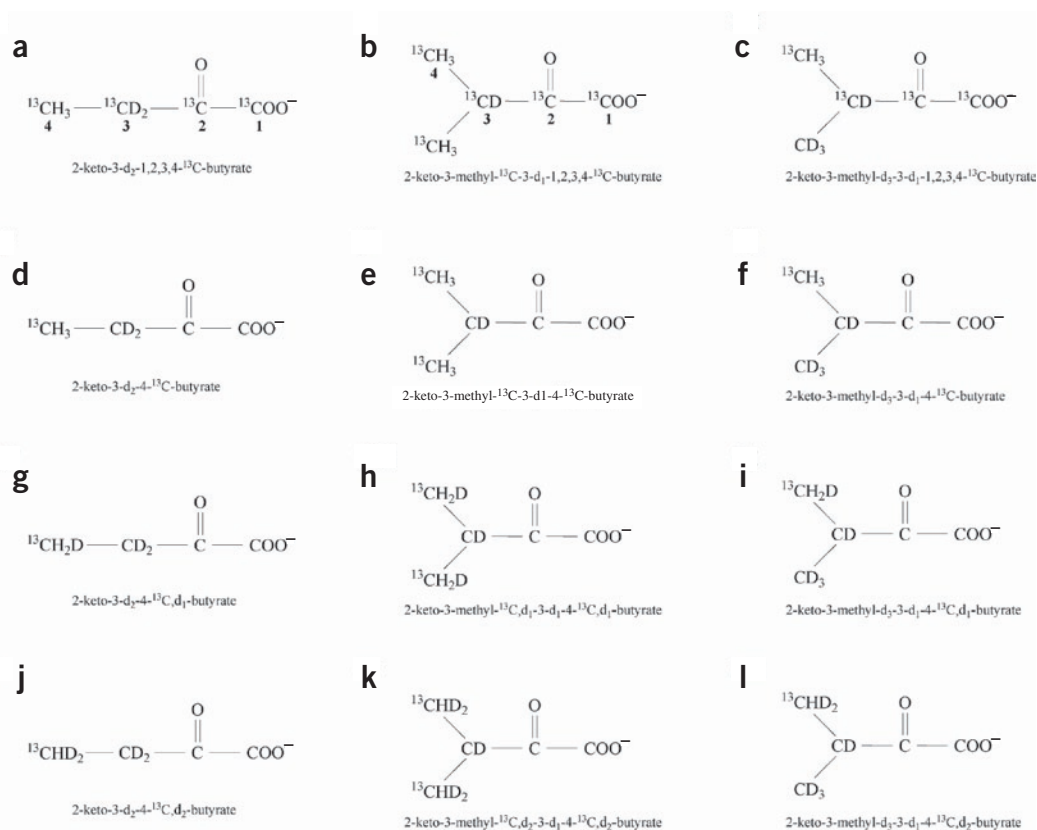
Methyl groups are of particular interest in NMR studies of proteins, because they occur frequently in the hydrophobic cores of these molecules<sup>13</sup> and thus often serve as sensitive reporters of molecular structure and dynamics<sup>14</sup>. Because of their favorable spectroscopic properties, methyl <sup>1</sup>H-<sup>13</sup>C correlation maps can be recorded with high sensitivity and resolution, such that methyl groups are the reporters of choice in many NMR applications that involve large proteins and macromolecular complexes. As has been demonstrated in many applications to date, certain  $\alpha$ -keto acids can serve as robust and efficient biosynthetic precursors for the incorporation of nearly any desired isotope labeling pattern into the side chains of isoleucine, leucine and valine residues in proteins that are overexpressed in *Escherichia coli* using minimal medium<sup>15–17</sup>.

## Choosing an appropriate labeling strategy for NMR studies of protein structure and dynamics

The large variety of different isotope labeling schemes (especially those involving selective methyl labeling) can make it difficult to decide what type of labeling strategy should be used and, indeed, the best approach for one system might be suboptimal for another. Below is a summary of what we consider to be the most useful labeling approaches for NMR studies of protein structure and dynamics, with an emphasis on applications that involve high-molecular-weight systems. Although the labeling strategies described here may be applied to smaller proteins as well, they become particularly important for studies of systems with molecular sizes greater than about 40 kDa.

1. Chemical shift assignments of backbone nuclei (and <sup>13</sup>C <sub>$\beta$</sub>  carbons) in high-molecular-weight protein systems are most readily obtained using [U-<sup>15</sup>N,<sup>2</sup>H,<sup>13</sup>C]-labeled samples or samples labeled as described in the second point below, that have been obtained from D<sub>2</sub>O-based growths. Proteins are dissolved in H<sub>2</sub>O, after <sup>2</sup>H→<sup>1</sup>H exchange; see the description of the refolding procedure used for MSG below.

2. Resonance assignments of isoleucine  $\delta$ 1, leucine and valine methyl groups can be accomplished using linearized <sup>13</sup>C spin-systems in {[U-<sup>15</sup>N,<sup>2</sup>H,<sup>13</sup>C]; Ile $\delta$ 1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>,<sup>12</sup>CD<sub>3</sub>]}-labeled samples dissolved in H<sub>2</sub>O that are produced using the compounds in **Figures 1a** and **c**. (See ref. 7 for NMR experiments designed specifically for this labeling scheme.) The same samples can also be used for the measurement of <sup>3</sup>J<sub>C $\gamma$ CO</sub> scalar couplings<sup>18</sup> as well as to obtain HN-CH<sub>3</sub> and HN-HN distance restraints from nuclear Overhauser effect (NOE) spectroscopy<sup>19</sup>. The <sup>13</sup>CH<sub>3</sub>/<sup>12</sup>CD<sub>3</sub> labeling of isopropyl groups in valine and leucine side chains (that can be achieved using the compounds shown in **Fig. 1c,f**) maximizes the dipolar TROSY effect<sup>20</sup> in <sup>1</sup>H-<sup>13</sup>C multiple-quantum NMR spectra of methyl



**Figure 1** | Isotopically labeled  $\alpha$ -keto acids that are commercially available and that can be used as biosynthetic precursors in *E. coli*-based growth of selectively methyl-labeled proteins. Sodium salts of  $\alpha$ -ketobutyric and  $\alpha$ -ketoisovaleric acids are available in protonated forms at position 3 and can be quantitatively exchanged to 3- $^2\text{H}$  at high pH in  $\text{D}_2\text{O}$  (see text). Each precursor is referred to in the text using its corresponding panel letter.

groups and is, therefore, highly recommended for large proteins, as opposed to a scheme where both methyls are  $^{13}\text{C}$ -labeled and protonated (Fig. 1b,e)<sup>17</sup>. We feel that it is advantageous, however, to use a separate perdeuterated sample (e.g.,  $[\text{U-}^{15}\text{N}, ^2\text{H}]$  or  $[\text{U-}^{15}\text{N}, ^2\text{H}, ^{13}\text{C}]$  labeled as in the first point above) for obtaining amide-amide distance restraints<sup>19</sup>.

3. Another sample with the methyl labeling scheme as above but where all other carbon positions are  $^{12}\text{C}$  ( $[\text{U-}^{15}\text{N}, ^2\text{H}]$ ; Ile $\delta$ 1- $[\text{}^{13}\text{CH}_3]$ ; Leu, Val- $[\text{}^{13}\text{CH}_3, ^{12}\text{CD}_3]$ )-labeled protein, obtained using compounds shown in Fig. 1d,f and dissolved in  $\text{D}_2\text{O}$ ) facilitates optimal measurement of  $^3J_{\text{C}_\gamma\text{N}}$  couplings as well as  $\text{CH}_3\text{-CH}_3$  NOEs<sup>19</sup>. This sample can also be useful for studies of methyl dynamics on a millisecond time scale<sup>21</sup>.

4. NMR studies of picosecond-nanosecond time-scale motions at methyl positions using  $^2\text{H}$  spin relaxation can be carried out on samples that contain either  $^{13}\text{CH}_2\text{D}$  or  $^{13}\text{CHD}_2$  moieties (see below)<sup>22</sup>. Methyl  $^{13}\text{C}$  relaxation rates for extraction of  $^{13}\text{C}$ -derived dynamics parameters can also be measured from samples with methyls labeled as  $^{13}\text{CHD}_2$  (i.e.,  $[\text{U-}^2\text{H}]$ ; Ile $\delta$ 1- $[\text{}^{13}\text{CHD}_2]$ ; Leu, Val- $[\text{}^{13}\text{CHD}_2, ^{13}\text{CHD}_2]$ )-labeled proteins; using compounds shown in Fig. 1j,k)<sup>23</sup>. The  $^2\text{H}$ -based experiments using  $^{13}\text{CH}_2\text{D}$  methyl groups have the advantage in that as many as five relaxation rates can be obtained per site<sup>24</sup>, so that the consistency of the data can be established rigorously. By contrast, only  $^2\text{H}$   $T_2$  (see ref. 22) and  $T_1$ -like (V.T. & L.E.K.,

manuscript in preparation) time constants can be obtained using  $^{13}\text{CHD}_2$  probes, although a comparison with measures of dynamics obtained from  $^{13}\text{C}$  relaxation experiments recorded on the same sample (i.e., using the same isotopomer probe) provides an excellent approach for validation of  $^{13}\text{CHD}_2$ -derived dynamics parameters.

## Experimental design

Here we describe an experimental protocol that is used for the production of highly deuterated, uniformly  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labeled NMR samples of proteins with optional incorporation of selective isotope labels into the methyl groups of isoleucine, leucine and valine residues. Examples are drawn from NMR studies of the structure and dynamics of an enzyme that our laboratory has focused on over the past several years, MSG from *E. coli*<sup>25</sup>. MSG is a monomeric protein that is devoid of disulfide linkages, comprised of 723 residues (82 kDa) and is almost twice as large as other proteins for which near-complete backbone chemical shift assignments have been obtained and for which global fold determination has been reported<sup>6,19</sup>. The rotational correlation time of MSG is sensitive to protein concentration; for a sample concentration of approximately 0.5 mM at 37 °C, a value of 39 ns was determined from  $^{15}\text{N}$  relaxation measurements (in  $\text{H}_2\text{O}$ )<sup>22</sup>. By contrast, a value of 54 ns was obtained for a 0.9 mM sample dissolved in  $\text{D}_2\text{O}$  (ref. 22).

**Protein overexpression in labeled minimal medium (M9/D<sub>2</sub>O) with optional selective labeling of isoleucine, leucine and valine methyl groups.** State-of-the-art NMR methodology for the backbone chemical shift assignment of high-molecular-weight proteins is predicated on uniform labeling with <sup>15</sup>N and <sup>13</sup>C isotopes and (most commonly full) deuteration at all nonexchangeable protein sites. The former can be achieved using [<sup>15</sup>N]ammonium chloride (<sup>15</sup>NH<sub>4</sub>Cl) and D-[<sup>13</sup>C]glucose as the main sources of nitrogen and carbon in minimal growth medium, respectively. The latter is achieved by growing the culture cells in D<sub>2</sub>O and using [<sup>2</sup>H,<sup>13</sup>C]glucose as the main source of carbon. The amount of glucose required for maximal protein expression per gram glucose should be tested using an unlabeled culture.

Complete deuteration at all nonexchangeable protein sites severely limits the number of available probes for both structural and dynamic studies by NMR. The selective protonation of a subset of methyl positions, however, inserts <sup>1</sup>H,<sup>13</sup>C labels into key elements of the molecule that can then be used as sensitive probes of protein structure and dynamics. The development of strategies for selective methyl protonation in a deuterated environment has a long history (see ref. 14 for a review of possible alternative labeling procedures). The most robust and cost-effective approach makes use of  $\alpha$ -keto acids that are added as supplements to minimal protein expression medium and that serve as biosynthetic precursors for the incorporation of methyl groups with the desired isotopic labeling pattern. Selective labeling of isoleucine  $\delta$ 1 methyls is achieved through the addition of  $\alpha$ -ketobutyrate (Fig. 1a,d,g,j)<sup>15</sup>, whereas addition of  $\alpha$ -ketoisovalerate (Fig. 1b,c,e,f,h,i,k,l) leads to the production of selectively labeled valine and leucine residues<sup>16</sup>. The primary advantage of using the  $\alpha$ -ketobutyrate and  $\alpha$ -ketoisovalerate precursors shown in Figure 1 is that many labeling patterns can be incorporated into the side chains of isoleucine, leucine and valine residues, and into methyl moieties themselves, and that scrambling of the label is not observed<sup>16</sup>. Although the number of methyl-containing sites is restricted to isoleucine  $\delta$ 1, valine  $\gamma$  and leucine  $\delta$  side chains, on average these three residues comprise approximately 20% of the total amino acid content in proteins<sup>26</sup>, so that, in general, a good distribution of probes is obtained throughout the molecule.

Sodium salts of  $\alpha$ -ketobutyric and  $\alpha$ -ketoisovaleric acids can be added to D<sub>2</sub>O-based minimal medium in amounts of about 50–70 mg l<sup>-1</sup> and about 100–120 mg l<sup>-1</sup>, respectively, about 1 h before induction of protein overexpression (see below). These keto acids can all be obtained commercially, many with either <sup>1</sup>H or <sup>2</sup>H at position 3. We prefer to purchase compounds that are 3-<sup>1</sup>H, as they are cheaper and because this position can be readily and quantitatively exchanged to 3-<sup>2</sup>H by incubation at high pH in D<sub>2</sub>O before use (at pH 12.5, ~2–3 h for  $\alpha$ -ketoisovalerate, and at pH 10.5, ~12–14 h for  $\alpha$ -ketobutyrate; lower pH is essential for the exchange of  $\alpha$ -ketobutyrate to avoid generating dimers through condensation of two ketoacid molecules)<sup>16</sup>.

***In vitro* refolding of proteins overexpressed in D<sub>2</sub>O-based medium.** Overexpression of proteins in pure D<sub>2</sub>O leads to the production of perdeuterated proteins, with deuterium incorporated also at exchangeable backbone and side-chain positions. Although subsequent protein purification and sample preparation steps are normally conducted in H<sub>2</sub>O, deuterium-to-proton exchange may be incomplete for amides in very stable structural motifs. Because state-of-the-art NMR methodologies for protein backbone assignments rely on the detection of protons at the backbone amide sites, incomplete exchange can lead to severe losses of NMR signals from amides that are deeply buried in protein cores. This necessitates the development of efficient protocols for unfolding and refolding of proteins in H<sub>2</sub>O to fully protonate the amide positions of these buried sites. *In vitro* refolding of large multidomain proteins can be difficult, however, as in most cases the thermodynamics and kinetics of the folding process will not be fully understood. The refolding protocol is often highly protein specific.

If only methyl groups are to be studied, the refolding step described below can be omitted, resulting in higher net yields of protein. Indeed, <sup>1</sup>H-<sup>13</sup>C spectra of methyl groups in high-molecular-weight proteins benefit from as low a density of protons as possible, and it is advisable to record spectra in D<sub>2</sub>O. In this case, incorporation of <sup>15</sup>N through the use of <sup>15</sup>NH<sub>4</sub>Cl in the growth medium is of no benefit. Because this precursor is so inexpensive, however, we routinely include it in protein production, so that if <sup>1</sup>H-<sup>15</sup>N correlation spectra are required at a later date, or HN-methyl NOE spectra, for example, a sample would be available, albeit one that would have to undergo refolding and exchange into H<sub>2</sub>O before use.

**The protocol.** Below we describe a general procedure for isotope labeling and deuteration of proteins for NMR studies with optional selective protonation of isoleucine, leucine and valine methyl groups. In this protocol, we use IPTG for induction of protein overexpression of genes regulated by the *lacI* gene product. This should be changed, if necessary, to the inducing agent that is suitable for the chosen system. We then briefly describe a refolding procedure that proved effective in the case of MSG, with a final yield of refolded monomeric (and fully exchanged) protein of approximately 60%<sup>1,6</sup>.

## Cost

The cost of an NMR sample is strongly dependent on the labeling scheme used and the final yield of protein. In this regard, refolding of protein (see below) can be quite costly because this step often is not particularly efficient. The following is the approximate expense per milligram of protein associated with the production of a number of labeled MSG samples, obtained from 1 l of growth medium: \$60 for [U-<sup>2</sup>H,<sup>15</sup>N,<sup>13</sup>C]MSG in H<sub>2</sub>O (i.e., including a final refolding step); \$85 for [U-<sup>2</sup>H,<sup>15</sup>N,<sup>13</sup>C]; Ile $\delta$ 1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>,<sup>12</sup>CD<sub>3</sub>]MSG in H<sub>2</sub>O; \$30 for [U-<sup>2</sup>H,<sup>15</sup>N,<sup>12</sup>C]; Ile $\delta$ 1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>,<sup>12</sup>CD<sub>3</sub>]MSG in D<sub>2</sub>O.

## MATERIALS

### REAGENTS

- IPTG
- LB rich growth medium
- M9 minimal medium (see REAGENT SETUP) without or with isotopically

labeled compounds (see REAGENT SETUP) as required

- BL21(DE3) cells with plasmid containing the gene of interest as well as the T7 promoter
- Denaturing buffer: 20 mM Tris, 100 mM NaCl, 10 mM  $\beta$ -mercaptoethanol,

6 M guanidinium hydrochloride (GuHCl), pH 7.8

- GuHCl-free buffer: 20 mM Tris, 100 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 5 mM  $\text{MgSO}_4$ , 2 mM benzimidazole, 10% sucrose, pH 7.8

## EQUIPMENT

- Sterile flasks for cell growth (preautoclaved)
- Purification column and protein concentrators
- Nalgene 0.45- $\mu\text{m}$  filtration unit (Nalgene Nunc) for filtering refolded protein

## REAGENT SETUP

**M9 minimal medium** For 1 l, mix together 6.5 g anhydrous  $\text{Na}_2\text{HPO}_4$ , 3.0 g of  $\text{KH}_2\text{PO}_4$ , 0.5 g of NaCl, 4 g of regular (nonisotopically labeled) D-glucose (for unlabeled medium only; depending on the protein, typically 2–3 g is sufficient, but because unlabeled glucose is inexpensive, slightly more is used), 1.0 g of  $\text{NH}_4\text{Cl}$  (for unlabeled medium only), 120 mg of  $\text{MgSO}_4$ , 11 mg of  $\text{CaCl}_2$ , vitamins (biotin and thiamine at a final concentration of 10  $\mu\text{g ml}^{-1}$  for both) and appropriate antibiotics. Filter sterilize. Media should be used within 24 h of adding glucose, vitamins

and antibiotics. Filter sterilized (or autoclaved) salts can be stored at room temperature for up to 6 months.

**Isotopically labeled compounds** The reagents used for this medium depend on the chosen labeling strategy. The carbon source is either 2–3 g of D- $[\text{}^2\text{H};^{13}\text{C}]$ glucose or 2–3 g of D- $[\text{}^2\text{H};^{12}\text{C}]$ glucose. In the case of MSG, 3 g of glucose was used; the amount required for maximal protein expression per gram glucose should be tested before this step.  $^{15}\text{NH}_4\text{Cl}$  (1 g) is used as the nitrogen source, and the reagents are dissolved in a final volume of 1.0 liter of 99%  $\text{D}_2\text{O}$ . If selective methyl labeling is desired, one or two of the compounds listed in **Figure 1** should be included in the growth medium, such as one of the compounds of the  $\alpha$ -ketobutyrate class (**Fig. 1a,d,g,j**) for production of isoleucine and one of the compounds of the  $\alpha$ -ketoisovalerate class (**Fig. 1b,c,e,f,h,i,k,l**) for production of leucine and valine with the labeling pattern of choice. Stock solutions of DCl and NaOD are required for pH (pD) adjustment in cases when compounds of **Figure 1** are acquired in the protonated form at position 3 and exchanged in  $\text{D}_2\text{O}$  at high pD.

## PROCEDURE

### Protein overexpression in labeled minimal medium (M9/D<sub>2</sub>O) with optional selective labeling of isoleucine, leucine and valine methyl groups

- 1| Start with a 5-ml culture of LB rich growth medium from a freshly transformed colony of BL21(DE3) cells and grow in a shaking incubator at 37 °C until the cells reach an  $\text{OD}_{600}$  of about 0.7–0.8.
- 2| Spin the cells down (1,200g at room temperature) and resuspend a small amount of cells in 20 ml sterile M9/H<sub>2</sub>O medium (prepared with unlabeled glucose and unlabeled  $\text{NH}_4\text{Cl}$ ) in a sterile flask to achieve a starting  $\text{OD}_{600}$  of about 0.05–0.1. Incubate the culture until the  $\text{OD}_{600}$  is about 0.6.
- 3| Spin all the cells down and resuspend in 100 ml M9/D<sub>2</sub>O medium (prepared with either D- $[\text{}^2\text{H};^{13}\text{C}]$ glucose or D- $[\text{}^2\text{H}]$ glucose and  $^{15}\text{NH}_4\text{Cl}$ ). The starting  $\text{OD}_{600}$  should be about 0.1. Incubate the culture until the  $\text{OD}_{600}$  is 0.4–0.5.
- 4| Dilute the culture to 200 ml with M9/D<sub>2</sub>O medium and grow until the  $\text{OD}_{600}$  is 0.4–0.5.
- 5| At this point, reagents for methyl labeling of side chains can be added (A) or the culture can be continued without side-chain labeling (B).

#### (A) Carry out methyl labeling of isoleucine, leucine and valine side chains

- (i) Dilute to a volume such that addition of keto acids in step (ii) below will give a volume of 1 l (see ref. 5) and continue growth until  $\text{OD}_{600}$  is about 0.25.
- (ii) Add 70 mg  $\text{l}^{-1}$  (final volume) of one of the  $\alpha$ -ketobutyrate (**Fig. 1a,d,g,j**) for production of isoleucine and/or 120 mg  $\text{l}^{-1}$  (final volume) of one of the  $\alpha$ -ketoisovalerates (**Fig. 1b,c,e,f,h,i,k,l**) for production of leucine and valine.
- (iii) Continue incubating the culture for approximately 1 h. The  $\text{OD}_{600}$  should be 0.3–0.4.

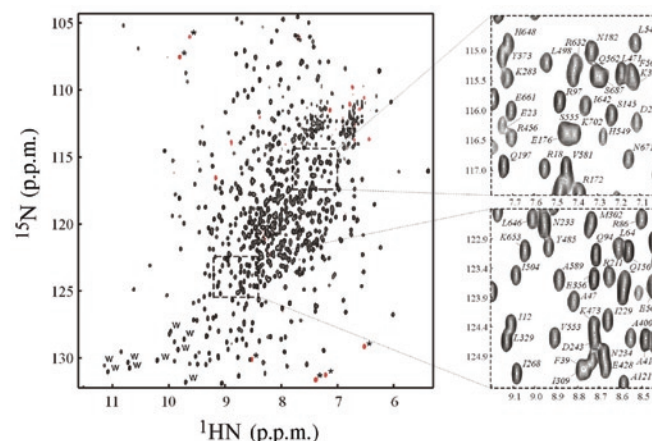
#### (B) Continue growth without side-chain labeling

- (i) Dilute the culture to 1 l with M9/D<sub>2</sub>O medium and continue growth until  $\text{OD}_{600}$  is 0.3–0.4.

- 6| Induce protein overexpression with IPTG. For MSG, a 1 mM final concentration of IPTG was used, and overexpression was carried out at 37 °C.

- 7| Continue the growth for 6–8 h. The final  $\text{OD}_{600}$  for expression of MSG is 1.0–1.2.

▲ **CRITICAL STEP** Excessively prolonged growth after induction should be avoided with selective methyl labeling to prevent generation of methyl groups with undesired isotopomers.





**8|** Harvest the cells by centrifugation at 5,000*g*, 4 °C for 15 min. Typical yields of MSG per 1 l of growth medium are about 50 mg (before the refolding step) and are not affected significantly by methyl labeling.

**9|** Purify the protein with the method that would normally be used.

### ***In vitro* refolding of proteins overexpressed in D<sub>2</sub>O-based medium, an approach used for MSG**

**10|** After initial purification, add denaturing buffer to achieve a protein concentration of 1.5 mg ml<sup>-1</sup>. Incubate 1 h at room temperature to ensure complete denaturation of the protein.

**11|** Refold the protein by rapid dilution into GuHCl-free buffer to a final protein concentration of about 35 μg ml<sup>-1</sup>. Carry out the dilution at 4 °C with vigorous stirring.

**12|** Incubate the solution for 2 h at room temperature.

**13|** Filter the solution from insoluble protein aggregates (if any) and concentrate the protein solution either on purification column(s) or using concentrators.

**14|** Proceed with the preparation of an NMR sample (further gel-filtration and reconcentration may be needed).

### ● **TIMING**

The time frame of the whole preparation, including the refolding step, is approximately 70 h.

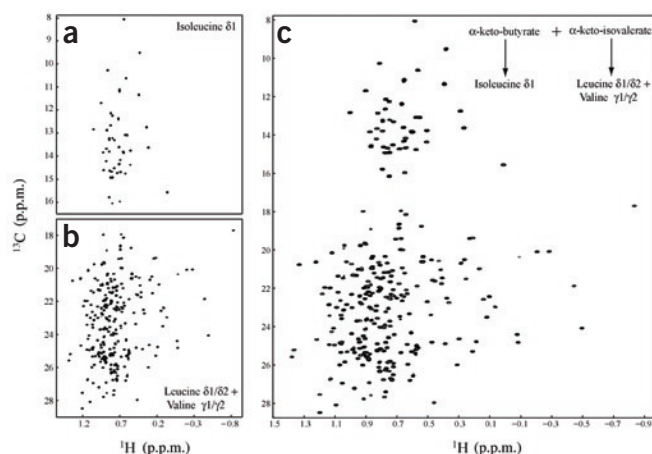
### **ANTICIPATED RESULTS**

Final yields of MSG after the refolding step are typically on the order of 30 mg l<sup>-1</sup> of medium. In the case of MSG, NMR sample concentrations that range from 0.5 to 0.9 mM were used with data recorded at spectrometer fields of 600 and 800 MHz using room temperature probes.

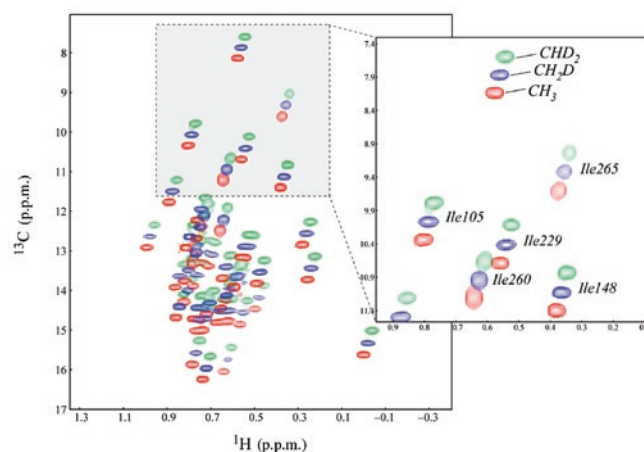
**Figure 2** shows a <sup>1</sup>H-<sup>15</sup>N heteronuclear single-quantum coherence (HSQC)-TROSY spectrum of [U-<sup>2</sup>H, <sup>15</sup>N]MSG refolded in H<sub>2</sub>O after D<sub>2</sub>O-based protein expression using the growth and refolding procedures described above. Comparison with spectra obtained without the refolding step (data not shown) indicates that a substantial (~100–120) number of cross-peaks are absent in data sets recorded from the preparation that omits refolding. These peaks correspond to amides that are buried deeply in the protein core and that exchange slowly with solvent, sometimes on a time scale of months<sup>6</sup>.

The preparation of uniformly <sup>13</sup>C-labeled proteins with selective isoleucine, leucine and valine methyl protonation can be achieved by addition of one of the compounds shown in **Figure 1a–c** to a growth medium containing [U-<sup>13</sup>C, <sup>2</sup>H]glucose, as described above. A {[U-<sup>2</sup>H, <sup>15</sup>N, <sup>13</sup>C]; Ileδ1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>, <sup>12</sup>CD<sub>3</sub>]}-labeled sample was critical for assignment of methyl <sup>1</sup>H and <sup>13</sup>C chemical shifts to specific sites in MSG<sup>7</sup>. Once methyl assignments are obtained, it is often beneficial to work with samples where <sup>13</sup>C labeling is restricted to the methyl groups themselves.

**Figure 3** illustrates several two-dimensional <sup>1</sup>H-<sup>13</sup>C



**Figure 3 |** Two-dimensional <sup>1</sup>H-<sup>13</sup>C HMQC spectra (D<sub>2</sub>O; 800 MHz; 37 °C) acquired with variously labeled MSG samples. (a–c) Spectra were generated using {[U-<sup>2</sup>H, <sup>15</sup>N]; Ileδ1-[<sup>13</sup>CH<sub>3</sub>]}MSG (a), {[U-<sup>2</sup>H, <sup>15</sup>N]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>, <sup>12</sup>CD<sub>3</sub>]}MSG (b) and {[U-<sup>2</sup>H, <sup>15</sup>N]; Ileδ1-[<sup>13</sup>CH<sub>3</sub>]; Leu,Val-[<sup>13</sup>CH<sub>3</sub>, <sup>12</sup>CD<sub>3</sub>]}MSG (c). α-ketoacid precursors were used to produce the samples whose spectra are shown in a (**Fig. 1d**), b (**Fig. 1f**) and c (**Fig. 1d,f**).



**Figure 4 |** An overlay of two-dimensional <sup>1</sup>H-<sup>13</sup>C correlation maps acquired on {[U-<sup>2</sup>H, <sup>15</sup>N]; Ileδ1-[<sup>13</sup>CH<sub>3</sub>]}MSG (peaks shown with red contours), {[U-<sup>2</sup>H, <sup>15</sup>N]; Ileδ1-[<sup>13</sup>CH<sub>2</sub>D]}MSG (blue contours) and {[U-<sup>2</sup>H, <sup>15</sup>N]; Ileδ1-[<sup>13</sup>CHD<sub>2</sub>]}MSG (green contours; D<sub>2</sub>O, 800 MHz, 37 °C). The type of correlation map acquired for samples that contain each isotopomer has been previously described<sup>27</sup>. Peak intensities from each correlation map have been scaled before overlay and do not reflect the relative sensitivities of each two-dimensional experiment. (Reproduced with permission from ref. 27.)

heteronuclear multiple-quantum coherence (HMQC) spectra of selectively methyl-labeled MSG samples. A  $\{[U-^2H, ^{15}N]; Ile\delta 1-[^{13}CH_3]\}$ MSG sample, generated from the compound shown in **Figure 1d** and D- $[U-^2H, ^{12}C]$ glucose as the principal carbon source, was used to record the spectrum of **Figure 3a**, whereas a  $\{[U-^2H, ^{15}N]; Leu, Val-[^{13}CH_3, ^{12}CD_3]\}$ MSG sample, obtained using the compound shown in **Figure 1f**, was used in recording the spectrum of **Figure 3b**. Although separate labeling of isoleucine and leucine plus valine side chains is possible (no ‘scrambling’ of the labels among different residue types is observed) and advantageous for certain NMR applications, the more common and widely used labeling strategy would include both isoleucine and leucine plus valine methyls. Labeling of both leucine and valine methyls occurs when using compounds from **Fig. 1 b,c,e,f,h,i,k,l**. The spectrum of  $\{[U-^2H, ^{15}N]; Ile\delta 1-[^{13}CH_3]; Leu, Val-[^{13}CH_3, ^{12}CD_3]\}$ MSG, obtained by addition of both compounds from **Figure 1d,f** to the growth medium, is shown (**Fig. 3c**).

In addition to the production of ‘isotopically pure’  $^{13}CH_3$ -labeled methyl moieties (i.e., labeling that is restricted to methyl groups), it is also possible to generate proteins with  $^{13}CH_2D$  or  $^{13}CHD_2$  labeling using biosynthetic precursors (**Fig. 1g–l**)<sup>27</sup>. As an example, **Figure 4** shows an overlay of separate two-dimensional  $^1H$ - $^{13}C$  correlation maps acquired using  $\{Ile\delta 1-[^{13}CH_3]\}$ MSG,  $\{Ile\delta 1-[^{13}CH_2D]\}$ MSG and  $\{Ile\delta 1-[^{13}CHD_2]\}$ MSG samples (obtained using the compounds shown in **Fig. 1d,g,j**, respectively). Because of (small) deuterium isotope shifts in both  $^1H$  and  $^{13}C$  dimensions of the spectra, the peaks of each methyl isotopomer have a unique set of chemical shifts (**Fig. 4**). Similar labeling patterns for leucine plus valine residues can be obtained using the compounds shown in **Figure 1e,h,k** (protons at both methyl groups) or in **Figure 1f,i,l** (if the second methyl is of the  $^{12}CD_3$  variety).

**ACKNOWLEDGMENTS** The authors gratefully acknowledge many of the former members of the Kay laboratory who have contributed to the development of the selective methyl protonation approaches described here. V.T. is a recipient of a Canadian Institutes of Health Research Postdoctoral Fellowship. L.E.K. is the recipient of a Canada Research Chair in Biochemistry.

**COMPETING INTERESTS STATEMENT** The authors declare that they have no competing financial interests.

Published online at <http://www.natureprotocols.com/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Tugarinov, V., Hwang, P.M. & Kay, L.E. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu. Rev. Biochem.* **73**, 107–146 (2004).
- Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin. Struct. Biol.* **9**, 594–601 (1999).
- Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated  $T_2$  relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
- Rosen, M.K. *et al.* Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **263**, 627–636 (1996).
- Gardner, K.H., Zhang, X., Gehring, K. & Kay, L.E. Solution NMR studies of a 42 kDa *E. coli* maltose binding protein/ $\beta$  cyclodextrin complex: chemical shift assignments and analysis. *J. Am. Chem. Soc.* **120**, 11738–11748 (1998).
- Tugarinov, V., Muhandiram, R., Ayed, A. & Kay, L.E. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *J. Am. Chem. Soc.* **124**, 10025–10035 (2002).
- Tugarinov, V. & Kay, L.E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* **125**, 13868–13878 (2003).
- Rajesh, S. *et al.* A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp. *J. Biomol. NMR* **27**, 81–86 (2003).
- Coughlin, P.E. *et al.* Improved resolution and sensitivity of triple-resonance NMR methods for the structural analysis of proteins by use of a backbone-labeling strategy. *J. Am. Chem. Soc.* **121**, 11871–11874 (1999).
- Metzler, W.J., Wittekind, M., Goldfarb, V., Mueller, L. & Farmer, B.T. Incorporation of  $^1H/^{13}C/^{15}N$ -[Ile,Leu,Val] into a perdeuterated,  $^{15}N$ -labeled protein: potential in structure determination of large proteins by NMR. *J. Am. Chem. Soc.* **118**, 6800–6801 (1996).
- Grzesiek, S., Anglister, J., Ren, H. & Bax, A.  $^{13}C$  line narrowing by  $^2H$  decoupling in  $^2H/^{13}C/^{15}N$ -enriched proteins. Applications to triple resonance 4D J-connectivity of sequential amides. *J. Am. Chem. Soc.* **115**, 4369–4370 (1993).
- Kainosho, M. *et al.* Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57 (2006).
- Janin, J., Miller, S. & Chothia, C. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–164 (1988).
- Tugarinov, V. & Kay, L.E. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBiochem* **6**, 1567–1577 (2005).
- Gardner, K.H. & Kay, L.E. Production and incorporation of  $^{15}N$ ,  $^{13}C$ ,  $^2H$  ( $^1H$ - $\delta 1$  methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119**, 7599–7600 (1997).
- Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. & Kay, L.E. A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta 1$ ) methyl-protonated  $^{15}N$ ,  $^{13}C$ ,  $^2H$ -labeled proteins. *J. Biomol. NMR* **13**, 369–374 (1999).
- Tugarinov, V. & Kay, L.E. An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* **28**, 165–172 (2004).
- Tugarinov, V. & Kay, L.E. Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G. *J. Am. Chem. Soc.* **126**, 9827–9836 (2004).
- Tugarinov, V., Choy, W.Y., Orekhov, V.Y. & Kay, L.E. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. USA* **102**, 622–627 (2005).
- Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E. & Kay, L.E. Cross-correlated relaxation enhanced  $^1H$ - $^{13}C$  NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).
- Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L.E. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J. Am. Chem. Soc.* **126**, 3964–3973 (2004).
- Tugarinov, V., Ollerenshaw, J.E. & Kay, L.E. Probing sidechain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: an application to an 82-kDa enzyme. *J. Am. Chem. Soc.* **127**, 8214–8225 (2005).
- Tugarinov, V. & Kay, L.E. Quantitative  $^{13}C$  and  $^2H$  NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface. *Biochemistry* **44**, 15970–15977 (2005).
- Millet, O., Muhandiram, D.R., Skrynnikov, N.R. & Kay, L.E. Deuterium spin probes of sidechain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in  $^{13}C$ -labeled and fractionally  $^2H$ -enriched proteins in solution. *J. Am. Chem. Soc.* **124**, 6439–6448 (2002).
- Howard, B.R., Endrizzi, J.A. & Remington, S.J. Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: mechanistic implications. *Biochemistry* **39**, 3156–3168 (2000).
- McDonald, P. & Argos, P. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99–122 (1988).
- Ollerenshaw, J.E., Tugarinov, V., Skrynnikov, N.R. & Kay, L.E. Comparison of  $^{13}CH_3$ ,  $^{13}CH_2D$ , and  $^{13}CHD_2$  methyl labeling strategies in proteins. *J. Biomol. NMR* **33**, 25–41 (2005).