Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy

Originally published in:
Journal of Biomolecular NMR 71(3), https://doi.org/10.1007/s10858-017-0156-z

Funding acknowledgement:
159707 - NMR studies in the Solid State (SNF)

Journal of Biomolecular NMR (2018) 71:141–150
https://doi.org/10.1007/s10858-017-0156-z

ARTICLE

Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy

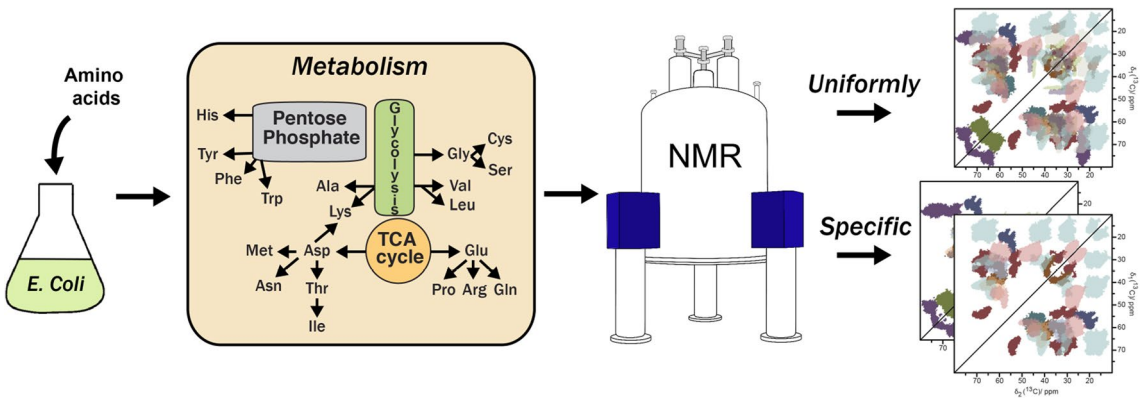
Denis Lacabanne[1] · **Beat H. Meier**[2] · **Anja Böckmann**[1]

Received: 28 July 2017 / Accepted: 23 November 2017 / Published online: 2 December 2017

Abstract

Selective isotope labeling is central in NMR experiments and often allows to push the limits on the systems investigated. It has the advantage to supply additional resolution by diminishing the number of signals in the spectra. This is particularly interesting when dealing with the large protein systems which are currently becoming accessible to solid-state NMR studies. Isotope labeled proteins for NMR experiments are most often expressed in *E. coli* systems, where bacteria are grown in minimal media supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as sole source of nitrogen and carbon. For amino acids selective labeling or unlabeling, specific amino acids are supplemented in the minimal medium. The aim is that they will be incorporated in the protein by the bacteria. However, *E. coli* amino-acid anabolism and catabolism tend to interconnect different pathways, remnant of a subway system. These connections lead to inter conversion between amino acids, called scrambling. A thorough understanding of the involved pathways is thus important to obtain the desired labeling schemes, as not all combinations of amino acids are adapted. We present here a detailed overview of amino-acid metabolism in this context. Each amino-acid pathway is described in order to define accessible combinations for ^{13}C or ^{15}N specific labeling or unlabeling. Using as example the ABC transporter BmrA, a membrane protein of 600 residues, we demonstrate how these strategies can be applied. Indeed, even though there is no size limit in solid-state NMR, large (membrane) proteins are still a challenge due to heavy signal overlap. To initiate resonance assignment in these large systems, we describe how selectively labeled samples can be obtained with the addition of labeled or unlabeled amino acids in the medium. The reduced spectral overlap enabled us to identify typical spectral fingerprints and to initiate sequential assignment using the more sensitive 2D DARR experiments with long mixing time showing inter-residue correlations.

Graphical Abstract



Keywords NMR · Amino-acid selective isotope labeling · Reverse labeling · *E. coli* metabolism · NMR assignments

Extended author information available on the last page of the article

Introduction

Selective labeling strategies have been crucial in NMR studies. With the advent of higher magnetic fields, solution-state NMR studies have concentrated on methyl-group selective labeling in deuterated proteins, which makes it possible to push the size limit of proteins to be investigated in order to observe only the most favorably relaxing component in TROSY spectroscopy (Pervushin et al. 1997). In solid-state NMR, relaxation is less an issue than spectral overlap, and notably glycerol-2 and 1–3 labeling schemes have been introduced to reduce spectral overlap (Higman et al. 2009) by reducing the number of NMR-active spins in the sample. These approaches have been shown to be precious for measuring distances used in structure calculations (Castellani et al. 2002; Tuttle et al. 2016).

Another way of reducing the number of NMR-visible residues and thus spectral overlap observed in the spectra is amino-acid selective labeling (Nuzzio et al. 2016). In combination with high-field NMR, it allows size limitations to be pushed forward and opens the way to the study of large and complex proteins, for example the ABC transporter discussed here. The success of amino acid selective labeling of proteins expressed in bacteria is intimately linked to the detailed knowledge of the amino-acid metabolism of the bacteria (Fig. 1). Before going into a detailed description of the individual pathways for each amino acid, it is important to point out that specific labeling or unlabeling (also called reverse labeling) needs to be done in minimal medium (Studier 2005), with ^{13}C labeled precursors, for example glycerol or glucose, and ammonium chloride as sole source of carbon and nitrogen. Labeled or unlabeled amino acids can then be added in the medium.

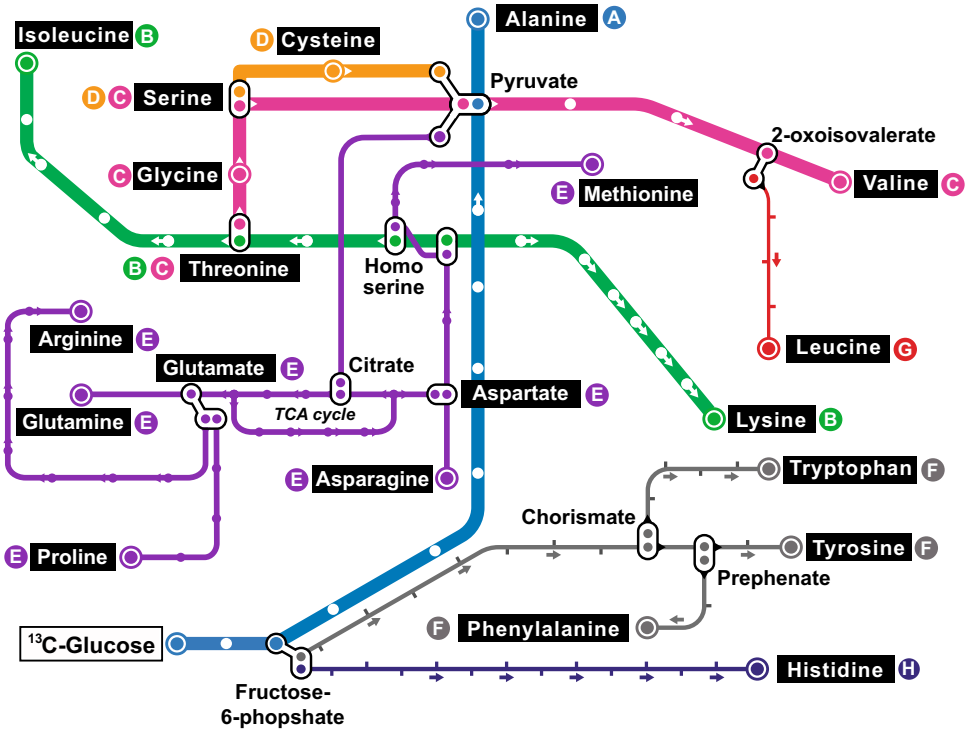


Fig. 1 General overview of amino-acid biosynthesis from *E. coli* metabolic pathways. The simplified metabolic pathways shown refer only to the amino-acid biosynthesis in *E. coli*. The overview was established considering the known *E. coli* metabolism, as for example reported in the databases KEGG (Kanehisa et al. 2017) and Ecocyc (Keseler et al. 2016) and represents a condensed version of the information found at http://www.genome.jp/kegg/ and https://ecocyc.org. Line A (blue line), for example, represents the glycolysis which leads to the formation of pyruvate, which is a precursor of alanine. Line B (green line) shows the catabolism of aspartate resulting in lysine, threonine and isoleucine. Line C (pink line) displays the catabolism of threonine which products are converted into glycine and serine. Serine can in turn be catabolized to yield cysteine, line D (orange line), and used, via the pyruvate station, for the biosynthesis of valine and alanine. Line E (purple line) shows the products formed after the TCA cycle: aspartate and glutamate, which can be catabolized leading to arginine, proline and glutamine. Line F (gray line) represents the aromatic biosynthesis pathway from fructose-6-phosphate. Line G (red line) shows the anabolism of leucine from a common precursor with valine, 2-oxoisovalerate. Finally, line H (dark blue) represents the anabolism of histidine from fructose-6-phosphate. Each metabolic step before the final amino acid is represented by a station, non-reversible reactions are indicated by arrows. Interconnections between lines are represented and the names of the interconnecting compounds are indicated. Amino acids are highlighted against black background

In the metabolic pathways, some amino acids are end products of their pathway, enabling selective labeling, and others are located on the same track, and cannot be labeled independently. Figure 1 gives an overview on the different synthesis (anabolic) and degradation (catabolic) pathways important in *E. coli* amino-acid metabolism, represented in the style of interconnecting subway lines. For example, Line A represents the glycolysis from glucose to two central metabolic products, fructose-6-phosphate, and pyruvate. From pyruvate, alanine is produced directly, and valine/leucine from a common precursor, 2-oxoisovalerate. A full description of the transport system can be found in the Fig. 1 caption, as well as in the following subsections which describe how every amino acids, or groups thereof, can efficiently be labeled in the context of these metabolic connections.

Isoleucine, leucine and valine

The catabolic pathways connecting the amino acids with aliphatic side chains are absent in *E. coli*, except for the alpha amino group (Kazakov et al. 2009). Consequently, the three branched amino acids isoleucine, leucine, and valine (respectively represented at the end of the line B, G and C in Fig. 1) can be selectively ^{13}C labeling without scrambling (Wiegand et al. 2017). However, this is not true for nitrogen labeling due to the action of transaminases. Indeed, the branched-chain-amino-acid transaminase catalyzes the reversible reaction of transfer of the alpha amino group between the branched amino acid and α-ketoglutarate to generate glutamate and the precursors of the branched amino acids (Rudman and Meister 1953; Muchmore et al. 1989; Shortle 1994; Waugh 1996). Moreover, the valine-pyruvate transaminase, which catalyzes a reversible transamination between L-alanine and valine, generates significant scrambling on alanine. Consequently, a supplementation of 100 mg/L (Shortle 1994) or 1 g/L (Bellstedt et al. 2013) of one of these amino acids in the minimal medium induces significant scrambling on the alpha amino groups of the other branched amino acids, as well as on alanine (strongly when valine is supplemented).

Isoleucine, valine and leucine are all targets for methyl labeling. Designing methyl labeling protocols in perdeuterated proteins requires thorough understanding of the *E. coli* metabolism. Actually, metabolic precursors can be used to incorporate $^{13}CH_3$ into branched amino acids in order to achieve minimum scrambling. Their type, necessary quantity and potential incompatibilities have recently been reviewed (Kerfah et al. 2015).

Alanine

Alanine is the end product of the line A that runs along the way of glycolysis (Fig. 1). In order to label alanine, the addition of 600 mg/L to the minimal medium enables a sufficient incorporation for both ^{15}N and ^{13}C as described in the literature (Muchmore et al. 1989; Ayala et al. 2009). However, selective ^{13}C labeling or ^{12}C reverse labeling remains difficult, due to the presence of alanine transaminases that transfer an amine group from glutamate (as amine donor) to pyruvate (produced from glucose, Line A), thereby generating ^{13}C alanine. Indeed, one would have to add large amounts of ^{12}C alanine in the medium to unlabel alanine, due to the high concentration of ^{13}C-pyruvate (a metabolic product of ^{13}C-glucose). Thus, this enzyme causes the presence of undesired ^{13}C signals from alanine in the case of incorporation of ^{12}C-Alanine (0.5 g/L) into the culture medium (Kaur et al. 2015). As the transaminase can also remove the amine group from alanine (resulting in the transfer of amine group to α-ketoglutate/2-oxoglutarate generating glutamate) to generate pyruvate, large amounts of ^{12}C alanine will result in large amounts of ^{12}C-pyruvate, causing signal loss by dilution, since pyruvate is at the origin of several amino-acids pathways. It is worth noting though that the addition of 1-[^{13}C]-alanine (labeled carbonyl) allows to obtain clean carbonyl labeling as shown by Takeuchi et al. (2007).

The specific labeling of 3-[^{13}C]-alanine, 2-3-[$^{13}C_2$]-alanine, and U-[^{13}C]-alanine is made possible by adding specific precursors (Ayala et al. 2009). The addition of the precursors is to prevent scrambling of isotopically labelled sites in the alanine that is also added.For this purpose, the addition of succinate (2.5 g/L), isoleucine (60 mg/L) and α-ketoisovalerate (200 mg/L) allows for a reduction of scrambling to a level lower than 5%. This protocol was developed in order to obtain a complete incorporation of methyl-protonated alanine into a perdeuterated protein using a supplement of 2-[^{2}H]-3-[^{13}C]-Alanine in combination with the precursors succinate-d_4, isoleucine-d_{10} and α-ketoisovalerate-d_7.

As already described for the branched amino acids, alanine ^{15}N specific labeling or reverse labeling is possible without scrambling; still, excessive amounts of alanine must be avoided ((1 g/L) (Bellstedt et al. 2013)), since they could induce scrambling with valines. Consequently, the intensity of the valine correlation peaks can only be decreased by 50% when the medium is supplemented with ^{14}N-alanine (1 g/L) (Bellstedt et al. 2013).

Arginine, lysine and histidine

These three amino acids are end products of their respective metabolic pathways represented respectively on the lines E, B and H (Fig. 1). Moreover, as they are end products without

possible reverse reactions, the specific or reverse labeling of these residues is simple, and scrambling is of no concern when they are added together or separately to the culture medium. Hiroaki and co-workers have shown that a supplementation of 100 mg/L of arginine or lysine is enough for a correct incorporation of these residues (95 and 82% respectively) in the case of nitrogen (un)labeling (Hiroaki et al. 2011). Bellstedt and co-workers have shown that there is no scrambling both in carbon and nitrogen atoms for these three residues even when they are added in a large excess (1 g/L) to the medium (Bellstedt et al. 2013). However, it was shown that arginine supplementation of 175 mg/L leads to about 50% of ^{13}C unlabeling (Shi et al. 2009a, b). Finally, a ^{12}C-arginine or ^{12}C-lysine supplementation of 400 mg/L is sufficient to remove the ^{13}C resonances (Wiegand et al. 2017). Arginine or lysine supplementation of respectively 380 or 160 mg/L is sufficient for ^{15}N/^{14}N (un)labeling (Muchmore et al. 1989; Nishida et al. 2006).

Histidine has the advantage that bacteria do not degrade it. Consequently, adding histidine to the medium causes no scrambling of ^{15}N and ^{13}C. A quantity of 400 mg/L (Wiegand et al. 2017) or 100 mg/L (Nishida et al. 2006) is sufficient for reverse labeling and selective labeling (1-^{13}C-His) respectively.

Aspartate/asparagine and glutamate/glutamine

These four amino acids are located together on line E (Fig. 1). Aspartate and glutamate are central players in the bacterial metabolism. Aspartate is the precursor of asparagine, isoleucine, methionine, lysine and threonine, while glutamate is the precursor of arginine, glutamine and proline. Glutamate is also the major nitrogen donor in protein bio-synthesis (Goux et al. 1995). As a consequence, the extensive interconnections and proximal locations of aspartate, glutamate, asparagine and glutamine on line E in Fig. 1 result in the fact that these residues remain a great challenge for a clean labeling strategy.

Still, Tate et al. have reported a strategy for selective labeling of the side-chain NH groups for asparagine and glutamine simultaneously (Tate et al. 1992) and Cao et al. have recently shown a strategy that is able to label selectively the side-chain NH group of asparagine residues, and also the side-chain NH groups of glutamine, asparagine and tryptophan simultaneously (Cao et al. 2014). We briefly describe these strategies in the following. The selective labeling of side-chain NH group of asparagine has to avoid at least three metabolic effects: (i) the transfer reaction of the side chain from glutamine to asparagine, and the reverse-reaction catalyzed by asparagine synthetase B; (ii) the degradation of glutamine in glutamate and ammonium catalyzed by glutaminase; and (iii) the incorporation of ammonium into asparagine catalyzed by aspartate-ammonia ligase when ammonium is abundant. Cao and co-workers obtained, with ubiquitin, a ^{15}N-labeling efficiency of the asparagine side-chain of 98% without significant scrambling by supplementing the minimal medium with 0.55 g/L $^{15}NH_4Cl$, 2 g/L ^{14}N-glutamate, 4 g/L ^{14}N-glutamine, 4 g/L ^{14}N-aspartate and all other ^{14}N-amino acids (except asparagine) 0.5 g/L.

Moreover, they equally obtained, with ubiquitin, a ^{15}N-labeling efficiency of the asparagine and glutamine side-chain of respectively 84 and 85% without significant scrambling. The approach consists in supplementing the minimal medium with 0.27 g/L $^{15}NH_4Cl$, 2 g/L ^{14}N-glutamate, 4 g/L ^{14}N-asparagine and all other ^{14}N-amino acids (except aspartate and glutamine) 0.5 g/L. These authors have noticed that, surprisingly, despite the addition of ^{14}N-asparagine in large excess, its ^{15}N-labeling remains quite efficient. They have suggested that the ^{15}N-labeling yield of the side-chain depends mainly on the intracellular concentration of aspartate and $^{15}NH_4Cl$.

Methionine

Methionine is located at the end of its metabolic pathway, it is the last station on line E (Fig. 1). Its catabolism by *E. coli* seems to be possible for nitrogen but this catabolic pathway remains unclear (Reitzer 2005). However, no scrambling was observed in ^{13}C and in ^{15}N when 1 g/L of labeled methionine was added, and when natural abundance methionine was added to the medium, the signal of methionine was 70% attenuated (Bellstedt et al. 2013). The methyl group of methionine can be used as an NMR probe in methyl labeling (Plevin and Boisbouvier 2012; Stoffregen et al. 2012) by adding 250 mg/L of [ε-^{13}C]-labeled methionine in the culture medium (Stoffregen et al. 2012; Gelis et al. 2007).

Proline

Proline is represented at the end of the line E (Fig. 1). Proline can be used as sole source of nitrogen and carbon for *E. coli* (Reitzer 2005). It can be catabolized to yield glutamate by the combination of two enzymes: proline dehydrogenase and 1-pyrroline-5-carboxylate dehydrogenase (Zhou et al. 2008). Consequently, the addition of natural-abundance proline in the culture medium induces general carbon scrambling, with amino acids located along line E (Gln, Glu and Arg) the most severely affected, especially when proline concentration is 250 mg/L (Rasia et al. 2012) or 1 g/L (Bellstedt et al. 2013). This general scrambling impedes a labeled-proline addition for specific labeling due to the apparitions of resonances from Gln, Glu and Arg. On

the other hand, the addition of natural abundance proline can be used for reverse labeling. Indeed, all resonances from proline will disappear, and the scrambling will induce a small signal loss of the other resonances, but will still allow 2D spectra recording (Wiegand et al. 2017; Kaur et al. 2015).

Glycine, serine, cysteine and threonine

Glycine and serine, neighboring stations on line C (Fig. 1), are linked by glycine-hydroxymethyltransferase-mediated interconversion. Threonine is represented at the end of line C (Fig. 1). Indeed, glycine and threonine are also linked by interconversion, which is, in this case, mediated by a threonine aldolase. Serine is the substrate of a deamination reaction, catalyzed by a serine dehydratase which converts the amino acid into pyruvate. Besides, this amino acid is involved in cysteine biosynthesis, as well as in tryptophan biosynthesis, where a tryptophan synthase combines a serine and an indole to form tryptophan.

In other words, (i) the addition of glycine induces scrambling with cysteine, serine and tryptophan for ^{15}N and ^{13}C; (ii) serine dilutes the labeling of all other amino acids; and (iii) threonine scrambles with cysteine, glycine, serine, tryptophan ^{15}N and ^{13}C, but only ^{13}C for isoleucine. Indeed, threonine is a precursor of the isoleucine biosynthesis pathway. Too complex to be fully represented in Fig. 1, the links between these residues are presented in detail in Fig. 2.

The addition of glycine and 2-ketobutyric acid (an isoleucine precursor), both in natural abundance, ensures an incorporation of labeled threonine without scrambling. Indeed, as labeled threonine is diverted to isoleucine and glycine

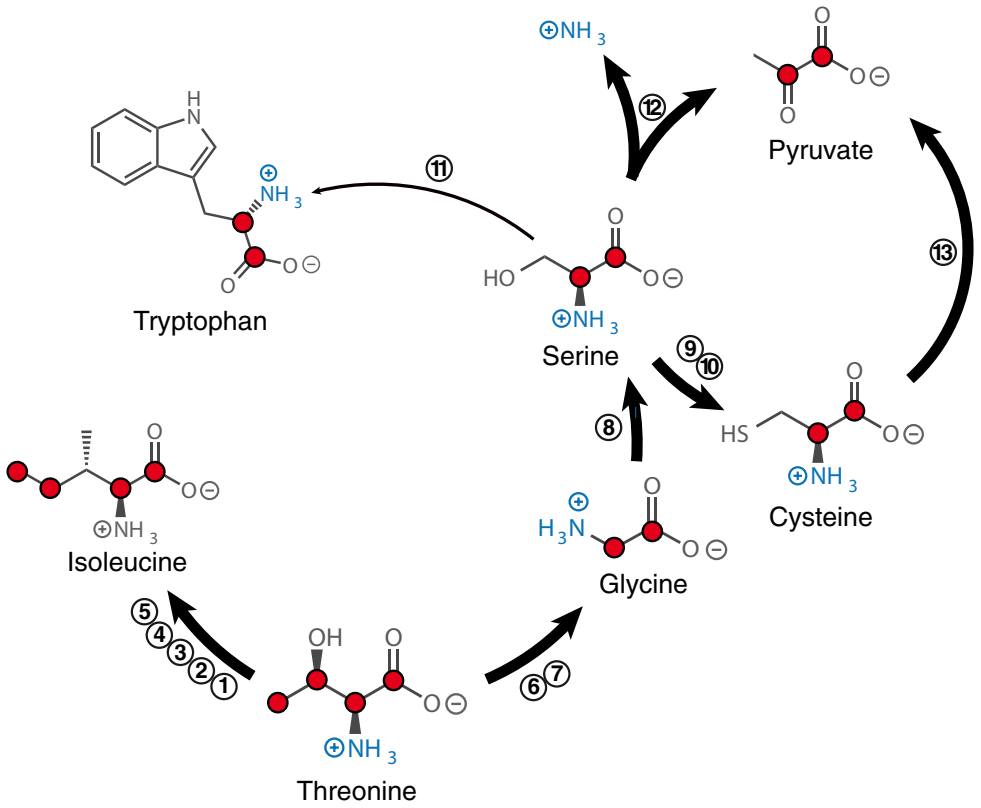


Fig. 2 Metabolic connections of threonine. Carbons from threonine are represented by red dots, whereas amines are colored in blue. Threonine is the precursor of isoleucine in the biosynthesis pathway. Threonine deaminase (1) leads to the formation of 2-oxobutanoate and ammonia, then acetohydroxy acid synthase (2) catalyzes a reaction of decarboxylation of pyruvate and converts 2-oxobutanoate into 2-aceto-2-hydroxybutanoate. This product is reduced by an oxidoreductase (3) and dehydrated by an acid dehydratase (4) forming the next-to-last compound, 3-methyl-2-oxopentanoate. Finally, isoleucine is formed by a reaction of transamination (5). The catabolism of threonine to yield glycine involves the threonine dehydrogenase (6) which catalyzes the reaction of oxidation of threonine into 2-amino-3-oxobutanoate, and the 2-amino-3-ketobutyrate CoA ligase (7) which catalyzes the reaction of condensation of 2-amino-3-ketobutyrate and coenzyme-A to form acetyl-CoA and glycine. Serine is a product of glycine catabolism; the reaction is catalyzed by a serine hydroxymethyltransferase (8). Serine catabolism forms pyruvate and ammonium via the serine deaminase (12), and serine is associated with the last stage of tryptophan anabolism catalyzed by a tryptophan synthase (11). The residue is also used for cysteine biosynthesis mediated by serine acetyltransferase (9) and cysteine synthase (10). Cysteine is degraded into pyruvate, ammonia and sulfide, this reaction is catalyzed by the action of cysteine desulfhydrases (13)

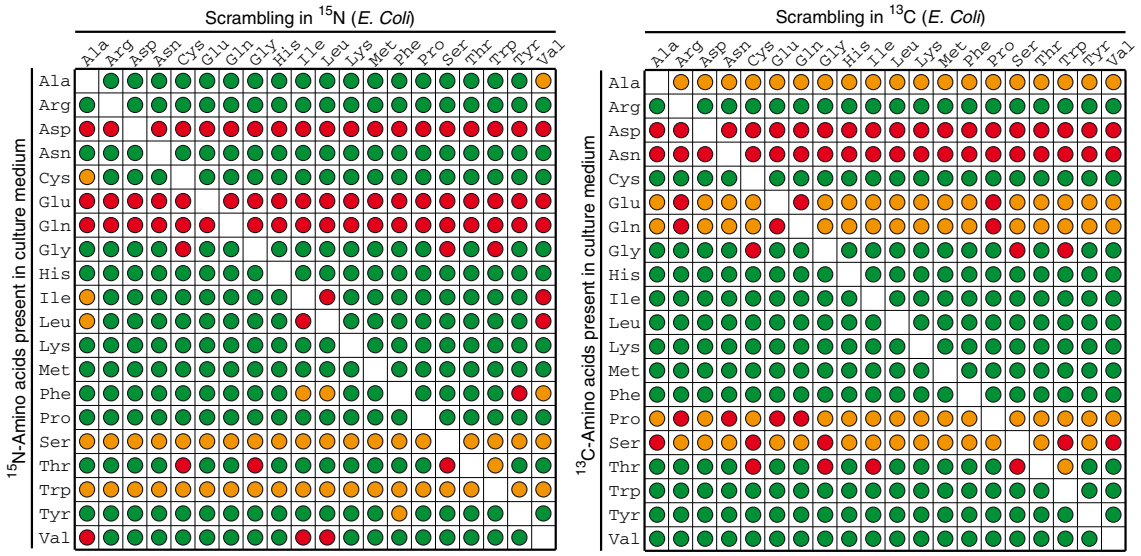


Fig. 3 Influence of an added amino acid on the (un)labeling of others with respect to ^{15}N and ^{13}C labeling. Combinations without scrambling are indicated with green dots; weak scrambling, i.e. suitable for reverse labeling and specific labeling, is indicated with orange dots. Substantial scrambling is indicated with red dots

(Fig. 2), the addition of glycine and 2-ketobutyric acid in natural abundance drastically reduces the scrambling effect from threonine. This solution has been presented for methyl labeling of threonine (Thr-γ2[$^{13}CH_3$]) and isoleucine (Ile-d_1[$^{13}CH_3$]) (Velyvis et al. 2012).

Phenylalanine, tyrosine and tryptophan

These aromatic residues use the end product of the pentose-phosphate pathway, erythrose-4-phosphate, a precursor of all aromatic residues, and which allows the fructose-6-phosphate to enter line F (Fig. 1). There is no degradation pathway for phenylalanine and tyrosine in *E. coli*. Regarding tryptophan, *E. coli* can utilize tryptophan as the sole source of carbon. Tryptophan is degraded into indole, pyruvate and ammonium by tryptophanase. Despite that pyruvate is found at the crossroad connecting several amino acids, the labeled or natural abundance pyruvate from tryptophan will be substantially diluted by the glucose contribution. As a consequence, these three amino acids can be labeled or unlabeled together or separately. Thus, supplementing the minimal medium with 150–200 mg/L (Shi et al. 2009a, b; Etzkorn et al. 2007) up to 1 g/L (Bellstedt et al. 2013) is sufficient to label one of these residues.

On the nitrogen side, phenylalanine and tyrosine are linked together by the aromatic and the aspartate aminotransferase. These two enzymes catalyze the reversible reaction of transamination of phenylpyruvate and 4-hydroxy-phenylpyruvate with glutamate as amino donor or

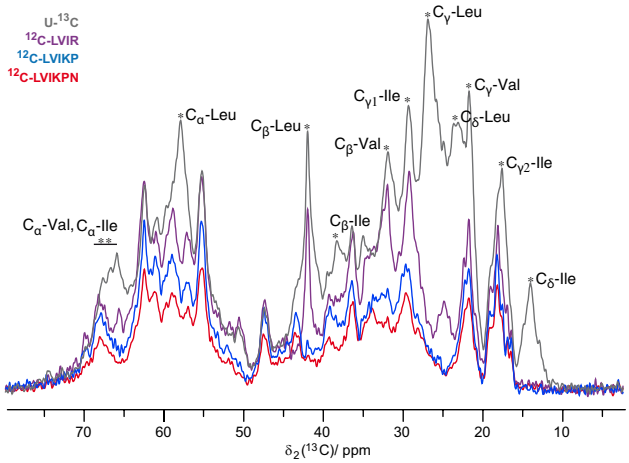


Fig. 4 Reverse labeling of the 60 kDa membrane protein BmrA. Aliphatic regions of 1D spectra of uniformly labeled BmrA (grey shade), BmrA with natural abundance LVIR (purple line), BmrA with natural abundance LVIKP (blue line) and BmrA with natural abundance LVIKPN (red shade). Sample preparation and acquisition parameters were described previously in the literature (Wiegand et al. 2017; Kunert et al. 2014). Briefly, the production of reverse labeled LVIKP-, LVIR- and LVIKPN- BmrA was performed using natural abundance Leu (0.23 g/L), Val (0.23 g/L), Ile (0.23 g/L), Lys (0.40 g/L), Pro (0.10 g/L), Arg (0.40 g/L) Gln (0.40 g/L) in presence of 2 g/L $^{15}NH_4Cl$ and ^{13}C-glucose as sole source of labeled nitrogen and labeled carbon. Natural-abundance amino acids were added to the culture medium one hour before induction

2-oxoglutarate as amino receptor (for the reverse reaction). As a result, there is non-negligible scrambling in nitrogen between these two amino acids.

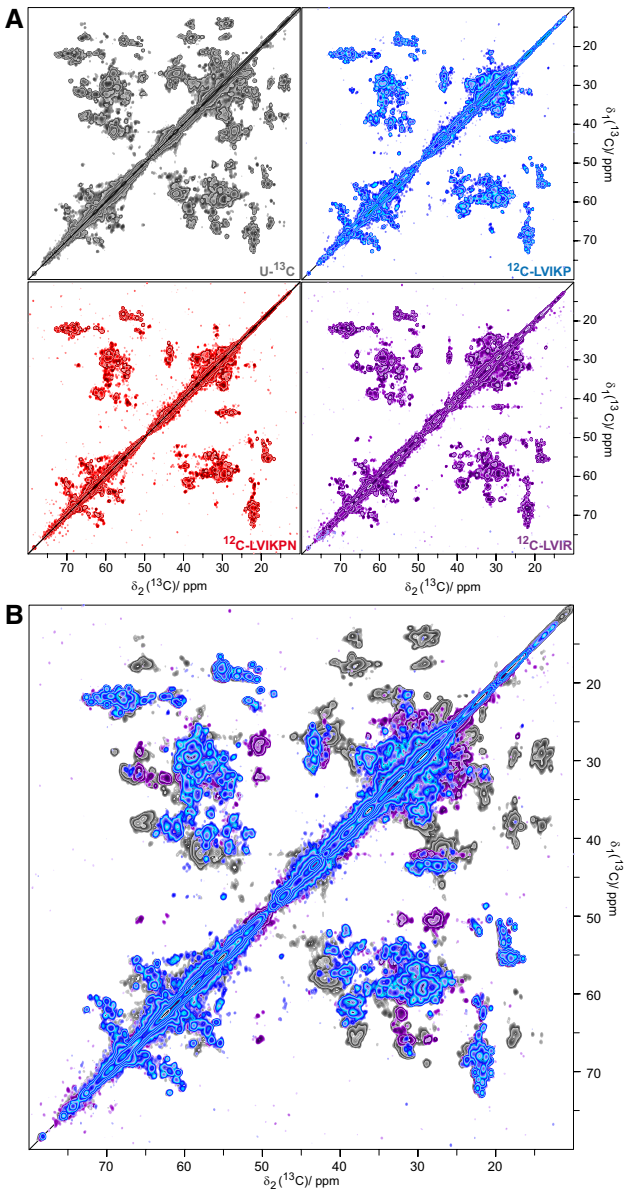


Fig. 5 2D spectra of reverse labeled BmrA. **a** Aliphatic regions of 2D ^{13}C–^{13}C DARR spectra of uniformly labeled BmrA (grey shade), BmrA with natural abundance LVIR (purple shade), BmrA with natural abundance LVIKP (blue shade) and BmrA with natural abundance LVIKPN (red shade). **b** Overlay of three spectra shown in **a** with uniformly labeled BmrA, BmrA with natural abundance LVIR (purple shade) and BmrA with natural abundance LVIKP (blue shade)

Furthermore, the branched-chain amino-acid transferase, can operate the same reversible reaction on phenylpyruvate or phenylalanine. For 3D experiments comprising a ^{15}N dimension, precursors of tyrosine or phenylalanine, respectively 4-hydroxy-phenylpyruvate and phenylpyruvate, can be used. Indeed, both precursors are located in the metabolism one step before transamination. Aminotransferases use the ^{15}N from the medium, which leads to ^{15}N labeling of tyrosine and phenylalanine without scrambling (Rasia et al. 2012).

Tryptophanase catalyzes the production of indole, pyruvate and ammonium. The latter can directly be used for the synthesis of amino acids. The action of this enzyme generates scrambling which leads to a general loss of the ^{15}N-backbone signal, when natural-abundance tryptophan is added (general loss estimated to 30% (Bellstedt et al. 2013)). To overcome this problem, solutions using precursors such as indole, or a mixture of indole and (un)labeled tryptophan can be used to reduce scrambling (Rodriguez-Mias and Pellecchia 2003; Berger et al. 2013; Schörghuber et al. 2015).

Figure 3 summarizes the scrambling induced by each other amino acid when added to the minimal medium.

Glycerol-2 and 1–3 labeling coupled with reverse labeling

Reverse labeling can as well be combined with selective labeling strategies, as for example glycerol selective labeling, which has been used in solid-state NMR in the context of structure determination of SH3 (Castellani et al. 2002) and is extensively described elsewhere (Castellani et al. 2002; Tuttle et al. 2016; Tang et al. 2011; Loquet et al. 2013; Melckebeke et al. 2010). A combination of glycerol labeling and amino-acid reverse labeling was described by Hong and Jakes (1999). There, [2-^{13}C]glycerol was added in the culture medium containing natural abundance amino acids which biosynthesis pathways go through the tricarboxylic acid cycle: Asp, Asn, Arg, Gln, Glu, Ile, Lys, Met, Pro and Thr. Addition of [2-^{13}C]glycerol thus resulted mainly in the labeling of amino acids produced by the glycolysis (Gly, Ala, Ser, Cys, Val and Leu) and pentose phosphate pathways (His, Phe, Trp and Tyr). One should note that the addition of ^{14}N Gln and Glu induces a general dilution of the ^{15}N labeling by 50%.

The method was extended to using [2-^{13}C]glycerol and [1,3-^{13}C]glycerol, this time with the combined addition of ^{15}N-labelled amino acids in order to obtain selective ^{13}C, but full ^{15}N labeling. Different combinations of amino acids yielding TEMPQANDSG-OmpG and SHLYGWAVF-OmpG labeled with [2-^{13}C] and [1,3-^{13}C] were described, and allowed the collection of distance restraints (Retel 2016). One could notice that the authors described partial scrambling in ^{13}C-^{15}N-SHLYGWAVF-OmpG, where ^{15}N-TARDNEDQIKMP was added, but peaks from QENDT were still observed in the spectrum. This indicates that the unlabeled amino-acid concentration of 50 mg/L used was likely too low, and addition of 100–200 mg/L would have been necessary to achieve optimum labeling.

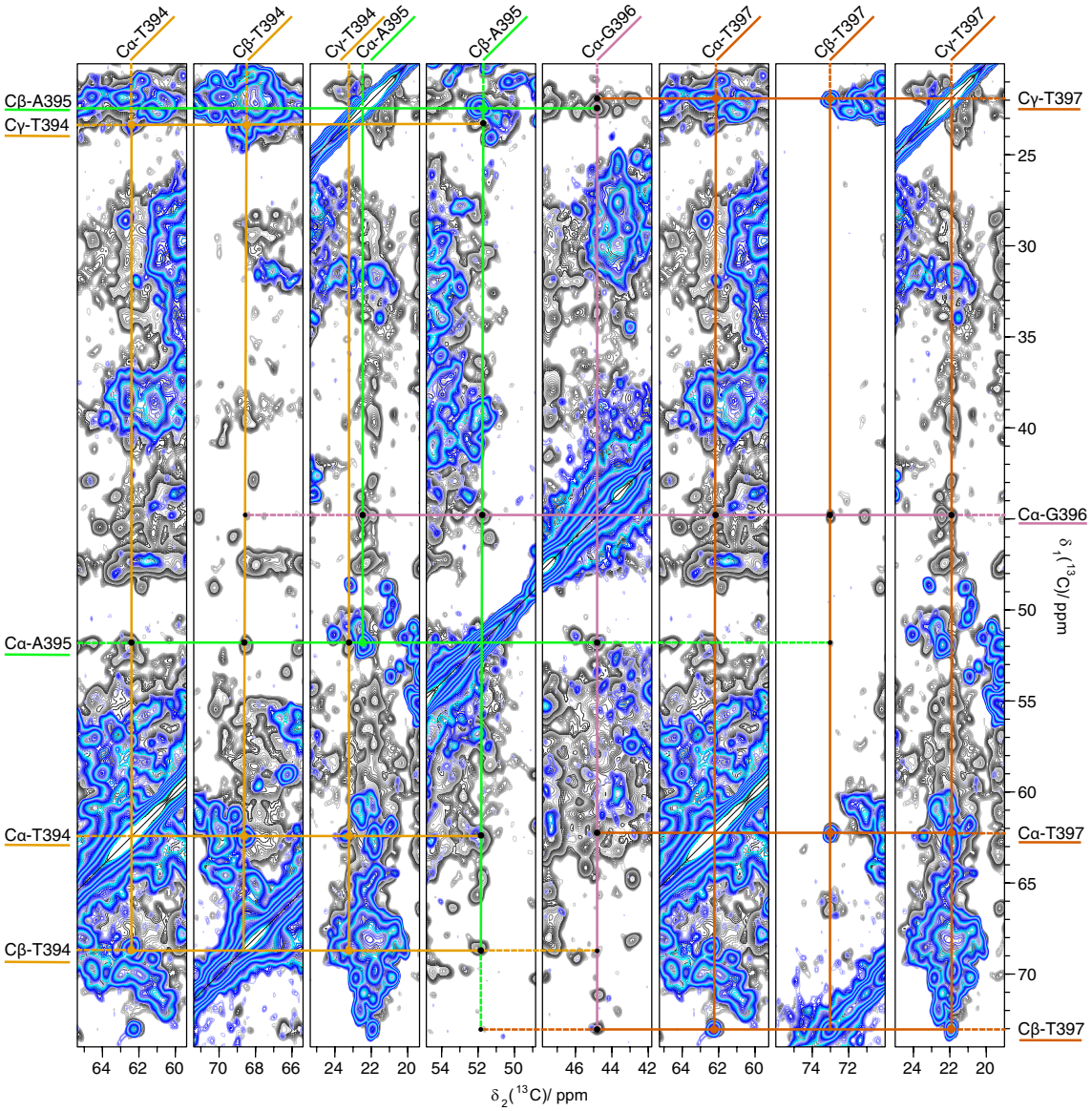


Fig. 6 Example for a sequentially assigned stretch including residues T394-A395-G396-T397, using 2D ^{13}C–^{13}C DARRs recorded with 20 ms (blue shades) and a 200 ms (grey shades) mixing time on a reverse labeled ^{12}C-^{14}N-[LVIKHP]-^{13}C-^{15}N BmrA. Inter-residue correlation peaks are represented by black dots for i + 1 and smaller black dots for i + 2. Broken lines indicate i + 2 connections. Most of the inter-residual peaks i ± 1 and i ± 2 are only isolated when using selective labeling. The residue combination TAGT is present only once in the BmrA sequence, which allows to assign sequence numbers to this stretch. For residue combinations which are present in several segments of the protein, secondary structure information obtained from a homology model in combination with secondary-structure typical chemical shifts allows to reduce assignment possibilities

Application to the ABC transporter BmrA

The ABC transporter BmrA is a dimeric membrane protein of 60 kDa per monomer. It was studied in lipid environment by solid-state NMR using a lipid-to-protein ratio of 0.5 (Kunert et al. 2014). Even though there is no inherent size limit in solid-state NMR, large membrane proteins are still a challenge due to low signal-to-noise ratios and signal overlap. The low signal-to-noise ratio makes the recording of 3D spectra difficult, and complicates sequential assignments of the protein. To alleviate overlap in these large systems, we also used paramagnetic relaxation enhancements in complement of selective labeling (Wiegand et al. 2017).

Alanine, glycine and branched residues (L, V, I) are over-represented in membrane proteins due to their hydrophobic proprieties (Ulmschneider and Sansom 2001). In the case of BmrA, they represent 45% of the protein (265 residues). In addition to this ^{13}C resonances from these residues are

spread all over the aliphatic region, leading to extensive overlap with other residues. For this reason, they are targets of choice for reverse labeling and allow simplification of the spectrum. Arginine and lysine share the same ^{13}C chemical-shift regions for their side-chain resonances, and specific labeling of only one of them allows their assignment. The branched side-chain residues, arginine and lysine do not lead to scrambling in the biosynthesis of the other amino acids, and thus allow for clean unlabeling. Figure 4 compares, in black and purple, the 1D spectra of fully labeled and LVIR-unlabeled BmrA, and one can see that the signal in regions which should remain unaffected, for example glycine around 48 ppm, remains conserved.

The 2D ^{13}C–^{13}C spectrum of BmrA without LVIR is presented in purple shade in Fig. 5. It is substantially less crowded than the spectrum recorded on uniformly labeled BmrA, and it displays many isolated peaks. With a view on further simplification of the spectrum, proline can be removed in order to clean up some regions in the spectrum to obtain several unique threonine, serine, glycine and alanine frequencies. This operation helps to identify certain amino acids and their neighbors in 2D spectra using longer mixing times (see below). However, as mentioned previously, adding natural abundance proline affects the signal intensity of the other residues. Even though natural abundance lysine was added instead of arginine, a 25% signal loss—due to natural abundance proline—of the other residues can be observed (Fig. 4). Despite the signal loss, the spectrum of BmrA with LVIKP in natural abundance (blue shade, Fig. 5) shows sufficient S/N and, in combination with the spectrum of LVIR-BmrA, allows the assignment and the characterization of several residues (Wiegand et al. 2017). Natural abundance asparagine was also added to the medium together with LVIKP in order to discriminate between aspartate and asparagine. However, as asparagine is converted into aspartate, and the latter is located at a metabolic crossroad, the result is substantial isotopic dilution and as a consequence NMR signal reduction of about 45% for natural abundance LVIKPN-BmrA, as seen in the 1D spectrum in Fig. 4.

The reduction of the number of resonances by unlabeling allowed for the recording of 2D ^{13}C spectra with a long mixing time displaying sequential correlations. These can be used, for isolated frequencies, to identify neighboring residues and in some cases partial sequential assignments. For BmrA, we used DARR spectra recorded with mixing time of 200 ms to obtain information about the $i \pm 1$ neighbors and even $i \pm 2$. A similar experiment recorded on fully-labeled samples yielded an overcrowded spectrum. Removal of Pro, Ile and Val cleans up the spectrum in the serine and threonine regions, whereas removing Leu, Ile and Pro gives a privileged access to the alanine and glycine regions. Figure 6 shows an example of the sequential assignment of a four residue amino-acid stretch.

Conclusions

We compiled and summarized information from the literature in order to highlight the strength of amino acids selective labeling or unlabeling in the context of solid-state NMR studies. Indeed, selective isotope labeling is central for the study of large proteins, and shall contribute to the further development of NMR of biological system by pushing the size limits imposed by signal overlap. We demonstrated the use of amino-acid selective labeling for the ABC transporter BmrA, a protein composed by a large number of hydrophobic residues, which unlabeling leads to decongestion of spectra and allows to initiate sequence-specific analyses.

Acknowledgements This work was supported by the Swiss National Science Foundation (Grant 200020_159707), the French ANR (ANR-14-CE09-0024B), and the LABEX ECOFECT (ANR-11-LABX-0048) within the Université de Lyon program Investissements d'Avenir (ANR-11-IDEX-0007).

References

Ayala I, Sounier R, Usé N, Gans P, Boisbouvier J (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J Biomol NMR 43:111–119

Bellstedt P et al (2013) Resonance assignment for a particularly challenging protein based on systematic unlabeling of amino acids to complement incomplete NMR data sets. J Biomol NMR 57:65–72

Berger C, Berndt S, Pichert A (2013) Efficient isotopic tryptophan labeling of membrane proteins by an indole controlled process conduct. Biotechnol Bioeng 110:1681–1690

Cao C et al (2014) Selective (15)N-labeling of the side-chain amide groups of asparagine and glutamine for applications in paramagnetic NMR spectroscopy. J Biomol NMR 59:251–261

Castellani F et al (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. Nature 420:98–102

Etzkorn M et al (2007) Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. Angew Chem Int Ed 46:459–462

Gelis I et al (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131:756–769

Goux WJ, Strong AA, Schneider BL, Lee WN, Reitzer LJ (1995) Utilization of aspartate as a nitrogen source in Escherichia coli. Analysis of nitrogen flow and characterization of the products of aspartate catabolism. J Biol Chem 270:638–646

Higman VA et al (2009) Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. J Biomol NMR 44:245–260

Hiroaki H, Umetsu Y, Nabeshima Y-I, Hoshi M, Kohda D (2011) A simplified recipe for assigning amide NMR signals using combinatorial 14N amino acid inverse-labeling. J Struct Funct Genom 12:167–174

Hong M, Jakes K (1999) Selective and extensive 13C labeling of a membrane protein for solid-state NMR investigations. J Biomol NMR 14:71–74

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361

Kaur H et al (2015) The ABC exporter MsbA probed by solid state NMR—challenges and opportunities. Biol Chem 396:1135–1149
Kazakov AE et al (2009) Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. J Bacteriol 191:52–64
Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J (2015) ScienceDirect Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. Curr Opin Struct Biol 32:113–122
Keseler IM, Mackie A, Santos-Zavaleta A (2016) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 45:543–550
Kunert B et al (2014) Efficient and stable reconstitution of the ABC transporter BmrA for solid-state NMR studies. Front Mol Biosci 1:5
Loquet A et al (2013) Atomic structure and handedness of the building block of a biological assembly. J Am Chem Soc 135:19135–19138
Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. Methods Enzymol 177:44–73
Nishida N et al (2006) Probing dynamics and conformational change of the GroEL-GroES complex by 13C NMR spectroscopy. J Biochem 140:591–598
Nuzzio KM et al (2016) High-resolution NMR studies of human tissue factor. PLoS ONE 11:e0163206
Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 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
Plevin MJ, Boisbouvier J (2012) Isotope-labelling of methyl groups for NMR studies of large proteins. Recent Dev Biomol NMR 38:1–24
Rasia RM, Brutscher B, Plevin MJ (2012) Selective isotopic unlabeling of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. ChemBioChem 13:732–739
Reitzer L (2005) Catabolism of amino acids and related compounds. EcoSal Plus. https://doi.org/10.1128/ecosalplus.3.4.7
Retel JS (2016) Structure determination of outer membrane protein G in native lipids by solid-state NMR spectroscopy. Thesis Freie Universität Berlin
Rodriguez-Mias RA, Pellecchia M (2003) Use of selective Trp side chain labeling to characterize protein-protein and protein-ligand interactions by NMR spectroscopy. J Am Chem Soc 125:2892–2893
Rudman D, Meister A (1953) Transamination in *Escherichia coli*. J Biol Chem 200:591–604
Schörghuber J et al (2015) Novel approaches in selective tryptophan isotope labeling by using *Escherichia coli* overexpression media. ChemBioChem 16:746–751
Shi L et al (2009a) Three-dimensional solid-state NMR Study of a seven-helical integral membrane proton pump—structural insights. J Mol Biol 386:1078–1093
Shi L, Lake EMR, Ahmed MAM, Brown LS, Ladizhansky V (2009b) Solid-state NMR study of proteorhodopsin in the lipid environment: secondary structure and dynamics. Biochim Biophys Acta 1788:2563–2574
Shortle D (1994) Assignment of amino acid type in 1H-15N correlation spectra by labeling with 14N-amino acids. J Magn Reson 105:88–90
Stoffregen MC, Schwer MM, Renschler FA, Wiesner S (2012) Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. Structure 20:573–581
Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234
Takeuchi K, Ng E, Malia TJ, Wagner G (2007) 1–13C amino acid selective labeling in a 2H15N background for NMR studies of large proteins. J Biomol NMR 38:89–98
Tang M et al (2011) High-resolution membrane protein structure by joint calculations with solid-state NMR and X-ray experimental data. J Biomol NMR 51:227–233
Tate S, Tate NU, Ravera MW, Jaye M, Inagaki F (1992) A novel 15N-labeling method to selectively observe 15NH2 resonances of proteins in 1H-detected heteronuclear correlation spectroscopy. FEBS Lett 297:39–42
Tuttle MD et al (2016) Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. Nat Struct Mol Biol 23:409–415
Ulmschneider MB, Sansom MS (2001) Amino acid distributions in integral membrane protein structures. Biochim Biophys Acta 1512:1–14
Van Melckebeke H, Wasmer C, Lange A (2010) Atomic-resolution three-dimensional structure of HET-s (218–289) amyloid fibrils by solid-state NMR spectroscopy. J Am Chem Soc 132:13765–13775
Velyvis A, Ruschak AM, Kay LE (2012) An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. PLoS ONE 7:e43725
Waugh DS (1996) Genetic tools for selective labeling of proteins with α-15N-amino acids. J Biomol NMR 8:184–192
Wiegand T, Lacabanne D, Keller K, Solid-state (2017) NMR and EPR spectroscopy of Mn2+-substituted ATP-fueled protein engines. Angew Chem. https://doi.org/10.1002/ange.201610551
Zhou Y, Zhu W, Bellur PS, Rewinkel D, Becker DF (2008) Direct linking of metabolism and gene expression in the proline utilization A protein from *Escherichia coli*. Amino Acids 35:711–718

Affiliations

Denis Lacabanne[1] · Beat H. Meier[2] · Anja Böckmann[1]

✉ Beat H. Meier
beme@ethz.ch

✉ Anja Böckmann
a.bockmann@ibcp.fr

[1] Molecular Microbiology and Structural Biochemistry, Labex Ecofect, UMR 5086 CNRS, Université de Lyon, 7 passage du Vercors, 69367 Lyon, France

[2] Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland