# Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins.

Rime Kerfah, Michael J Plevin, Rémy Sounier, Pierre Gans, Jérôme Boisbouvier

▶ **To cite this version:**

Rime Kerfah, Michael J Plevin, Rémy Sounier, Pierre Gans, Jérôme Boisbouvier. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins.. Current Opinion in Structural Biology, 2015, 32, pp.113-122. 10.1016/j.sbi.2015.03.009 . hal-01162354

**HAL Id: hal-01162354**

**https://hal.univ-grenoble-alpes.fr/hal-01162354v1**

Submitted on 21 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

# Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins

Rime Kerfah[1,2,3,6], Michael J Plevin[4], Remy Sounier[5], Pierre Gans[1,2,3] and Jerome Boisbouvier[1,2,3]

Nuclear magnetic resonance (NMR) spectroscopy is a uniquely powerful tool for studying the structure, dynamics and interactions of biomolecules at atomic resolution. In the past 15 years, the development of new isotopic labeling strategies has opened the possibility of exploiting NMR spectroscopy in the study of supra-molecular complexes with molecular weights of up to 1 MDa. At the core of these isotopic labeling developments is the specific introduction of [$^{1}$H,$^{13}$C]-labeled methyl probes into perdeuterated proteins. Here, we describe the evolution of these approaches and discuss their impact on structural and biological studies. The relevant protocols are succinctly reviewed for single and combinatorial isotopic-labeling of methyl-containing residues, and examples of applications on challenging biological systems, including high molecular weight and membrane proteins, are presented.

**Addresses**
[1] University Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38044 Grenoble, France
[2] CNRS, F-38044 Grenoble, France
[3] CEA, DSV, F-38044 Grenoble, France
[4] Department of Biology, University of York, York YO10 5DD, UK
[5] Institut de Génomique Fonctionnelle – CNRS – INSERM – University Montpellier I & II, F-34094 Montpellier, France

Corresponding author: Boisbouvier, Jerome (jerome.boisbouvier@ibs.fr)
[6] Current address: NMR-Bio, IBS/CEA, 71 Avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9, France.

**Current Opinion in Structural Biology** 2015, **32**:113–122

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Imre Berger** and **Roslyn M Bill**

http://dx.doi.org/10.1016/j.sbi.2015.03.009

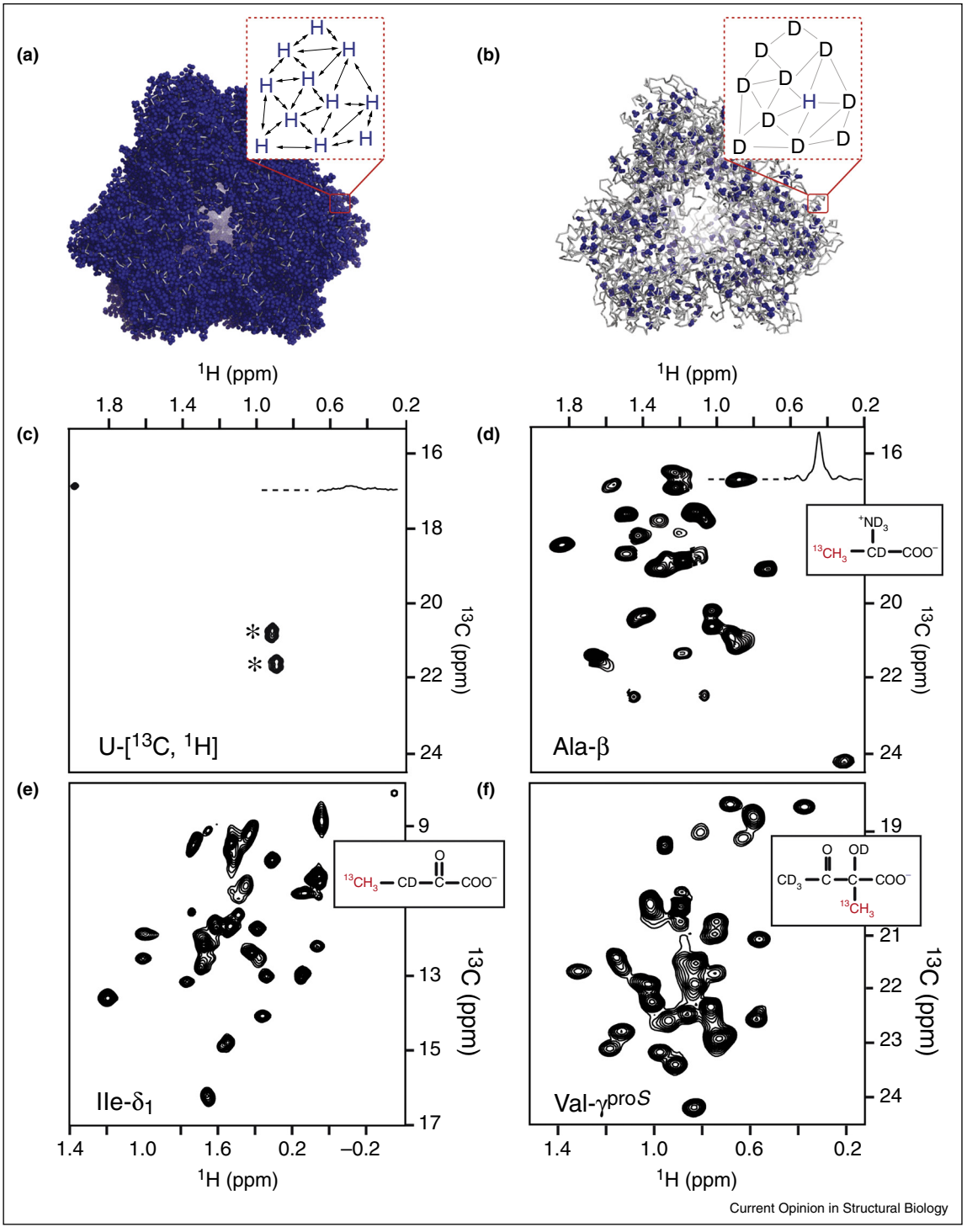
0959-440X/© Elsevier Ltd. All rights reserved.

Nuclear magnetic resonance (NMR) spectroscopy is an established and powerful tool for studying the structure, dynamics and interactions of biomolecules at atomic resolution. The contribution of NMR spectroscopy to the study of high molecular weight proteins has evolved considerably since the mid-1990s. Advances in NMR hardware (e.g., cryogenic probes and higher field NMR spectrometers), the introduction of optimized NMR experiments (e.g., TROSY [1]) and the use of deuterated [$^{13}$C,$^{15}$N]-labeled samples [2–6] have enabled NMR studies of proteins up to ca. 50 kDa. Strategies for the specific introduction of [$^{1}$H,$^{13}$C]-labeled methyl groups in perdeuterated proteins have substantially extended that upper molecular weight limit. Indeed, these advances have permitted solution NMR studies of supra-molecular complexes up to 1 MDa, protein targets that were previously inaccessible to the technique.

NMR-driven studies of high molecular weight systems can provide site-resolved information on local structure, thermodynamics, kinetics, molecular dynamics, post-translational modifications and function. Much of the groundbreaking work in the NMR analysis of supramolecular assemblies has been conducted on the proteasome. The 20S core proteasome particle consists of two subunits, α and β. Each subunit forms a heptameric ring, which assemble to form a $\alpha_7\beta_7\beta_7\alpha_7$ oligomer. Methyl-specific labeling and methyl-TROSY NMR [7•] approaches developed and applied by Kay and colleagues have been used to characterize oligomerization, substrate binding and protein dynamics as well as screening for small molecule inhibitors [8••,9–12]. One elegant example, which also highlights the enormous experimental flexibility that comes from analyzing these protein complexes in solution, concerns the measurement of p$K_a$ values for catalytic groups in the 20S core proteasome particle [13•]. Despite a number of high-resolution structural studies, there had been ongoing debate about the identity of the base that deprotonates the catalytic hydroxyl group of Thr-1 of the β-subunit. NMR analysis of ionization equilibria in smaller proteins is routine. Application of this approach to the 20S core particle required a sample with threonine-specific methyl labeling [13•]. The pH-dependent changes in the chemical shift of the $\gamma_2$-methyl group of Thr-1 were fitted to different models describing the ionization of neighboring functional groups. These analyses were consistent with a reaction mechanism in which the N-terminal α-amino group deprotonates the γ-OH group of the catalytic threonine.

NMR spectroscopy can report information on molecular dynamics across a range of time scales, from picoseconds to hours. Typically, these data are accessed by measuring how an NMR signal relaxes or changes with time. Methyl-specific labeling and methyl-TROSY experiments permit NMR relaxation studies of larger proteins and supramolecular assemblies. A further strength of NMR spectroscopy is

**Figure 1**




Examples of specific methyl-labeling of the 468 kDa amino peptidase TET2 from *Pyrococcus horikoshii*. Proton densities presented on the 3D structure of the homododecameric TET2 protein (PDB code: 1Y0R) for **(a)** U-[$^1$H,$^{13}$C] TET2 sample and **(b)** U-[D] TET2 with specific [$^1$H,$^{13}$C]-labeling of the β-methyl group of alanine. 2D ($^1$H,$^{13}$C)-HMQC spectra of samples of **(c)** U-[$^1$H,$^{13}$C] TET2 compared to U-[D] TET2 with specific [$^1$H,$^{13}$C]-labeling of **(d)** the β-methyl group of alanine using alanine as precursor [37•], **(e)** the $\delta_1$-methyl group of isoleucine using 2-ketobutyrate [20••], and **(f)** the pro*S* methyl group of valine using 2-acetolactate [30]. The 1D traces extracted at the $^{13}$C frequency of Ala-344 are shown in (c) and (d). In the U-[$^1$H, $^{13}$C] TET2 sample, the intense dipolar interactions with nearby protons broaden the lines beyond the threshold of detection (c); whereas in the U-[D] TET2 sample specifically protonated on the β-methyl group of alanine, the

its ability to provide site-resolved information for intrinsically disordered proteins or regions of proteins. These two strengths were combined in a study of the gating mechanism of the 20S proteasome [14•]. The degradation of protein substrates occurs in a central cavity formed by the two 7-mer β rings. The catalytic cavity is accessed via pores formed by the two 7-mer α rings. Access through the pore is regulated by the N-termini of the α-subunits, which are flexible and which had not been resolved in crystallographic studies. Methionine ε-methyl-specific labeling was used to probe the dynamics of the N-terminal gating regions. Resonance assignment was performed through site-by-site Met-to-Ala mutation. Methyl-TROSY spectra revealed that certain methionine residues yielded multiple NMR signals that corresponded to whether the N-terminus was inside or outside the proteasome pore. Analysis of the stoichiometry revealed that, on average, two of the seven termini were in the pore, while five were not. NMR EXchange SpectroscopY (EXSY) demonstrated that these states interconvert on the seconds time-scale as well as revealing longer-distance allosteric affects, which were detected over 80 Å from the proteasome pore entrance.

The study of longer time-scale protein dynamics can also be probed by NMR spectroscopy. Real-time analysis of enzyme turnover [15] or protein folding [16] can be performed at site-resolved resolution through the collection of a series of 2D NMR experiments. The upper molecular weight limit of this approach was enhanced greatly by the optimization of the SOFAST HMQC experiment [17] for sparsely-protonated, methyl-labeled high molecular weight proteins. Consequently, 2D ($^{1}$H,$^{13}$C) correlation spectra of very large proteins can be acquired on a second time-scale [18]. Rizo and colleagues reported a nice example of the use of real-time NMR studies of protein function [19]. Using 2D methyl TROSY NMR experiments, they showed that the rate at which the Syntaxin-1 protein restructures to form an active SNARE complex is affected by other protein components of the synaptosome.

Below, we will review the methods for selective protonation and isotopic labeling of methyl groups that have permitted such studies. We will also highlight how methyl-labeling can be applied to study the structure of larger proteins.

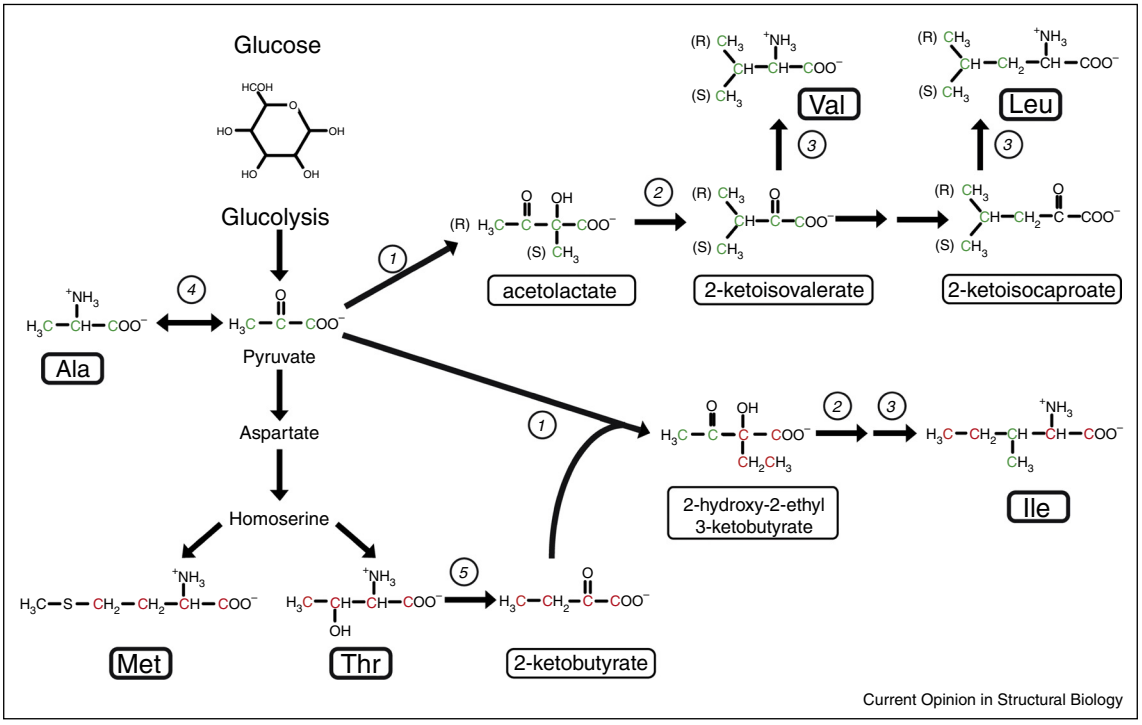
## Development of protocols for isotopic labeling of methyl groups in proteins

NMR spectroscopy is a well-established technique for characterizing the structure, dynamics and function of proteins of less than 30 kDa. Spectroscopists have access to an enormous array of multidimensional heteronuclear NMR experiments designed for [$^{13}$C,$^{15}$N]-labeled proteins. Above 30 kDa, the reduction in molecular tumbling rate increases the rate at which NMR signal relaxes, causing signal broadening and compromising experimental sensitivity. In higher molecular weight proteins, the major mechanism that drives $^{1}$H NMR relaxation is dipolar interactions between the large number of neighboring protons (Figure 1a). For proteins of 30–80 kDa, detrimental dipolar interactions can be reduced by expressing the protein in a perdeuterated culture medium. A basal level of protons, typically around 20%, is then reintroduced at labile sites (e.g., NH, OH, $NH_3^+$, among others), by purifying or, if necessary, refolding the protein in $H_2O$-based buffers. The resulting protein has protonated amide groups, which ensures that backbone-directed NMR experiments can still be applied to proteins in this size range.

To apply solution NMR spectroscopy to proteins above 100 kDa requires an even greater reduction in overall level of protonation (Figure 1a–d). However, to benefit from the high intrinsic sensitivity of the $^{1}$H nucleus it is necessary to retain some residual protonation at key sites. Methyl groups are ideal candidates for NMR studies of high molecular weight proteins, as they yield highly intense and well-resolved NMR signals due to the multiplicity of protons and the rapid rotation about the three-fold methyl symmetry axis. Furthermore, methyl groups are enriched in the hydrophobic cores of proteins and at protein–protein interfaces. To reduce $^{1}$H signal relaxation rates requires a protein sample in which only selected methyl groups are protonated while the remainder of the protein is completely deuterated. Such a labeling scheme can be achieved by supplementing a fully deuterated *Escherichia coli* minimal culture medium with one or more specifically [$^{13}CH_3$]-labeled amino acids or biosynthetic precursors (Figure 2). The approach used to produce methyl labeled proteins depends on the biosynthetic pathway of the target methyl group, with the simplest cases involving amino acids that are the products of irreversible metabolic pathways. The earliest methyl-specific labeling protocol reported in the literature concerned the $\delta_1$-methyl group of isoleucine [20••]. The precursor 2-keto-3-[$D_2$],4-[$^{13}$C]-butyrate was used as the sole source of protons in a perdeuterated culture media, to produce a [U-D], Ile-$\delta_1$-[$^{13}CH_3$]-labeled protein (Figure 1e). A similar strategy was subsequently developed to isotopically label the prochiral methyl groups of leucine and valine. This protocol used 2-keto-3-[D]-[$^{13}CH_3$,$^{13}CH_3$]-isovalerate, a precursor of both leucine and valine, which labels both methyl groups in both amino

**(Figure 1 Legend Continued)** reduction of local proton density makes it possible to recover a narrow NMR signal detectable by solution NMR (d). In (c), the two observable signals annotated by * correspond to the methyl resonances of valine-4 located on the flexible N-terminal end of the TET2 sequence. The insets of (d) to (f) show the chemical structure of the precursor or amino acid used for specific labeling. NMR spectra were acquired at 50 °C on a NMR spectrometer operating at a proton resonance frequency of 800 MHz.

**Figure 2**




Simplified biosynthetic pathway for methyl-containing amino acids in *E. coli* (adapted from KEGG http://www.genome.jp/kegg/kegg2.html). The different precursors cited in the text are indicated. Carbon atoms derived directly from pyruvate or alanine (via deamination of alanine catalyzed by *E. coli* transaminases) are indicated in green, while carbon atoms derived from aspartate are shown in red. Enzymes: (1): α-hydroxy acid synthase (AHAS); (2): ketol-acid reductoisomerase (KARI) and dihydroxy-acid dehydratase (DHAD); (3): branched-chain amino acid aminotransferase (BCAT); (4): *E. coli* alanine transaminases (AlaA, AlaC and AvtA); (5): biosynthetic threonine deaminase (IlvA).

acids [21–23]. Later, 2-keto-3-[D]-[$^{13}CH_3$,$^{12}CD_3$]-isovalerate was used to label only one of the prochiral methyl groups in Leu/Val, which greatly reduced the local proton concentration and thereby methyl relaxation rates [24•, 25,26]. However, this precursor is produced as a racemic mixture and consequently each prochiral methyl group in the protein is labeled at 50%, which means that both sites are detectable by NMR but with reduced sensitivity. Stereospecific labeling of leucine and valine was achieved using 2-acetolactate (2-hydroxy-2-methyl-3-ketobutyrate), a more upstream precursor in the Leu/Val biosynthetic pathway (Figure 2) that could be chemically synthesized to give 100% pro*S* or pro*R* labeling in the final protein product [27••]. Protocols for the labeling of leucine but not valine, and vice versa, have also been proposed. [$^{13}CH_3$]-labeled leucine [28] or 2-ketoisocaproate [29] have been introduced for either stereo-specific or non-stereospecific labeling of leucine. Alternatively, 2-acetolactate [30] or valine [28] can be supplemented in conjunction with deuterated leucine to label solely valine methyl groups (Figure 1f). The Ile-$\gamma_2$ methyl group can be labeled using 2-hydroxy-2-ethyl-3-keto-butanoic acid [31,32], while the ε-methyl group of methionine can be labeled by supplementing the medium with the residue itself [33••,34,35].

Specific protonation of either alanine (Figure 1d) or threonine is more complicated, as both of these amino acids are either synthesized in reversible reactions or are precursors of other methyl-containing amino acids (Figure 2). Typically, alanine [36,37•] and threonine [38•] can be supplemented directly in the M9/$D_2O$ medium with an appropriate cocktail of deuterated metabolic intermediates (Table 1) to suppress scrambling (i.e., leak of $^{13}C$ or/and $^{1}H$ into undesired atomic positions).

## Combinatorial labeling of methyl-containing residues increases the number of NMR probes in large proteins

The high level of deuteration that is needed to detect signals of protonated methyl groups leads to a substantial reduction in the number of NMR-visible probes and consequently to a significant loss of structural information. The six naturally occurring methyl-containing amino acids represent up to 30–40% of the amino acids in proteins, including up to 50% in hydrophobic cores [39] and up to 24% at protein–protein interfaces [40]. Simultaneous labeling of several types of methyl group-containing amino acids is an obvious way to increase the number of NMR-visible probes from which useful information

**Table 1**

**Incorporation of $^{13}CH_3$ groups in perdeuterated proteins**

| Methyl groups | Name of precursors | Quantity (mg/L)[c] | Scrambling/incompatibilities | Ref. |
|---|---|---|---|---|
| Ile-$\delta_1$ | 2-Ketobutyrate or<br>2-(S)-hydroxy-2-ethyl-3-ketobutyrate | 60<br>60 | Both precursors have co-incorporation incompatibilities with 2-acetolactate[d] | [20••]<br>[44•] |
| Ile-$\gamma_2$ | 2-Hydroxy-2-ethyl-3-ketobutyrate (racemic) | 100 | Scrambling (7%) in L/V$^{pro\text{-}R}$ group[e]. Co-incorporation incompatibility with 2-acetolactate[d] | [31], [32] |
| Met-ε | L-Methionine or<br>4-Methyl-thio-2-ketobutyrate | 100–250 | – | [33••]<br>[34] |
| Leu/Val | 2-Ketoisovalerate[a] | 80–120 | – | [24•], [26] |
| Leu/Val$^{proS}$ | 2-Acetolactate[b] | 300 | Co-incorporation i ncompatibility with isoleucine precursors[d] | [27••] |
| Leu | 2-Ketoisocaproate[a] | 150 | – | [29] |
| Leu$^{pro\text{-}S}$ | Stereospecifically labeled L-Leucine[b] | 20 | – | [28] |
| Val$^{pro\text{-}S}$ | 2-Acetolactate[b] +<br>U-[D] L-Leucine | 300<br>20 | Scrambling in L$^{pro\text{-}S}$ groups (<1%). Co-incorporation incompatibility with isoleucine precursors[d] | [30] |
| Val$^{pro\text{-}S}$ | Stereospecifically labeled Valine[b]<br>+ U-[D] Leucine | 100<br>20 | – | [28] |
| Ala-β | L-Alanine | 600 | Scrambling in Ile-$\gamma_2$, L/V* (up to 25%), minor scrambling in other CH and $CH_2$ sites[f] | [37•], [43], [44•] |
| Thr-$\gamma_2$ | L-Threonine | 50–100 | Scrambling in Ile-$\delta_1$ (ca. 50%), minor scrambling in glycine[g] | [38•] |

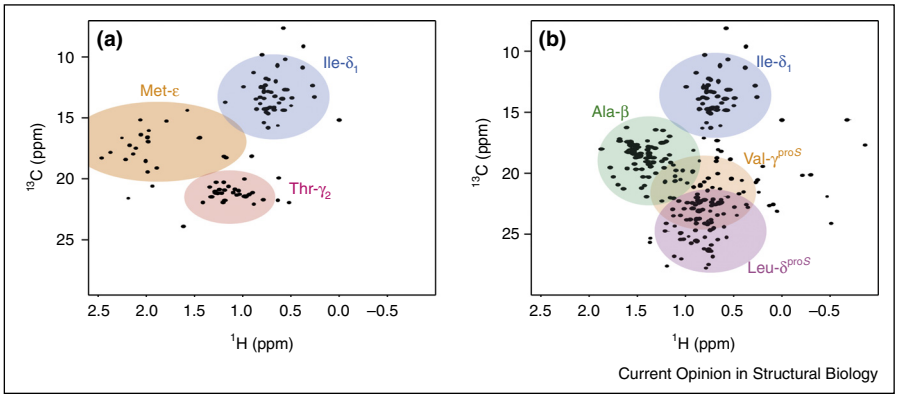
[a] This precursor is produced as a racemic mixture, and consequently overall labeling of each prochiral methyl is 50%, which means that both pro-*S* and pro-*R* sites are detectable by NMR but with reduced sensitivity.
[b] This precursor permits stereospecific labeling of pro-*S* (or pro-*R*) methyl groups. Selected methyl groups are labeled at ca. 100%, enhancing sensitivity for the detection of structural meaningful NOEs between remote methyl groups by up to a factor of 4.
[c] Quantity of precursors to add 1 hour prior to induction of protein expression in M9/$D_2O$ media containing 2 g/L of deuterated glucose. Variation of carbon sources may interfere with incorporation of specifically labeled precursors into overexpressed protein.
[d] Co-incorporation incompatibilities can be suppressed by addition of 2-acetolactate 40 minutes before isoleucine precursor (added 20 minutes before induction) [44•].
[e] Scrambling suppressed by co-addition of 200 mg/L U-[D] 2-ketoisovalerate [32].
[f] Scrambling suppressed by co-addition of 200 mg/L U-[D] 2-ketoisovalerate, 60 mg/L U-[D]-isoleucine or 2-(S)-hydroxy-2-ethyl-3-ketobutanoic acid, 2.5 g/L U-[D]-succinate and 2.5 g/L U-[D]-glycerol [37•,44•].
[g] Scrambling suppressed by co-addition of 60 m/L of U-[D]-isoleucine + 100 mg/L U-[D] glycine [38•].

can be obtained. A number of combinatorial [$^{13}CH_3$]-methyl-labeling schemes have been reported, including ILV (Ile-$\delta_1$/Leu-δ/Val-γ) [22,26,41,42], MILV (Met-ε/Ile-$\delta_1$/Leu-δ/Val-γ) [33••], AILV (Ala-β/Ile-$\delta_1$/Leu-δ/Val-γ) without [43] or with stereospecific labeling of leucine and valine [44•], and MILVT (Met-ε/Ile-$\delta_1$/Leu-δ/Val-γ/Thr-$\gamma_2$) [45••].

Methyl groups resonate in a narrow region of 2D ($^1H$,$^{13}C$) NMR spectra. As a consequence, the motivations for combining different methyl groups must be considered:

- *The residue type-dependence of [$^{13}CH_3$] resonance dispersion*. Signal overlap can be alleviated by choosing suitable amino acids to label simultaneously [46]. Figure 3a illustrates an example of a well-resolved 2D methyl-TROSY spectrum of malate synthase G (MSG; 82 kDa) in which Met-ε, Ile-$\delta_1$ and Thr-$\gamma_2$ methyl groups have been labeled simultaneously. These methyl groups typically resonate in non-overlapping regions of the spectrum.
- *Specific labeling of one methyl group per residue*. The regio-specific or stereo-specific labeling of a single methyl group in Ile, Val or Leu can greatly increase sensitivity and resolution [20••,27••,28,30–32,44•] (Figure 3b). Optimal resonance dispersion is achieved by labeling the pro*S* methyl groups of leucine and valine at the same time as the Ile-$\delta_1$ methyl group, as these groups

**Figure 3**




Scrambling-free combinatorial isotopic labeling to optimize sensitivity and resolution. 2D ($^{1}H$,$^{13}C$) HMQC spectra of **(a)** U-[D], (Ile-$\delta_1$, Met-ε, Thr-$\gamma_2$)-[$^{13}CH_3$] or **(b)** U-[D], (Ala-β, Ile-$\delta_1$, Leu/Val$^{proS}$)-[$^{13}CH_3$]-labeled MSG [44•]. Spectra were acquired at 37 °C on a NMR spectrometer operating at a proton resonance frequency of 800 MHz. The colored ellipses represent the location of methyl resonances for each type of methyl group. The *x* and *y* boundaries of the semi-transparent ellipses demonstrate the second standard deviations associated with each chemical shift. ($^{1}H$,$^{13}C$) chemical shifts were taken from the BioMagResBank (BMRB, http://www.bmrb.wisc.edu).

tend to have wider chemical shift dispersion [47] and limited overlap.

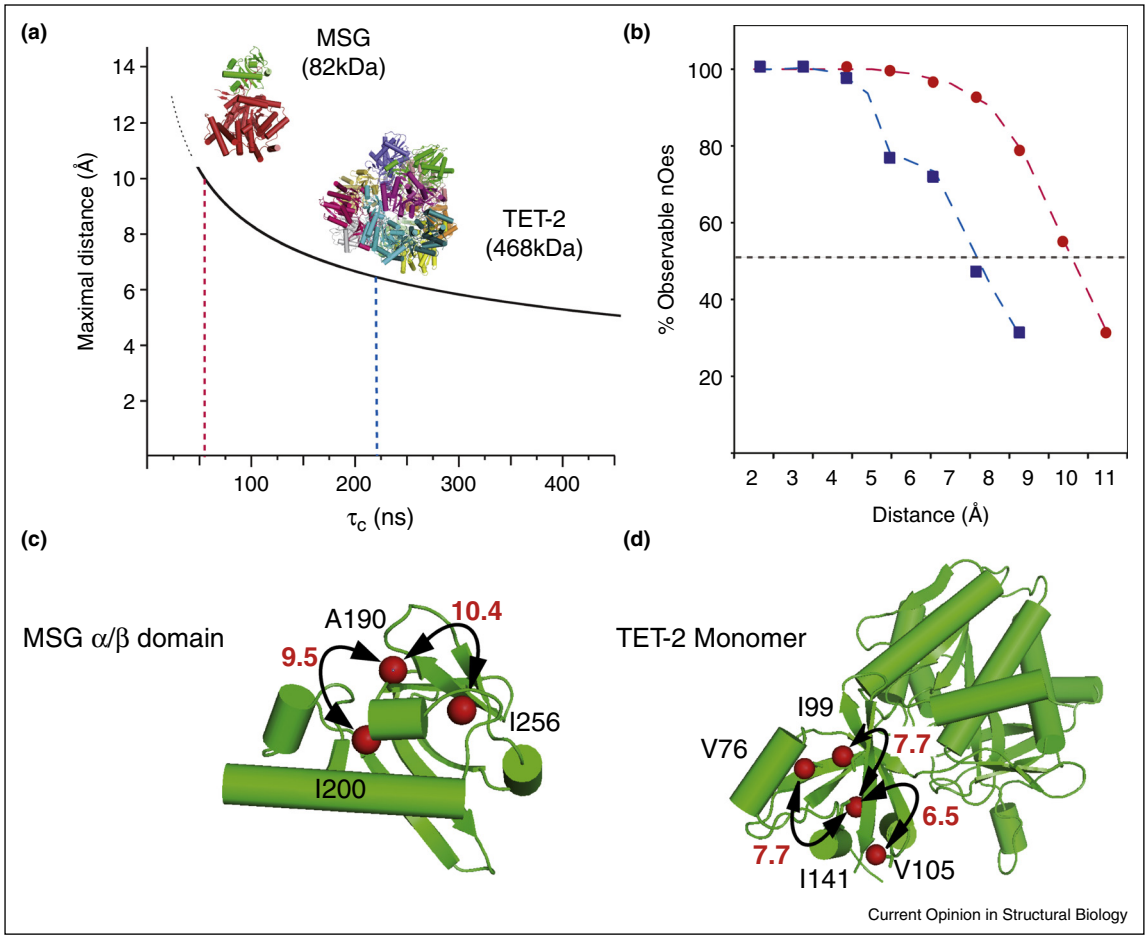
- *Prevention of co-incorporation incompatibilities of precursors*. If the metabolic precursors used for combinatorial labeling share the same metabolic pathway, the enzymatic machinery can exhibit preference towards one precursor over others. Combinatorial labeling of Ile/Val/Leu with 2-ketobutyrate and 2-acetolactate is affected by this phenomenon. Incompatibility of the two precursors (Table 1) leads to a 2-fold reduction in enrichment at the Leu/Val$^{proS}$ site unless 2-acetolactate is added to the media before 2-ketobutyrate [44•].
- *Prevention or exploitation of isotope scrambling*. Amino acids that are intermediates or involved in reversible metabolic pathways generate isotope scrambling. This isotopic leak can be suppressed, or exploited if advantageous (Table 1). For example, isotopic scrambling to Ile-$\gamma_2$ methyl groups occurs when 3-[$^{13}$C]-Ala and 2-ketobutyrate are used simultaneously [37•]. 3-[$^{13}$C]-Ala is converted into 3-[$^{13}$C]-pyruvate, which causes 2–5% scrambling in Isoleucine $\gamma_2$ position [44•]. The replacement of 2-keto-3-[$D_2$]-4-[$^{13}$C]butyrate with 2-hydroxy-2-[2′-($^{13}$C)-1′-($D_2$)]ethyl-3-keto-4-[$D_3$]-butyrate resolves this problem. Likewise, the activity of threonine deaminase causes enrichment of Ile-$\delta_1$ methyl groups when the medium is supplemented with 2-[D]-3-[D]-4-[$^{13}$C]-Thr [38•]. This scrambling can be suppressed by adding deuterated isoleucine to retro-inhibit threonine deaminase or exploited by labeling both Ile-$\delta_1$ and Thr-$\gamma_2$ sites simultaneously.

## NMR structural analysis of methyl labeled protein samples

The Nuclear Overhauser Effect (NOE) is a key observable for 3D structure determination by solution NMR spectroscopy. In smaller proteins, NOEs can be detected between $^{1}H$–$^{1}H$ pairs separated by up to 6–7 Å. However, the sparse density of protons in selectively methyl-protonated proteins has allowed the detection of accurate long-range methyl/methyl distances of up to 12 Å in 10–20 kDa protein [48]. Although NOE magnetization transfer is more efficient for large proteins, the detection of methyl/methyl NOEs in such systems is limited by the signal broadening that results from faster relaxation (Figure 4). Nonetheless, NOE cross-peaks can be detected for methyl groups separated by up to 10 Å, in MSG (82 kDa, 37 °C; $\tau_c \approx 50$ ns; [44•]), while in a 0.5 MDa protein complex, it is possible to detect NOEs between methyl groups separated by 7–8 Å (50 °C; $\tau_c \approx 220$ ns [30]).

Combinatorial selective protonation of methyl groups has underpinned numerous structural studies of high molecular weight proteins. The first large protein whose structure was determined using restraints from ($^{13}CH_3$)-NMR spectroscopy was the 82 kDa monomeric protein MSG [49••]. Methyl-TROSY NMR studies have also contributed significantly to structural studies of challenging membrane proteins. Examples include the structure of the voltage-dependent anion channel (VDAC), an integral membrane protein that allows the diffusion of small molecules across the eukaryotic outer mitochondrial membrane [50], and the hexameric p7 cation channel from hepatitis C virus [51]. Methyl NMR approaches were also used to elucidate the 3D structure of the phototaxis receptor sensory rhodopsin II, a seven transmembrane-helix protein that exhibits many structural similarities to G-protein coupled receptors [52]. In larger protein assemblies, specific labeling of methyl probes has been used to model the interaction between functional

**Figure 4**




Detection of long-range methyl–methyl NOEs in high molecular weight proteins. **(a)** Variation of the *maximal distance* for which methyl/methyl NOEs are detectable as a function of the correlation time of the protein. Methyl probes separated by the *maximum distance* have a 50% probability of yielding observable NOEs. **(b)** Distribution of experimentally detected inter-methyl NOEs in U-[D], (Ala-β, Ile-$\delta_1$, Leu/Val$^{proS}$)-[$^{13}CH_3$]-MSG [44•] and U-[D], (Ile-$\delta_1$, Val$^{proS}$)-[$^{13}CH_3$]-TET2 systems [30]. 3D HMQC-NOESY data sets were recorded on a Varian (Agilent) DirectDrive spectrometer operating at a proton frequency of 800 MHz equipped with a cryogenic triple resonance probehead. The experiments were recorded at 37 °C for 96 hours with a 1 mM U-[D], (Ala-β, Ile-$\delta_1$, Leu/Val$^{proS}$)-[$^{13}CH_3$]-MSG sample and a NOE mixing time of 500 ms, (50 °C, 64 hours and 400 ms for U-[D], (Ile-$\delta_1$, Val$^{proS}$)-[$^{13}CH_3$]-TET2 sample). Each point represents the percentage of experimentally detected NOEs between pairs of methyl groups for a given distance (±0.5 Å). The total number of methyl/methyl pairs for a given distance was predicted using the 3D structures of MSG (82 kDa; red curve; PDB code: 1D8C) or TET2 (468 kDa; blue curve; PDB code: 1Y0R). Examples of methyl/methyl pairs for which long-range NOEs were experimentally detected in MSG **(c)** and TET2 **(d)**.

signal-peptides and the 204 kDa translocase SecA [33••], to characterize the interaction between the trigger factor chaperone and substrates at the atomic level [45••], and to model the structure of the 390 kDa box C/D RNP enzyme bound to substrate RNA [53••].

## Extending methyl-specific labeling strategies

Specific isotopic labeling and protonation of methyl groups in very large perdeuterated proteins is essential for their investigation by solution NMR spectroscopy. However, the possible drawbacks of this approach need to be considered. The first concerns the restricted number of observable NMR probes compared to standard uniform labeling. As discussed, it is possible to reduce this effect through combinatorial labeling schemes. In addition, non-native methyl groups can be introduced at strategic sites through mutation [14•,35] or by chemical modification of lysine [54] or cysteine [55] residues.

Perhaps the largest obstacle is the requirement to produce milligram quantities of protein in deuterated culture medium. $D_2O$ negatively affects bacterial growth and can severely impact recombinant protein expression and solubility to such an extent that methyl labeling ceases to be economically viable. Protocols for the methyl labeling of perdeuterated proteins in the yeast *Kluyveromyces lactis*

have been proposed [56] as well as approaches for methyl-specific labeling via the insect cell/baculovirus expression system [57••]. Cell-free protein expression alleviates the requirement that the expression organism must be viable in perdeuterated expression medium while also offering tremendous control over the isotopic composition of the reaction mixture and therefore the final protein. In addition, cell-free protein synthesis can be tuned to maximize the production of challenging systems, for example, improving membrane protein production through the addition of lipids and detergents to the reaction mixture [58–60].

Residue-selective protonation of methyl groups in perdeuterated proteins opens the possibility of NMR investigations of the structure, dynamics and function of challenging biological systems at atomic resolution. A growing array of isotopically-enriched metabolic precursors is now available for isotopic-labeling of different types of methyl groups in protein targets in *E. coli*. In the next decade, advances in emerging approaches for methyl labeling, such as cell-free protein synthesis or eukaryotic expression systems, will be key to ensuring that greater numbers of biologically important protein targets can be investigated by solution NMR spectroscopy.

## Conflict of interest

None declared.

## Acknowledgements

We would like to thank I. Ayala for stimulating discussions and assistance in sample preparation, G. Mas for providing NOEs restraints observed in TET2, and Dr O. Hamelin for providing [$^{13}$C, D]-labeled Threonine. This work used the high-field NMR and the isotopic labeling facilities at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 Grant Agreement no. 260887.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Pervushin K, Riek R, Wider G, Wuthrich 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* 1997, **94**:12366-12371.

2. Yamazaki T, Lee L, Arrowsmith CH, Muhandiram DR, Kay LE: **A suite of triple resonance NMR experiments for the backbone assignment of $^{15}$N, $^{13}$C, $^{2}$H labeled proteins with high sensitivity**. *J Am Chem Soc* 1994, **116**:11655-11666.

3. Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, Nettesheim D, Chang BS, Thompson CB, Wong SL, NG SL, Fesik SW: **X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death**. *Nature* 1996, **381**:335-341.

4. Venters RA, Huang CC, Farmer BT, Trolard R, Spicer LD, Fierke CA: **High-level $^{2}$H/$^{13}$C/$^{15}$N labeling of proteins for NMR studies**. *J Biomol NMR* 1995, **5**:339-344.

5. Garrett DS, Seok YJ, Liao DI, Peterkofsky A, Gronenborn AM, Clore GM: **Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR**. *Biochemistry* 1997, **36**:2517-2530.

6. Gardner KH, Kay LE: **The use of $^{2}$H, $^{13}$C, $^{15}$N multidimensional NMR to study the structure and dynamics of proteins**. *Annu Rev Biophys Biomol Struct* 1998, **27**:357-406.

7. • Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE: **Cross-correlated relaxation enhanced $^{1}$H–$^{13}$C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes**. *J Am Chem Soc* 2003, **125**:10420-10428.
Description of interference effects in $^{13}CH_3$ groups of perdeuterated proteins and introduction of methyl TROSY experiment.

8. •• Sprangers R, Kay LE: **Quantitative dynamics and binding studies of the 20S proteasome by NMR**. *Nature* 2007, **445**:618-622.
Application of methyl-specific labeling methods and methyl TROSY experiments to the study of the dynamics of the 670 kDa proteasome and its interaction with the 11S activator.

9. Sprangers R, Li X, Mao X, Rubinstein JL, Schimmer AD, Kay LE: **TROSY-based NMR evidence for a novel class of 20S proteasome inhibitors**. *Biochemistry* 2008, **47**:6727-6734.

10. Baldwin AJ, Religa TL, Hansen DF, Bouvignies G, Kay LE: **$^{13}CHD_2$ methyl group probes of millisecond time scale exchange in proteins by $^{1}$H relaxation dispersion: an application to proteasome gating residue dynamics**. *J Am Chem Soc* 2010, **132**:10992-10995.

11. Hansen DF, Kay LE: **Determining valine side-chain rotamer conformations in proteins from methyl $^{13}$C chemical shifts: application to the 360 kDa half-proteasome**. *J Am Chem Soc* 2011, **133**:8272-8281.

12. Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE: **The proteasome antechamber maintains substrates in an unfolded state**. *Nature* 2010, **467**:868-871.

13. • Velyvis A, Kay LE: **Measurement of active site ionization equilibria in the 670 kDa proteasome core particle using methyl-TROSY NMR**. *J Am Chem Soc* 2013, **135**:9259-9262.
Investigation of the active site of the proteasome using specific labeling of threonine methyl groups and application to the study of catalytic mechanism.

14. • Religa TL, Sprangers R, Kay LE: **Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR**. *Science* 2010, **328**:98-102.
Dynamics studies of the N-terminal sequence of proteasome a-subunit using methyl-specific labeling of methionine and application to the study of gating mechanism.

15. Marshall CB, Ho J, Buerger C, Plevin MJ, Li G, Li Z, Ikura M, Stambolic V: **Characterization of the intrinsic and TSC2-GAP-regulated GTPase activity of Rheb by real-time NMR**. *Sci Signal* 2009, **2**:ra3 http://dx.doi.org/10.1126/scisignal.2000029.

16. Rennella E, Cutuil T, Schanda P, Ayala I, Forge V, Brutscher B: **Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate**. *J Am Chem Soc* 2012, **134**:8066-8069.

17. Schanda P, Brutscher B: **Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds**. *J Am Chem Soc* 2005, **127**:8014-8015.

18. Amero C, Schanda P, Dura MA, Ayala I, Marion D, Franzetti B, Brutscher B, Boisbouvier J: **Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies**. *J Am Chem Soc* 2009, **131**:3448-3449.

19. Ma C, Li W, Xu Y, Rizo J: **Munc13 mediates the transition from the closed Syntaxin–Munc18 complex to the SNARE complex**. *Nat Struct Mol Biol* 2011, **18**:542-549.

20. •• Gardner KH, Kay LE: **Production and incorporation of $^{15}$N, $^{13}$C, $^{2}$H ($^{1}$H-$\delta_1$ methyl) isoleucine into proteins for multidimensional NMR studies**. *J Am Chem Soc* 1997, **119**:7599-7600.
First description of an efficient protocol for specific labeling of methyl groups with $^{13}CH_3$ isotopomer in perdeuterated proteins using 2-keto acid precursors.

21. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE: **A robust and cost-effective method for the production of Val, Leu, Ile (delta-1) methyl-protonated $^{15}$N-, $^{13}$C-, $^{2}$H-labeled proteins**. *J Biomol NMR* 1999, **13**:369-374.

22. Gross JD, Gelev VM, Wagner G: **A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes**. *J Biomol NMR* 2003, **25**:235-242.

23. Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang J, Betz SF, Fesik SW: **NMR-based screening of proteins containing $^{13}$C-labeled methyl groups**. *J Am Chem Soc* 2000, **122**:7898-7904.

24. • Tugarinov V, Kay LE: **An isotope labeling strategy for methyl TROSY spectroscopy**. *J Biomol NMR* 2004, **28**:165-172.
Introduction of 2-keto-3-[D]-[$^{13}CH_3$,$^{12}CD_3$]-isovalerate, which suppresses intra-residue methyl/methyl dipolar interactions and enhances resolution of methyl-TROSY spectra.

25. Tugarinov V, Kanelis V, Kay LE: **Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy**. *Nat Protoc* 2006, **1**:749-754.

26. Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R: **Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing**. *J Am Chem Soc* 2004, **126**:5348-5349.

27. •• Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J: **Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins**. *Angew Chem Int Ed* 2010, **49**:1958-1962.
Description of a protocol to stereo-selectively label leucine and valine methyl groups. This strategy reduces overlap and enhances NMR signal intensity by up to a factor of 4.

28. Miyanoiri Y, Takeda M, Okuma K, Ono A, Terauchi T, Kainosho M: **Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereo-specifically methyl labeled amino acids**. *J Biomol NMR* 2013, **57**:237-249.

29. Lichtenecker R, Coudevylle N, Konrat R, Schmid W: **Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds**. *Chembiochem* 2013, **14**:818-821.

30. Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J: **Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins**. *J Biomol NMR* 2013, **57**:251-262.

31. Ruschak AM, Velyvis A, Kay LE: **A simple strategy for $^{13}$C, $^{1}$H labeling at the Ile-gamma-2 methyl position in highly deuterated proteins**. *J Biomol NMR* 2010, **48**:129-135.

32. Ayala I, Hamelin O, Amero C, Pessey O, Plevin MJ, Gans P, Boisbouvier J: **An optimized isotopic labelling strategy of isoleucine-gamma2 methyl groups for solution NMR studies of high molecular weight proteins**. *Chem Commun (Camb)* 2012, **48**:1434-1436.

33. •• Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG: **Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR**. *Cell* 2007, **131**:756-769.
Introduction of the combinatorial labeling of ILVM methyl groups and application to the structural study of a complex of 204 kDa.

34. Fischer M, Kloiber K, Hausler J, Ledolter K, Konrat R, Schmid W: **Synthesis of a $^{13}$C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy**. *Chembiochem* 2007, **8**:610-612.

35. Stoffregen MC, Schwer MM, Renschler FA, Wiesner S: **Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces**. *Structure* 2012, **20**:573-581.

36. Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P, Matthews S: **A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues**. *J Am Chem Soc* 2007, **129**:15428-15429.

37. • Ayala I, Sounier R, Use N, Gans P, Boisbouvier J: **An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein**. *J Biomol NMR* 2009, **43**:111-119.
Scrambling-free and optimal incorporation of $^{13}CH_3$ isotopomer of alanine in proteins overexpressed in M9/$D_2O$ medium.

38. • Velyvis A, Ruschak AM, Kay LE: **An economical method for production of $^{2}$H, $^{13}CH_3$-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome**. *PLoS One* 2012, **7**:e43725.
Introduction of an enzymatic scheme to prepare 2-[D]-3-[D]-4-[$^{13}$C]-Thr for incorporation into perdeuterated proteins.

39. Miller S, Janin J, Lesk AM, Chothia C: **Interior and surface of monomeric proteins**. *J Mol Biol* 1987, **196**:641-656.

40. Lo Conte L, Chothia C, Janin J: **The atomic structure of protein–protein recognition sites**. *J Mol Biol* 1999, **285**:2177-2198.

41. Tugarinov V, Kay LE: **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* 2003, **125**:13868-13878.

42. Velyvis A, Schachman HK, Kay LE: **Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase**. *J Am Chem Soc* 2009, **131**:16534-16543.

43. Godoy-Ruiz R, Guo C, Tugarinov V: **Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins**. *J Am Chem Soc* 2010, **132**:18340-18350.

44. • Kerfah R, Plevin MJ, Pessey O, Hamelin O, Gans P, Boisbouvier J: **Scrambling free combinatorial labeling of alanine-β, isoleucine-$\delta_1$, leucine-proS and valine-proS methyl groups for the detection of long range NOEs**. *J Biomol NMR* 2015, **61**:73-82.
Scrambling-free use and optimization of co-incorporation of precursors for the labeling of $A^{\beta}I^{\delta1}(LV)^{proS}$ methyl groups in protein overexpressed in *E. coli*.

45. •• Saio T, Guan X, Rossi P, Economou A, Kalodimos CG: **Structural basis for protein anti-aggregation activity of the trigger factor chaperone**. *Science* 2014, **344**:1250494.
Structure determination of 100 kDa complexes, based on the combinatorial labeling of ILVMT methyl groups.

46. Plevin MJ, Boisbouvier J: **Isotope-labelling of methyl groups for NMR studies of large proteins**. In *Recent Developments in Biomolecular NMR*. Edited by Clore M, Potts J. Royal Society of Chemistry; 2012:1-24 http://dx.doi.org/10.1039/9781849735391-00001. (Chapter 1).

47. London RE, Wingad BD, Mueller GA: **Dependence of amino acid side chain $^{13}$C shifts on dihedral angle: application to conformational analysis**. *J Am Chem Soc* 2008, **130**:11097-11105.

48. Sounier R, Blanchard L, Wu Z, Boisbouvier J: **High-accuracy distance measurement between remote methyls in specifically protonated proteins**. *J Am Chem Soc* 2007, **129**:472-473.

49. •• Tugarinov V, Choy WY, Orekhov VY, Kay LE: **Solution NMR-derived global fold of a monomeric 82-kDa enzyme**. *Proc Natl Acad Sci USA* 2005, **102**:622-627.
Structure determination of the largest monomeric protein using methyl-based NMR.

50. Hiller S, Malia TJ, Garces RG, Orekhov VY, Wagner G: **Backbone and ILV side chain methyl group assignments of the integral human membrane protein VDAC-1**. *Biomol NMR assignments* 2010, **4**:29-32.

51. OuYang B, Xie S, Berardi MJ, Zhao X, Dev J, Yu W, Sun B, Chou JJ: **Unusual architecture of the p7 channel from hepatitis C virus**. *Nature* 2013, **498**:521-525.

52. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D: **Structure determination of the seven-helical transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy**. *Nat Struct Mol Biol* 2010, **17**:768-774.

53. •• Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlomagno T: **The structure of the box C/D enzyme reveals regulation of RNA methylation**. *Nature* 2013, **502**:519-523.
Structure determination of a 390 kDa RNP complex using methyl-specific labeling of ILV methyl groups with PREs and SANS.

54. Bokoch MP, Zou Y, Rasmussen SG, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi HJ, Thian FS, Kobilka TS *et al.*: **Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor**. *Nature* 2010, **463**:108-112.

55. Religa TL, Ruschak AM, Rosenzweig R, Kay LE: **Site-directed methyl group labeling as an NMR probe of structure and dynamics in supramolecular protein systems: applications to the proteasome and to the ClpP protease**. *J Am Chem Soc* 2011, **133**:9063-9068.

56. Miyazawa-Onami M, Takeuchi K, Takano T, Sugiki T, Shimada I, Takahashi H: **Perdeuteration and methyl-selective $^{1}H$, $^{13}C$-labeling by using a *Kluyveromyces lactis* expression system**. *J Biomol NMR* 2013, **57**:297-304.

57. •• Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Mizumura T, Suzuki S, Shimada I: **Functional dynamics of deuterated β2-adrenergic receptor in lipid bilayers revealed by NMR spectroscopy**. *Angew Chem Int Ed* 2014, **53**:13376-13379.
First reported use of methyl-group labeling protocol using insect/baculovirus expression system and its application to the study of a functional GPCR protein.

58. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono AM, Guntert P: **Optimal isotope labelling for NMR protein structure determinations**. *Nature* 2006, **440**:52-57.

59. Reckel S, Gottstein D, Stehle J, Lohr F, Verhoefen MK, Takeda M, Silvers R, KainoshoM, Glaubitz C, Wachtveitl J *et al.*: **Solution NMR structure of proteorhodopsin**. *Angew Chem Int Ed* 2011, **50**:11942-11946.

60. Linser R, Gelev V, Hagn F, Arthanari H, Hyberts SG, Wagner G: **Selective methyl labeling of eukaryotic membrane proteins using cell-free expression**. *J Am Chem Soc* 2014, **136**:11308-11310.