Supplementary Information for the manuscript:

Selective methyl labeling of eukaryotic membrane proteins using cell-free expression

By Rasmus Linser, Vladimir Gelev, Franz Hagn, Haribabu Arthanari, Sven G. Hyberts, and Gerhard Wagner

Details of cell growth, inclusion body preparation and hydrolysis, and cell-free synthesis

The OmpX gene from *E.coli* (SwissProt: P0A917) lacking its signal sequence (residues 1-24) was PCR-amplified from *E.coli* genomic DNA and inserted into a pET21a expression vector (EMD) via NdeI/BamHI restriction sites, as described previously.[1] Crude labeled amino acids were prepared by growing *E. coli* BL21 (DE3) cells containing the vector in M9 media containing 50 μg/l Carbenicilin, using 1 g/l of $^{15}NH_4Cl$ and either ^{13}C, ^{2}H-glucose or ^{12}C, ^{2}H-glucose (2 g/l) as the sole carbon source in D_2O. For amino acids with $^{1}H/^{13}C$ labeling at all ILV methyls, 125 mg α-ketoisovalerate and 75 mg α-ketobutyrate powders were added to the culture one hour prior to induction. For stereoselective methyl labeling, 300 mg per liter culture of the Leu/Val precursor ethyl 2-hydroxy 2-^{13}C-methyl 3-oxobutanoate (incubated for 30 min in 25 ml 0.1 M NaOD in D_2O and neutralized with 0.5 eq. DCl in 10 ml D_2O) was added to the growth medium at an OD_{600} of 0.4. At an OD_{600} of 0.5, 75 mg of the Ile precursor α-ketobutyrate was added in its powder form. At an OD_{600} of 0.65, the cells were induced with 1 mM IPTG and grown at 37°C. Since no stagnation of cell growth or a decrease in protein levels were observed after induction of protein expression for extended times, cells were only harvested after 20 h at 37°C to reach maximum culture densities. (The D_2O used in this work had been recycled once or twice after usage in previous bacterial cultures, incurring a deuteration level of roughly 95 and 90 %, respectively.)
Inclusion bodies were prepared by repeated (4x) extensive sonication in the presence of Triton-X100. The cell pellets were suspended in 40 ml 20 mM Tris, pH 7.5, 2 % Triton-X100 and sonicated for several minutes. The suspension was spun down at 30 000 g and the pellet resuspended in new Triton-X100 buffer. The final sonication and spin down omits Triton-X100. For storage, the final pellets were suspended in 1 ml of ddH_2O and transferred to reaction tubes, spun down at 20 000 g and weighed. Inclusion bodies are innocuous for the expressing bacteria, which can be grown to high cell densities after induction (see Supplementary Figure 2), and can be purified and hydrolyzed easily. On the order of 250 to 350 mg of inclusion bodies were obtained per liter of cell culture.

Crude amino acids were prepared mostly according to Hansen et al./Lemaster and Richards.[2,3] A 15-fold excess of methanesulfonic acid was added to the inclusion bodies as well as imidazole (20% of the inclusion body mass) as a suicide base, the reaction vessel was sealed and the reaction was stirred at 115°C for 3 d.

The reaction mixture was diluted 25x with water and filtered to remove insoluble particles and then poured into a column filled with Dowex AG50W-X8 resin (>10x the inclusion body wet weight). The column was washed with 10 CV water and eluted with 20 CV 100mM piperidine in water. The amino acid-containing fractions were evaporated to dryness on a rotary evaporator and dissolved in water, the pH carefully adjusted to 9.5 using $NaOH_{aq.}$ and evaporated to dryness repeatedly. When the pH remained unchanged (after 2-3 times), the solution was neutralized and passed through a 500-Da filter

before lyophilization. This filtration slightly increased cell-free-expression yields as judged by SDS page. Final yields amounted to roughly 200 mg/l culture of the crude ILV amino acids for cell-free protein expression.

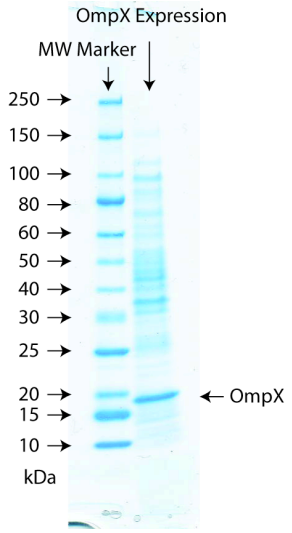
Cell-free protein expression was done according to the Schwarz et al.[4] Trp, Asn, Cys, and Gln, which are known to be most vulnerable to acid hydrolysis at high temperatures,[5] were added in an unlabeled form. In the first cell-free reaction of this kind, the crude ILV amino acids were supplemented with 5% deuterated, ^{15}N commercial algal amino acids (Sigma-Aldrich) in addition. This was done for reasons of prudence in order to ensure a minimal amount of all amino acids also in case that single amino acids other than W, N, C, and Q had not survived the process well enough to provide the critical amounts needed. In later preparations, this addition was omitted without obvious loss of protein yield.

Sequence:

```
         10         20         30         40         50         60
TSTVTGGYAQ SDAQGQMNKM GGFNLKYRYE EDNSPLGVIG SFTYTEKSRT ASSGDYNKNQ
         70         80         90        100        110        120
YYGITAGPAY RINDWASIYG VVGVGYGKFQ TTEYPTYKHD TSDYGFSYGA GLQFNPMENV
        130        140
ALDFSYEQSR IRSVDVGTWI  AGVGYRF
```

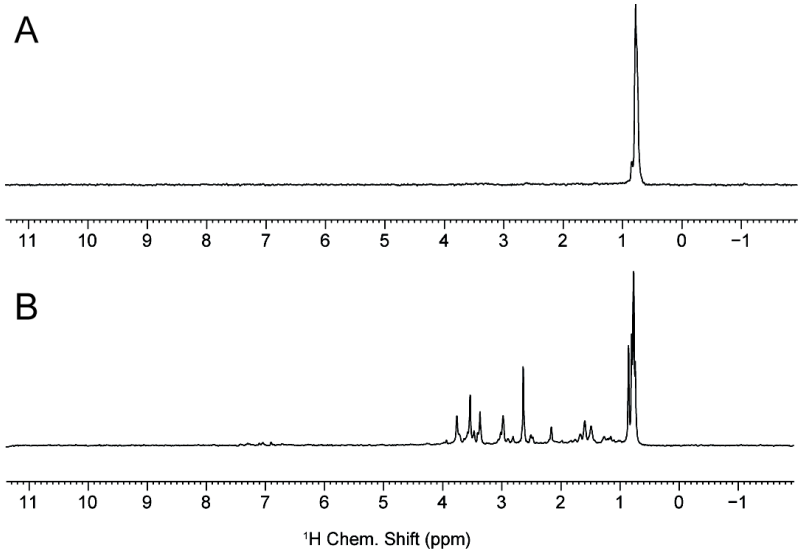
Total amino acids:	***Composition:***		
147	Ala 9 (6.1 %)	Glu 6 (4.1 %)	Phe 7 (4.8 %)
Molecular weight:	Arg 6 (4.1 %)	Gly 20 (13.6%)	Pro 4 (2.7 %)
16311.8 kDa	Asn 8 (5.4 %)	His 1 (0.7 %)	Ser 13 (8.8 %)
	Asp 8 (5.4 %)	Ile 6 (4.1 %)	Thr 12 (8.2 %)
	Cys 0 (0.0 %)	Leu 4 (2.7 %)	Trp 2 (1.4 %)
	Gln 7 (4.8 %)	Lys 6 (4.1 %)	Tyr 16 (10.9%)
		Met 3 (2.0 %)	Val 9 (6.1 %)

Supplementary Figure 1: Amino-acid sequence and composition of OmpX – a native *E. coli* membrane protein. The amino-acids Cys, Trp, Gln, and Asn get degraded upon acid hydrolysis and need to be supplemented for the cell-free reaction.[5]

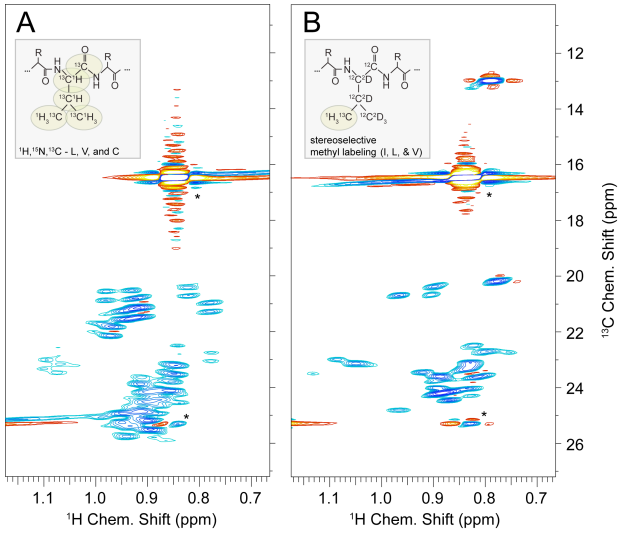


Supplementary Figure 2: Expression of OmpX in *E. coli* BL21 (DE3) cells. The right lane shows crude cell mass containing unlabeled OmpX in LB medium after 12h at 37C. Cells were induced at an OD_{600} of 0.8, reaching an OD_{600} of 5.7 after 12h and mainly consisting of OmpX inclusion bodies. For

deuterated M9 media, on the order of 250 to 350 mg of inclusion bodies were obtained per liter of culture.

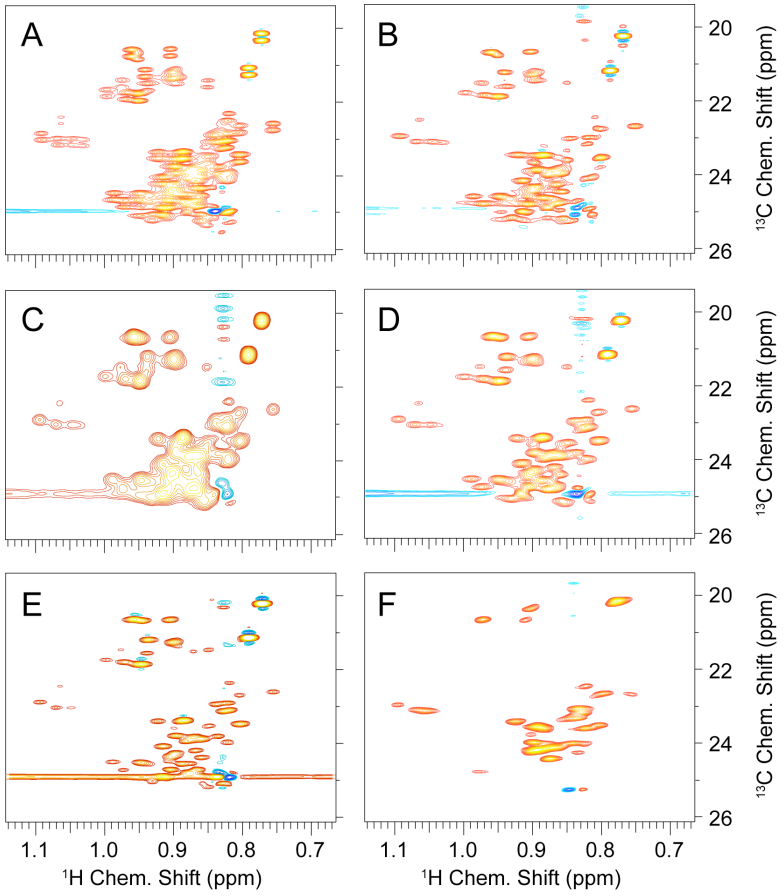


Supplementary Figure 3: ^{13}C-filtered ^{1}H spectra of crude ILV amino acids. A) Amino acids with ethyl 2-hydroxy 2-^{13}C-methyl 3-oxobutanoate and α-ketobutyrate precursors and u-^{12}C, ^{2}H glucose. B) Amino acids produced from α-ketoisovalerate and α-ketobutyrate precursors and u-^{13}C, ^{2}H glucose. (An expected residual protonation of non-methyl sites in all 20 amino acids[6] was observed as a result of usage of ~ 90% D_2O for *E.coli* culture media).

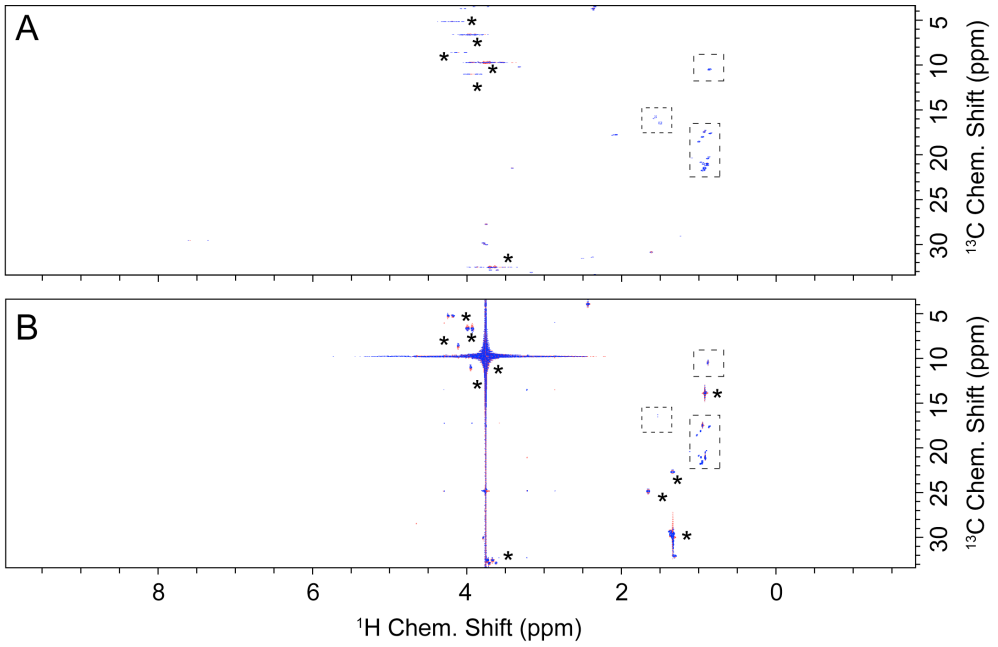


Supplementary Figure 4: Comparison of A) methyl labeling using combinatorial-labeling strategies (u-^{1}H,^{13}C,^{15}N-labeled L, V, and C in otherwise u-^{2}H,^{12}C,^{15}N-labeled protein) vs. B) labeling using stereo-selectively ILV-methyl-^{1}H/^{13}C (and otherwise u-^{2}H,^{12}C,^{15}N-labeled) amino acids. Even though fully protonated I, L, and V amino acids may be used as a route towards methyl correlations, the marked reduction in spectral overlap gained from fewer correlations with stereoselective labeling cannot be

obtained by combinatorial labeling using commercially available algal amino acids at this time. These spectra were recorded at 500 MHz ^{1}H Larmor frequency. The strong peak at 0.9/16.5 ppm and the artifacts at 25 ppm ^{13}C chemical shift result from unlabeled LMPG detergent (1-myristoyl-2-hydroxy-*sn*-glycero-3-phospho-(1'-*rac*-glycerol)).



Supplementary Figure 5: Comparison of obtainable spectral resolution of a fully ^{13}C, non-stereoselectively ^{1}H-methyl labeled SREBP membrane anchor (A – E) with the one obtained with stereoselective methyl labeling (F). Resolution lost due to doublet patterns can in principle be recovered by constant-time experiments (even though less sensitive) or doublet deconvolution. Spectral simplification due to stereospecific labeling, however, cannot be achieved without this labeling. A) Standard HSQC, B) the same spectrum with doublet deconvolution[7,8], C, D, and E represent constant-time HSQCs recorded using 13.3, 26.6, and 39.9 ms constant-time periods. F) Standard HSQC recorded on a stereoselectively methyl-labeled sample. The primary sequence of the two samples is identical apart from 5 deviating residues constituting a different cleavage site between the anchor and an uncleaved N-terminal His tag. All spectra were recorded at 800 MHz ^{1}H Larmor frequency, acquisition times for spectra A, C, D, E are 4.5 h, 45 min, 3 h, and 9.5 h, respectively, on the same sample. F was recorded in 1.5 h on a sample with only a quarter the concentration. Only the Val/Leu region is shown here due to the presence of only one Ile. The artifacts at 25 ppm ^{13}C chemical shift result from unlabeled LMPG detergent.



Supplementary Figure 6: Detergent artifacts from 1-*myristoyl*-2-hydroxy-sn-glycero-3-phospho-(1'-rac-glycerol) (LMPG), comparing deuterated (A) and protonated (B) detergent. Detergent signals are marked by asterisks, the protein methyl resonances of interest (stereo-ILV-labeled S1P-cleaved SREBP membrane anchor) are boxed by dashed lines. The indirect spectral width of 30 ppm leads to folding of peaks above 33 ppm in the ^{13}C dimension.

References Supporting Information:

(1) Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G. *J. Am. Chem. Soc.* **2013**, *135*, 1919.

(2) LeMaster, D. M.; Richards, F. M. *Anal. Biochem.* **1982**, *122*, 238.

(3) Hansen, A. P.; Petros, A. M.; Mazar, A. P.; Pederson, T. M.; Rueter, A.; Fesik, S. W. *Biochemistry* **1992**, *31*, 12713.

(4) Schwarz, D.; Junge, F.; Durst, F.; Frölich, N.; Schneider, B.; Reckel, S.; Sobhanifar, S.; Dötsch, V.; Bernhard, F. *Nat. Protocols* **2007**, *2*, 2945.

(5) Pickering, M. V.; Newton, P. *LC-GC* **1992**, *8*, 778.

(6) Asami, S.; Schmieder, P.; Reif, B. *J. Am. Chem. Soc.* **2010**, *132*, 15133.

(7) Shimba, N.; Kovacs, H.; Stern, A. S.; Nomura, A. M.; Shimada, I.; Hoch, J. C.; Craik, C. S.; Dötsch, V. *J. Biomol. NMR* **2004**, *30*, 175.

(8) Hyberts, S. G.; Milbradt, A. G.; Wagner, A. B.; Arthanari, H.; Wagner, G. *J. Biomol. NMR* **2012**, *52*, 315.