Electronic Supplementary Material (ESI) for Chemical Communications


# An Optimized Isotopic Labelling Strategy of Isoleucine-$\gamma_2$ Methyl Groups for solution NMR Study of High Molecular Weight Proteins

**Isabel Ayala[‡a], Olivier Hamelin[‡b], Carlos Amero[a,c], Ombeline Pessey[a], Michael J. Plevin[a], Pierre Gans[d], Jérôme Boisbouvier[a,*]**

[a] *CNRS, Institut de Biologie Structurale, Grenoble, 41 rue Jules Horowitz - 38027 Grenoble Cedex – France,*
[b] *Université de Grenoble, Laboratoire de Chimie et Biologie des Métaux, 17 avenue des martyrs, 38054 Grenoble Cedex 9 – France.*
[c] *Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos. Cuernavaca. México*
[d] *CEA, Institut de Biologie Structurale, Grenoble, 41 rue Jules Horowitz - 38027 Grenoble Cedex – France.*
[‡] *These authors contributed equally to this work*
*Correspondence: jerome.boisbouvier@ibs.fr*

## *Supporting Information*

# 1. Synthesis of selectively labelled 2-hydroxy-2-ethyl-3-oxobutanoate

Ethyl 2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate

A mixture of 5.0 g (37.3 mmol) of ethyl [U-$^{13}$C]-3-oxobutanoate (**1**) (Cambridge Isotope Laboratories, Inc.), 5.7 g (41.0 mmol) of $K_2CO_3$ and 3.28 mL (41.0 mmol) [$^{2}H_5$]ethyl iodide (Cambridge Isotope Laboratories, Inc.) in 50 mL of absolute ethanol was heated at 60°C under argon for 23 h. After filtration, the filtrate was concentrated in vacuo to afford 5.15 g (82 % yield) of product, which was sufficiently pure to be used without further purification. NMR spectroscopy: $^{1}$H NMR ($CDCl_3$) 4.13 (dq, $J$ = 7.1, 3.0 Hz, 2H), 3.25 (ps dt, $J$ =130.5, 7.2, 6.9 Hz, 1H), 2.15 (ddd, $J$ = 127.8, 6.0, 1.2 Hz, 3H), 1.21 (t, $J$ = 7.1 Hz, 3H)

Ethyl 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate (**2**)

Hydroxylation reaction was carried out using a freshly prepared dimethyldioxirane solution. To a mixture of 2.0 g (12.3 mmol) of ethyl 2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate in 10 mL of distilled water, was added 305 mg (1.23 mmole) of $Ni(OAc)_2.4H_2O$ and, at 0°C, 70 mL of an untitrated solution of dimethyldioxirane (0.05-0.10 M) in acetone (Adam and Smerz, 1996). The resulting solution was allowed to warm to room temperature and stirred for 24 hours before addition of a second portion (70 mL) of dimethyldioxirane at 0°C. After additional 24h, a third portion of oxidant was added at 0°C and the resulting mixture was stirred for 24h. The organic solvent was then evaporated *in vacuo*. Water was added and the resulting aqueous residue was extracted with dichloromethane (four times). The organic extract was dried over $Na_2SO_4$ and concentrated *in vacuo* to give 2.1 g (11.75 mmol; 96 % yield) of ethyl 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate as a colourless liquid which was pure enough to be used without further purification. NMR spectroscopy: $^{1}$H NMR ($CDCl_3$) 4.30-4.16 (m, 2H), 4.10 (q, $J$ =

6.3, 3.3 Hz, 1H, OH), 2.24 (ddd, $J$ = 128.7, 6.0, 1.2 Hz, 3H), 1.26 (t, $J$ = 7.1 Hz, 3H).

2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate (3)

Deprotection of ethyl 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate (2) was achieved in $H_2O$ by addition of NaOH. Typically, 100 mg of (2) was placed in 0.5 mL of $H_2O$ and a stoichiometric quantity of NaOH (2.5M) was added step by step by aliquot of 10 μL. The deprotection was immediate as determined by NMR spectroscopy. Once the deprotection was complete, the solution was adjusted to neutral pH by addition of 100 μL of a TRIS solution (1 M - pH=8). The solution of (3) was then stored at –20°C until required.

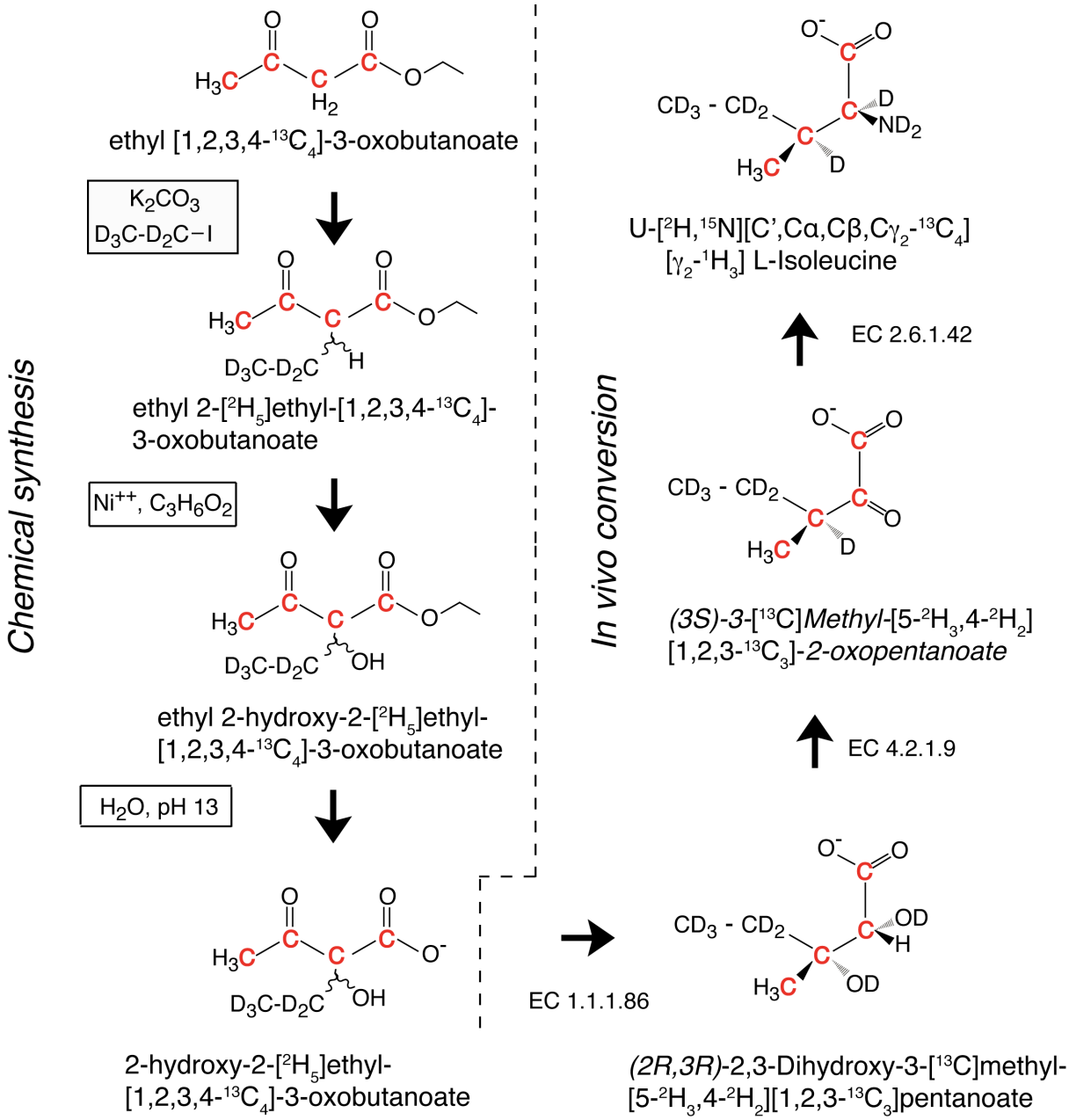


Figure S1: Reaction scheme of the protocol for the production of U-[$^{2}$H,$^{12}$C,$^{15}$N], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^{1}$H$_3$-γ2]Ile-labelled proteins. A detailed protocol for the chemical synthesis of 2-hydroxy-2-[$^{2}$H$_5$]ethyl-[1,2,3,4-$^{13}$C$_4$]-3-oxobutanoate is provided on page 2. $^{13}$C nuclei are displayed in red. The stereochemistry following the incorporation of 2-hydroxy-2-[$^{2}$H$_5$]ethyl-[1,2,3,4-$^{13}$C$_4$]-3-oxobutanoate in the different intermediates of Ile biogenesis pathway is indicated on the figure (assuming growth in D$_2$O-based M9 culture medium). Each biosynthetic intermediate has been named according to the Kyoto Encyclopedia of Genes and Genomes (KEGG). The enzymes catalyzing each reaction are indicated by EC number. EC 1.1.1.86: ketol-acid reductoisomerase; EC 4.2.1.9: dihydroxy-acid dehydratase; EC 2.6.1.42: branched-chain amino acid aminotransferase. Further information on the Ile metabolic pathway can be found online: http://www.genome.jp/kegg/.

## 2. Overexpression of methyl specifically labelled proteins in *E. coli*.

Optimization of the incorporation of 2-hydroxy-2-ethyl-3-oxobutanoate in protein.

Initial experiments to determine the level of 2-hydroxy-2-ethyl-3-oxobutanoate incorporation into overexpressed proteins were performed using ubiquitin as a model system. *E. coli* BL21(DE3) cells were transformed with a pET41c plasmid carrying the human His-tagged ubiquitin (pET41c-His-Ubi) gene and transformants were grown in M9/$H_2O$ media containing 1 g/L $^{15}ND_4Cl$, and 2 g/L of U-[$^{13}C$], glucose. When the optical density (O.D.) at 600 nm reached 0.7, a solution containing 2-hydroxy-2-ethyl-3-oxobutanoate was added. After an additional 1 h, protein expression was induced by the addition of IPTG to a final concentration of 1 mM. Induction was performed for 3 hours at 37°C. Ubiquitin was purified by Ni-NTA (Qiagen) chromatography in a single step.

The optimal quantity of 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate required to achieve near complete incorporation in the overexpressed protein was assessed in a series of cultures (50 mL each) in which different amounts of labelled precursor were added 1 hour prior induction to final concentrations of 0, 40, 80, 120, 160 and 250 mg/L. The level of incorporation into the purified protein was monitored using 2D $^{13}C$-HSQC spectra. When the labelled precursor is incorporated into the overexpressed protein, $^{12}C^{2}H_3$ isotopomer replaces the $^{13}C^{1}H_3$ group at positions $\delta_1$ of Isoleucine, and the corresponding methyl correlations disappear from spectra. Conversely, the incorporation of 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate do not change the labelling pattern of methyl at position $\gamma_2$. The quantification was performed by comparing the intensities of signals corresponding to Ile-$\delta_1$ and Ile-$\gamma_2$ methyl groups with respect to the signals of Methionine, Alanine and Threonine methyl groups. The addition of pure 2-hydroxy-2-ethyl-3-oxobutanoate, at a concentration higher than 100 mg per liter of M9 culture medium, achieves an incorporation level of 95 % in Ile side chains (Figure S2). Surprisingly, we observe a scrambling of the protonated methyl into the Leu/Val amino-acid biogenesis pathway (Figure S3), giving rise to low-level labelling of the Val/Leu proR methyl position. The mechanism(s) responsible for this undesired labelling remain(s) to be elucidated. Practically, isotope scrambling to proR methyl groups could be totally suppressed by adding a saturating concentrations (200 mg/L) of α-ketoisovalerate-$d_7$ (CDN Isotopes Inc) to the expression medium at the same time as 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate.

Production of U-[$^{2}$H,$^{12}$C,$^{15}$N], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^{1}$H$_3$-γ2]Ile-labelled proteins.

*E. coli* BL21(DE3) carrying the plasmid of the overexpressed protein (TET2 or MSG) were progressively adapted, in three stages, over 24 h, to M9/$D_2O$ media containing 1 g/L $^{15}ND_4Cl$ and 2 g/L D-glucose-$d_7$ (Isotec). In the final culture, the bacteria were grown at 37°C in M9 media prepared with 99.85% $D_2O$ (Eurisotop). When the O.D. (600 nm) reached 0.8, a solution containing labelled 2-hydroxy-2-ethyl-3-oxobutanoate (≥ 100mg / L) prepared with the protocol described above) and perdeuterated isovalerate (200 mg/L – CDN isotopes Inc) was added. 1 hour later, TET2 (MSG) expression was induced by the addition of IPTG to a final concentration of 1 mM (0.1 mM). Expression was performed for 3 hours (12 hours) at 37°C (20°C) before harvesting.

In early stages of this study we were using concentration of 2-hydroxy-2-ethyl-3-oxobutanoate as high as 300 mg/L of culture. To reduce cost of labelled materials we have used in later stages lower concentrations, and we have finally demonstrated that concentrations ≥ 100 mg/L are enough to achieve an incorporation level ≥ 95% (figure S2), provide that 2g/L of glucose are used as the source of carbon. Using higher concentrations of glucose will require higher amounts of the biosynthetic precursor to saturate the Isoleucine pathway (Plevin et al., 2011).

Proteins Purification.

Malate Synthase G (MSG) was purified initially by Chelating Sepharose chromatography (GE Healthcare) followed by gel filtration chromatography (Superdex 200pg GE Healthcare). Typical final yields after purification were 60-80 mg/L of methyl specific protonated MSG. The concentration of MSG in typical NMR samples was 1 mM in 100% $D_2O$ buffer containing 25 mM MES (pH 7.0 uncorrected), 20mM $MgCl_2$, 5mM DTT. NMR data were acquired at 37°C.

TET2 was purified using two anion exchange chromatography steps (DEAE Sepharose CL-6B, and Resource Q 6 mL, GE Healthcare) followed by gel filtration (Sephacryl S-300 HR, GE Healthcare). The typical final yield after purification was 20 mg/L of methyl specific protonated TET2. The final NMR samples of TET consisted of ~80 μM TET2 dodecamer (~1 mM monomer) in 20 mM Tris (pH 7.4 uncorrected), 20 mM NaCl dissolved in 300 μL $D_2O$. NMR data were acquired at 50°C.

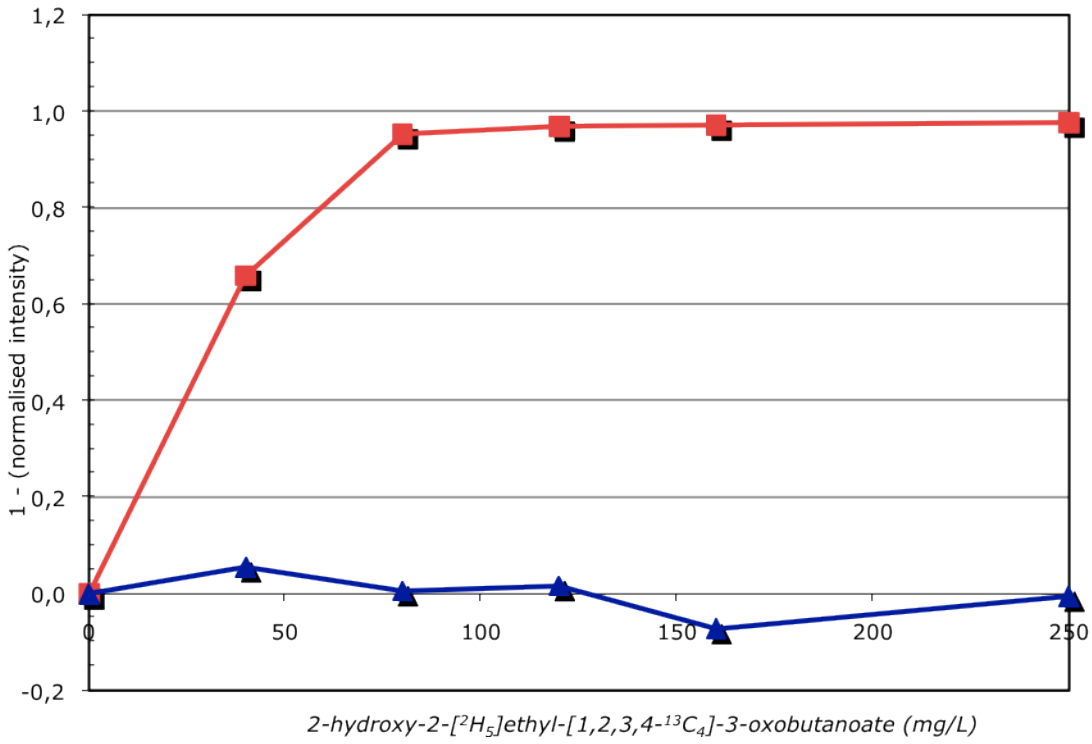


Figure S2: Level of incorporation of 2-hydroxy-2-ethyl-3-oxobutanoate in overexpressed proteins as a function of the amount of exogeneous precursor added. Ubiquitin was expressed in *E. coli* in M9/$H_2O$ culture medium with 2g/L of U-[$^{13}$C], glucose. Different amounts of 2-hydroxy-2-[$^{2}H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate were added one hour before induction. Quantification was performed by comparing the intensities of signals corresponding to Ile-$\delta_1$ (red squares) and Ile-$\gamma_2$ (blue triangles) methyl groups with respect to signals of Methionine, Alanine and Threonine. A level of incorporation in Ile side chains of 95 % is obtained by adding ≥ 100 mg of 2-hydroxy-2-ethyl-3-oxobutanoate per litter of M9/$D_2O$ culture medium using 2 g/L of glucose as the sole source of carbon.

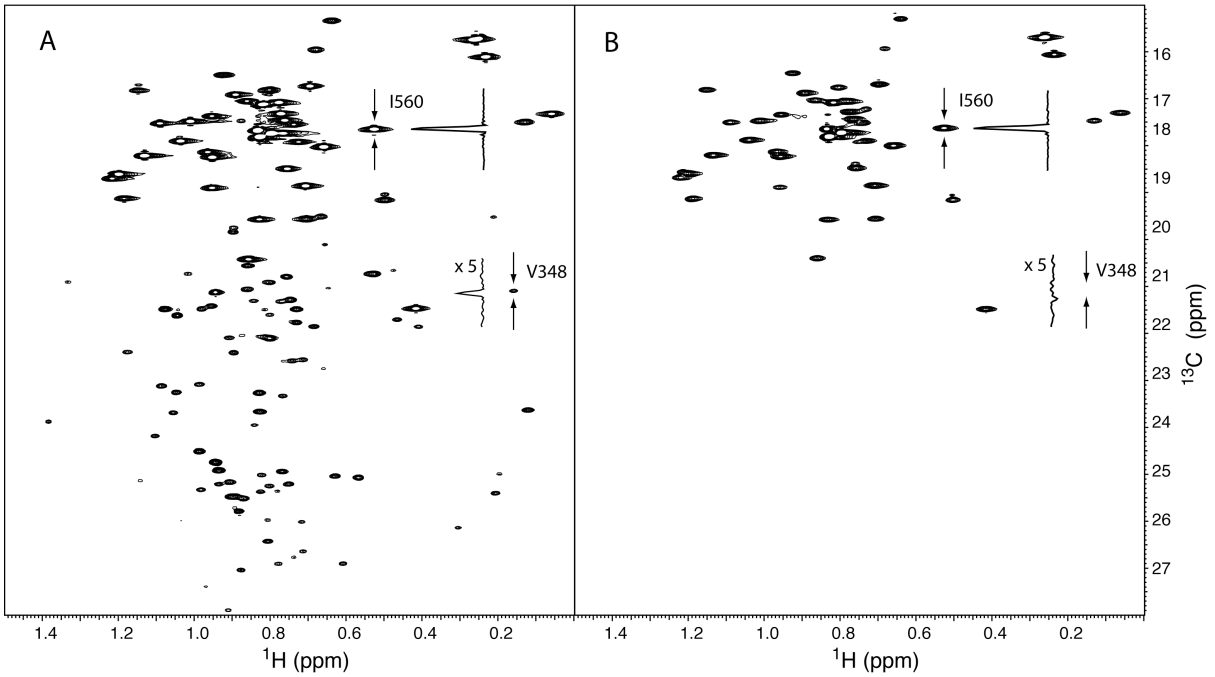


Figure S3 : Effect of α-ketoisovalerate-$d_7$ on isotopic scrambling. 2D $^{13}$C-HMQC spectra of U-[$^2$H,$^{12}$C,$^{15}$N], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^1$H$_3$-γ2]Ile-labelled MSG. MSG was produced in presence of 240 mg/L of 2-hydroxy-2-[$^2H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-oxobutanoate (prepared as described above) and either in absence (A) or in presence of 200 mg/L perdeuterated α-ketoisovalerate (B). $^{13}$C-HMQC spectra were recorded at 37°C in $D_2O$ on a NMR spectrometer operating at a proton frequency of 600 MHz. In panel A most intense signals for methyls of Val/Leu were 4-8% of those of the Ile-$\gamma_2$ methyl.

## 3. NMR Spectroscopy.

Experimental Details.

All NMR spectra of ubiquitin were recorded on a Varian DirectDrive spectrometer operating at a proton frequency of 600 MHz equipped with a cryogenic triple resonance pulsed field gradient probehead.

For MSG and PhTET2, 2D Methyl-TROSY, 3D $^{13}$C-HMQC-NOESY and 3D HCC spectra were recorded on a Varian DirectDrive spectrometer operating at a proton frequency of 800 MHz equipped with a cryogenic triple resonance pulsed field gradient probehead. The 3D HC(C)C spectra were recorded on 600 MHz spectrometer equipped with a cryogenic probehead. Each 3D experiments were recorded in 60 h using 1 mM U-[$^{2}$H,$^{12}$C,$^{15}$N], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^{1}$H$_3$-γ2]Ile-labelled sample of PhTET2 or MSG. The HCC "out and back" experiments were collected with 40 complex points in indirect $F_1$ dimension (Cβ - $t_1^{max}$ = 12 ms) and 56 complex points in indirect $F_2$ dimension (Cγ2- $t_2^{max}$ = 14 ms). The HC(C)C "out and back" experiments were collected with 80 complex points in indirect $F_1$ dimension (Cα - $t_1^{max}$ = 9 ms) and 109 complex points in indirect $F_2$ dimension (Cγ2- $t_2^{max}$ = 12 ms). The 3D HMQC-NOESY experiment was recorded with a mixing time of 300 ms, 40 complex points in indirect $F_1$ dimension ($^{1}$H - $t_1^{max}$ = 16 ms) and 36 complex points in indirect $F_2$ dimension (Cγ2- $t_2^{max}$ = 11 ms).

All data were processed and analyzed using nmrPipe/nmrDraw (Delaglio et al., 1995) and NMRView (Johnson, 2004).

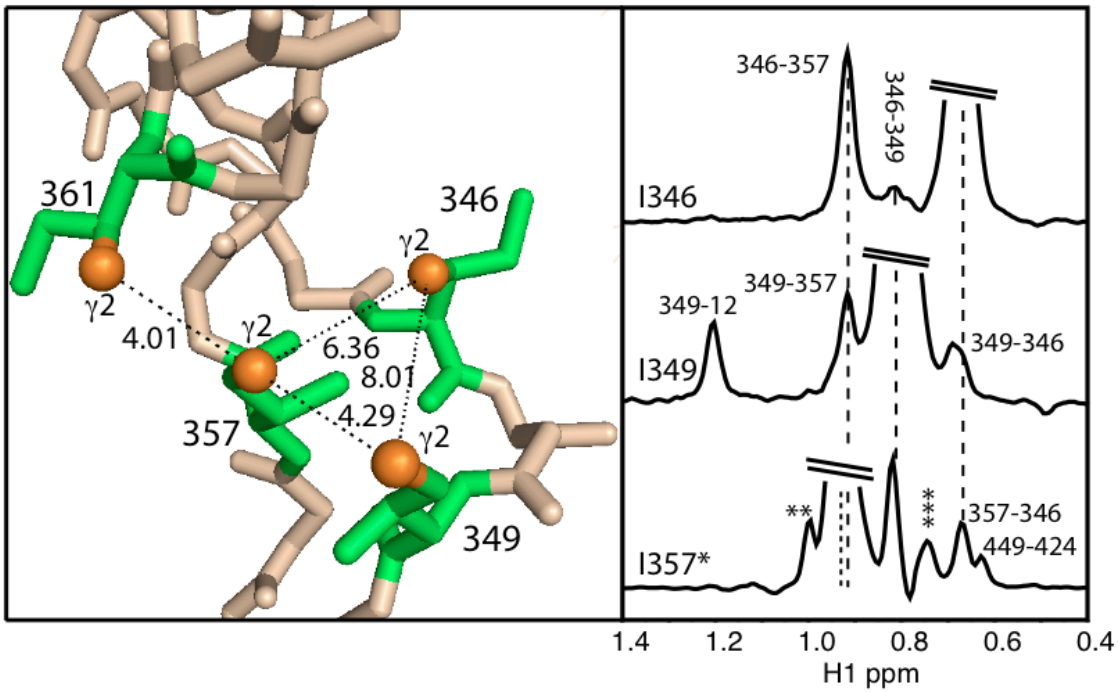


**Figure S4**: Observation of long-range NOEs between methyl-$\gamma_2$ of Ile-357 and neighbour Ile methyl groups in MSG. (a) Expansion of the region of MSG structure around Ile-357 side chain. Dashed lines indicate detected NOE (distance in Å are indicated on the side of the dash lines). Only side chains of Isoleucine 346, 349, 357 and 361 are displayed (green). The methyl-$\gamma_2$ groups are presented by orange balls. (b) 1D traces extracted from a 3D HMQC-NOESY spectra acquired, with a 1 mM U-[$^{2}$H,$^{12}$C,$^{15}$N], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^{1}$H$_3$-γ2]Ile MSG sample, at 37°C, on a NMR spectrometer operating at a proton frequency of 800 MHz. Traces have been extracted in the F3 dimension (acquisition) at the (F1,F2) frequencies of diagonal peaks of I346, I349 and I357. Signals of diagonal peaks have been truncated, and cross-peaks are labelled by their corresponding residue numbers. Shoulders of diagonal peak located in adjacent planes are indicated by asterisks (note that I357 is partially overlapped with intense peaks corresponding to I449).

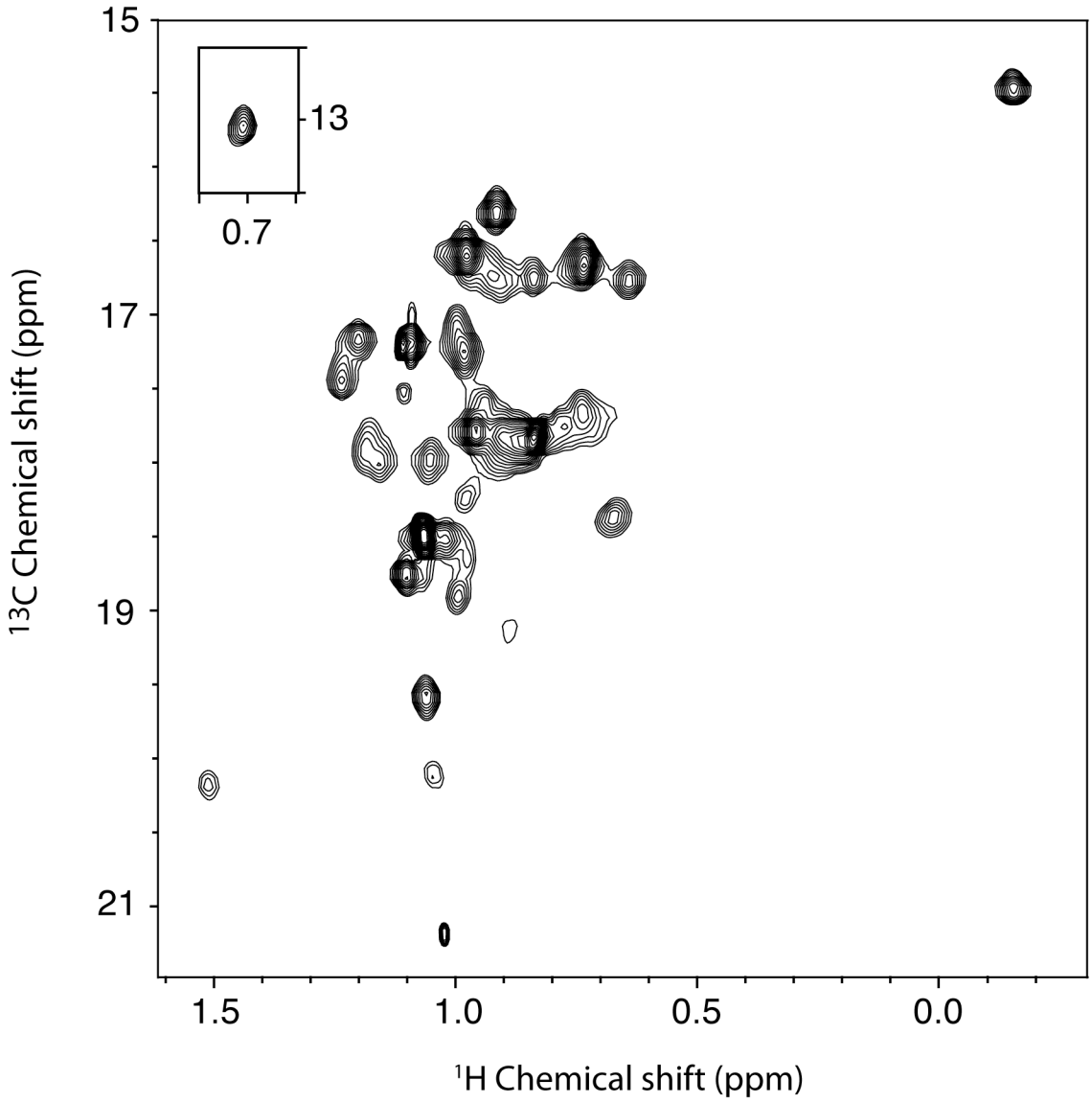


Figure S5 : Methyl-TROSY spectra acquired of a 1 mM U-[$^{15}$N,$^{12}$C,$^{2}$H], U-[$^{13}$C',$^{13}$Cα,$^{13}$Cβ,$^{13}$C$^{1}$H$_3$-γ2]Ile-labelled PhTET2 sample prepared using 2-hydroxy-2-[$^{2}$H$_5$]ethyl-[1,2,3,4-$^{13}$C$_4$]3-oxobutanoate. The spectrum was acquired at 50 °C on an 800 MHz spectrometer

Table S1 : Assignment of MSG isoleucine γ2 methyl groups.

| Residue | | Hγ2 | Cγ2 |
|---|---|---|---|
| | | | |
| I5 [a] | | 0.82 | 17.8 |
| I12 [a] | | 1.21 | 18.6 |
| I42 [a] | | 0.88 | 16.9 |
| I60 [a] | | 1.19 | 18.5 |
| I105 [a] | | 1.18 | 19.1 |
| I109 [a] | | 0.72 | 17.9 |
| I147 [a] | | 0.22 | 16.0 |
| I148 [a] | | 0.72 | 17.2 |
| I167 [a] | | 1.00 | 17.4 |
| I200 [b] | | 0.40 | 21.4 |
| I229 [a] | | 0.94 | 18.2 |
| I238 [a] | | 0.68 | 16.7 |
| I242 [a] | | 1.08 | 17.5 |
| I248 [a] | | 1.12 | 18.2 |
| I256 [c] | | 0.76 | 17.7 |
| I260 [a] | | 0.95 | 18.8 |
| I265 [a] | | 0.65 | 18.0 |
| I268 [a] | | 0.49 | 19.1 |
| I284 [a] | | 0.25 | 15.7 |
| I309 [a] | | 0.82 | 17.6 |
| I327 [c] | | 0.78 | 17.7 |

| Residue | | Hγ2 | Cγ2 |
|---|---|---|---|
| | | | |
| I349 [a] | | 0.82 | 19.5 |
| I357 [d] | | 0.75 | 17.4 |
| I361 [a] | | 0.91 | 16.4 |
| I370 [a] | | 0.85 | 17.0 |
| I388 [a] | | 0.77 | 17.0 |
| I409 [a] | | 0.70 | 18.8 |
| I424 [a] | | 0.63 | 15.3 |
| I439 [a] | | 0.75 | 18.4 |
| I449 [a] | | 0.95 | 17.3 |
| I482 [a] | | 0.76 | 17.3 |
| I504 [a] | | 0.73 | 17.5 |
| I560 [a] | | 0.51 | 17.6 |
| I579 [a] | | 1.03 | 17.8 |
| I592 [a] | | 0.96 | 18.1 |
| I603 [b] | | 0.79 | 16.7 |
| I615 [a] | | 0.81 | 17.1 |
| I623 [a] | | 0.25 | 15.7 |
| I637 [d] | | 1.13 | 16.8 |
| I642 [a] | | 0.69 | 19.5 |
| I650 [a] | | 0.05 | 17.3 |
| I697 [a] | | 0.12 | 17.4 |

[a] Regular: Residues assigned by Sheppard et al (2009) and confirmed by this study.
[b] Assignment was made on the basis of the Cβ and Cα values (BMRB number 5471).
[c] Assignments for I256 and I327 were inverted compared to Sheppard et al (2009) on the basis of the Cβ and Cα values ( BMRB number 5471; Tugarinov et al. 2003) [12].
[d] Assignment based on 3D-methyl-methyl-NOESY experiment and the three-dimensional structure of MSG (PDB number 1P7T) .

References:

W. Adam, A. K. Smerz, *Tetrahedron* 1996, *52*, 5799

F. Delaglio, S. Grzesiek , G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J Biomol NMR* 1995, *6*, 277-293.

B.A. Johnson, *Methods Mol. Biol* 2004, 278, 313.

M.J. Plevin, O. Hamelin, J. Boisbouvier & P. Gans *J. Biomol. NMR,* 2011, *49*, 61.

D. Sheppard; C. Guo; V. Tugarinov, *J. Biomol. NMR,* 2009, *43*, 229.

V. Tugarinov, D.R. Muhandiram DR, A Ayed, L.E. Kay, *J. Am. Chem. Soc.*, 2002, *124*, 10025