

Anthranilic acid, the new player in the ensemble of aromatic residue labeling precursor compounds

Julia Schörghuber¹ · Leonhard Geist² · Marilena Bisaccia¹ · Frederik Weber¹ · Robert Konrat² · Roman J. Lichtenecker¹

Received: 14 June 2017 / Accepted: 28 August 2017 / Published online: 31 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract The application of metabolic precursors for selective stable isotope labeling of aromatic residues in cell-based protein overexpression has already resulted in numerous NMR probes to study the structural and dynamic characteristics of proteins. With anthranilic acid, we present the structurally simplest precursor for exclusive tryptophan side chain labeling. A synthetic route to ¹³C, ²H isotopologues allows the installation of isolated ¹³C–¹H spin systems in the indole ring of tryptophan, representing a versatile tool to investigate side chain motion using relaxation-based experiments without the loss of magnetization due to strong ¹J_{CC} and weaker ²J_{CH} scalar couplings, as well as dipolar interactions with remote hydrogens. In this article, we want to introduce this novel precursor in the context of hitherto existing techniques of in vivo aromatic residue labeling.

Keywords Isotope labeling · Protein NMR · Tryptophan · Anthranilic acid

Introduction

The progress in stable isotope labeling has driven the field of protein NMR spectroscopy, ever since the first spectra of small randomly deuterated or uniformly ¹⁵N-labeled proteins have been recorded (Campbell 2013). Numerous experimental techniques have been developed in the meantime to address a wide range of structure-biological questions (Lian and Middleton 2001; Ohki and Kainosho 2008; Zhang and van Ingen 2016). However, poor signal resolution and complex magnetization transfer causes severe problems in the case of high molecular weight target proteins, which can only be addressed using highly defined patterns of heavy isotopes (¹³C, ¹⁵N, ²H). Early examples of selective stable isotope labeling almost exclusively focused on aliphatic side chains (Goto and Kay 2000; Kerfah et al. 2015). Their high abundance in the core of globular proteins, as well as their intense NMR signals due to the threefold symmetry of a methyl group rotation result in straightforward data acquisition for structure calculation. These aliphatic side chains play an important role in protein folding, creating a well-defined hydrophobic microenvironment for substrate binding and solvent exclusion.

Although some very early examples can be found in literature (Wüthrich and Wagner 1975), it was only in recent years, that studies of aromatic side chains increasingly attracted the attention of the biomolecular NMR community. Aromatic side chains provide valuable additional restraints for structure calculation and exhibit a number of characteristics, which turn them into important sensors to study enzymatic mechanisms or binding events. They are prone to participate in non-covalent π–π or π–cation interactions, which significantly contribute to the protein's tertiary structure (Dougherty 1996). Aromatic residues additionally control enzymatic reaction trajectories by stabilizing charged

Electronic supplementary material The online version of this article (doi:10.1007/s10858-017-0129-2) contains supplementary material, which is available to authorized users.

✉ Roman J. Lichtenecker
roman.lichtenecker@univie.ac.at

¹ Institute of Organic Chemistry, University of Vienna, Währingerstr. 38, 1090 Vienna, Austria

² Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology, Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Dr-Bohr-Gasse 9, 1030 Vienna, Austria

intermediates or direct structural changes to close hydrophobic cavities and protect them from water molecules. Due to their distinct physicochemical properties, aromatic side chains play a very prominent role at binding interfaces to other proteins, nucleic acids, lipid membranes or small ligands (Moreira et al. 2013; Rahman et al. 2015). These binding events, as well as conformational changes during enzymatic catalysis, are frequently connected to extensive aromatic side chain motion in a wide range of time scales from picoseconds to seconds.

Side chain dynamics, induced by binding events or changes in temperature or pressure, have been studied using different NMR based techniques (Boehr et al. 2006). The corresponding experiments, such as longitudinal- and transverse- ^{13}C -relaxation (Weininger et al. 2012a, 2014a; Kasinath et al. 2015), the ^{13}C -TROSY-Carr-Purcell-Meiboom-Gill (CPMG) experiment (Weininger et al. 2012b), the ^{13}C -TROSY-rotating frame relaxation ($R_{1\rho}$) sequence (Weininger et al. 2014b), intra-residue NOE interpretation (Miyanoiri et al. 2011), or the recording of aromatic residual dipolar couplings (Sathyamoorthy et al. 2013) have been optimized for applications in aromatic spin systems. These methods resulted in the calculation of rate constants, energy barriers, as well as activation volumes. As mentioned above, the data of such experiments can only be interpreted when spectra have been simplified and magnetization transfer pathways optimized using defined isotope distributions in the protein samples.

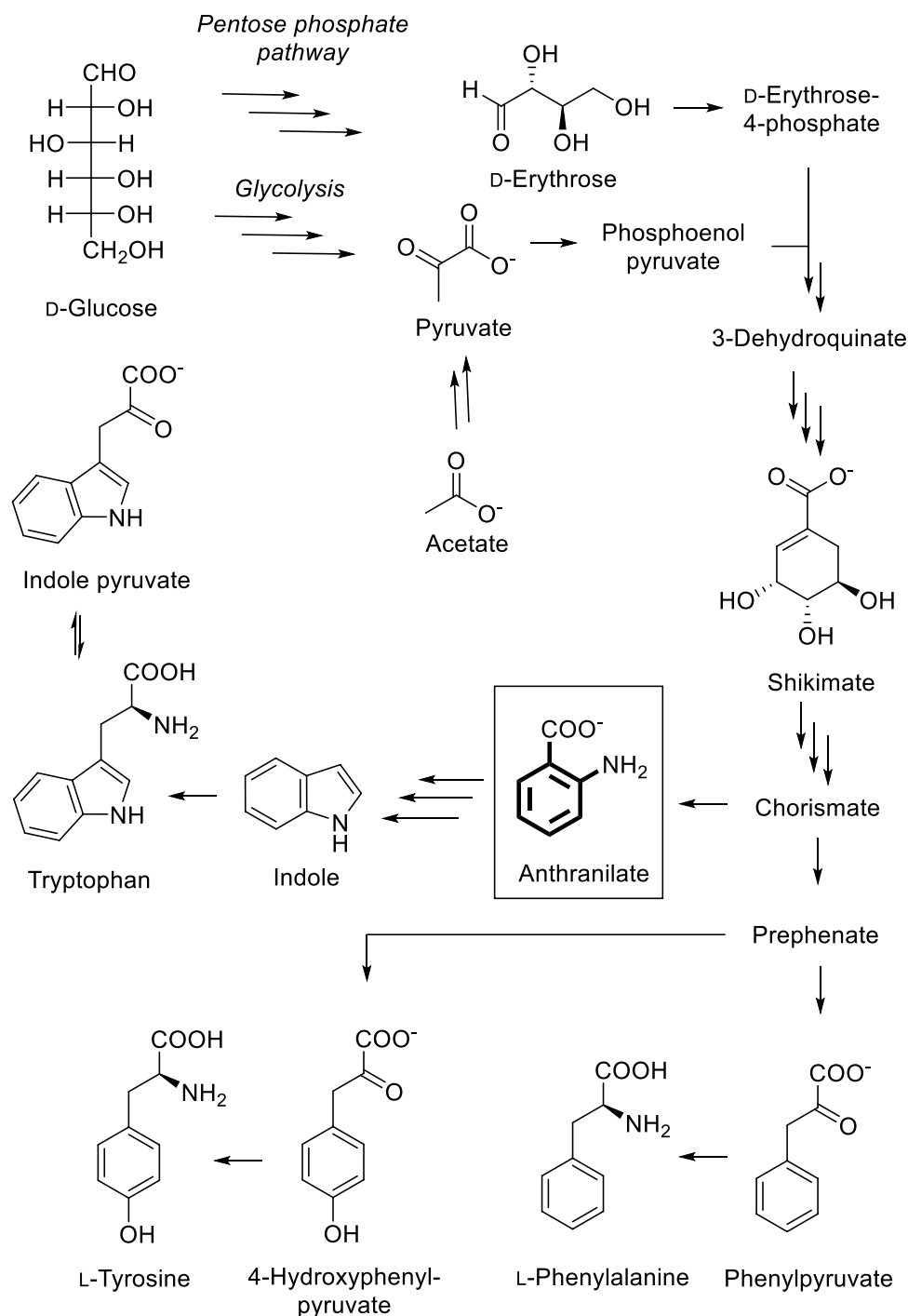
Two main strategies can be pursued to generate defined isotope patterns in a protein of interest: In cell-free overexpression methods all components needed for the transcription and the translation process are provided to an *in vitro* system. This approach is well suited to generate proteins, which contain posttranslational modifications or modified amino acids, as well as proteins that are toxic to *in vivo* systems. Selective labeling devoid of cross labeling to other residues is assured, but the method requires the application of labeled amino acids (Torizawa et al. 2005; Kainosho et al. 2006; Takeda et al. 2010; Miyanoiri et al. 2011). Moreover, cell-free protein expression exhibits major shortcomings in providing high protein yields and is considerably expensive (Staunton et al. 2006).

The second strategy features isotope labeled precursor compounds as additives in the growth medium of an overexpressing microorganism, which has been established as the method of choice to yield in high concentrations of the corresponding labeled molecular targets (Hoogstraten and Johnson 2008). Potential pitfalls include poor precursor uptake or cross labeling to other residues. Scheme 1 summarizes the main metabolites, which have been used to overexpress proteins containing labeled phenylalanine, tyrosine or tryptophan residues. Certain ^{13}C -patterns of D-glucose (Teillum et al. 2006; Lundström et al. 2007), acetate (Wand et al.

1995), erythrose (Kasinath et al. 2013; Weininger 2017), pyruvate (Guo et al. 2009; Milbradt et al. 2015), or glycerol (LeMaster and Kushlan 1996; Ahlner et al. 2015) have been applied as carbon sources to avoid direct ^{13}C - ^{13}C coupling in the target residues. These compounds are commercially available and additional organic synthesis is usually not required. However, the costs are considerable regarding the concentrations needed for high incorporation levels (up to 3 g/L medium). These methods partially exhibit poor selectivity, as the corresponding isotope sources represent early metabolic intermediates residing upstream of the shikimate pathway. This inevitably leads to the undesired loss of heavy isotopes at metabolic branches. The metabolites of the shikimate pathway 3-dehydroquinate to prephenate feature multiple chiral centers and are therefore rather too complex to be considered as isotope containing targets in economic multistep organic synthesis. In only one example unlabeled shikimate was thus used to introduce protonated aromatic residues into an otherwise deuterated protein (Rajesh et al. 2003). As soon as the aromatic ring system is biosynthetically formed, the corresponding metabolic intermediates represent much more attractive target molecules for organic synthesis. In rare cases, aromatic amino acids have been utilized in combination with *in vivo* overexpression media for selective labeling/unlabeling (Vuister et al. 1994; Kelly et al. 1999; Wang et al. 1999). However, these examples often suffer from poor metabolic uptake and increased effects on metabolic network regulation, since many of the enzymes involved are inhibited by the final product of the corresponding biosynthetic pathway. These methods additionally lack the possibility of simple uniform ^{15}N labeling by overexpressing the target protein in presence of ^{15}N -ammonium chloride, which is beneficial for many NMR applications. α -Ketoacids can be considered as late metabolic intermediates and their use as precursors for aliphatic residue labeling has a long history due to the effective uptake and the straightforward aminotransferase-catalyzed conversion to the target amino acids (Goto et al. 1999; Lichtenecker et al. 2004, 2013a). In a similar approach for aromatic residue labeling, phenylpyruvate has been shown to be an effective precursor for phenylalanine, whereas the use of 4-hydroxyphenylpyruvate leads to tyrosine labeling (Lichtenecker et al. 2013b). Isotopologues of indole have been used to selectively label tryptophan side chains (Schörghuber et al. 2015; Rodriguez-Mias and Pellecchia 2003). Alternatively, indole pyruvate can be applied for backbone Trp-labeling, which is the first intermediate in the tryptophan degradation pathway, but is effectively converted to the target amino acid due to the highly reversible character of the Trp-aminotransferase.

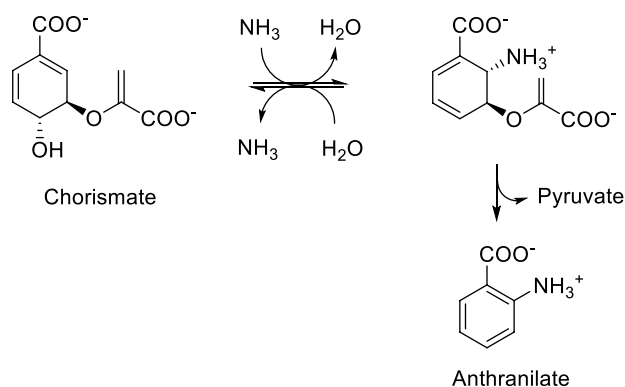
The structurally simplest precursor, which holds promise to result in exclusive tryptophan labeling is anthranilic acid. Although ^{15}N -isotopologues of this compound have been used in the past to produce $[1-^{15}\text{N}]\text{L}$ -tryptophan by

Scheme 1 Overview of aromatic residue labeling precursor compounds in *E. coli* overexpression media



fermentation (Liu et al. 2009) and protocols for the synthesis of different stable isotope patterns have been reported (van Liemt et al. 1996), methods to apply anthranilic acid as a direct metabolic precursor in protein overexpression media are still not available, at least to our knowledge. We hypothesized that none of the heavy isotopes present when using anthranilic acid as a tryptophan precursor should be

metabolized back into the shikimic acid pathway. Since the second reaction step in the anthranilate synthase (*EC 4.1.3.27*) catalyzed conversion of chorismate to anthranilate is an irreversible process (elimination of pyruvate, Scheme 2), we expected exclusive labeling of the tryptophan target residues without any cross labeling to phenylalanines and tyrosines.



Scheme 2 Anthranilate synthase catalyzed conversion of chorismate to anthranilate

Materials and methods

Organic synthesis

Detailed protocols for the synthesis of the anthranilic acid isotopologues [^{15}N]anthranilic acid **4**, 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11**, and 4,6-dideuterio[5- ^{13}C]anthranilic acid **18**, as well as NMR and MS characterization of intermediates and final products are provided in the supporting information.

Protein overexpression

His-tagged GB1

The pET-M11 plasmid containing the sequence for a his-tagged GB1 protein (H6-GB1) was transformed into an *E. coli* BL21(DE3) strain. Bacterial cultures were grown in LB medium and then transferred to a M9 minimal medium supplemented either with unlabeled NH_4Cl (1 g/L) and increasing amounts of [^{15}N]anthranilic acid **4** to acquire selectively labeled protein or with $^{15}\text{NH}_4\text{Cl}$ (1 g/L) to acquire uniformly labeled protein (Marley et al. 2001). Protein overexpression was induced by the addition of isopropyl- β -D-thiogalactopyranosid and pursued overnight. The crude proteins were purified using Ni^{2+} affinity chromatography. NMR samples contained 1 mM protein and 10% D_2O .

Brd4-BD1

Recombinant human Brd4-BD1 (bromodomain 1 of Bromodomain containing protein 4) was expressed in *E. coli* BL21(DE3) containing an N-terminal TEV-cleavable His6-tag (the expression plasmid was kindly provided by Boehringer Ingelheim). Uniformly ^{15}N /selectively $^{13}\text{C}_2$ -tryptophan labeled H6-TEV-Brd4-BD1 was expressed following the expression protocol for efficient isotopic labeling

of recombinant proteins using a fourfold cell concentration in M9 minimal medium supplemented with $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11** (50 mg/L) (Marley et al. 2001). Uniformly ^{15}N / uniformly ^{13}C -labeled Brd4-BD1 was overexpressed using 1 g/L $^{15}\text{NH}_4\text{Cl}$ and 3 g/L $^{13}\text{C}_6$ -glucose in the minimal medium. Cells were harvested by centrifugation, lysed by sonication and the lysates were subsequently centrifuged. Proteins were purified from the resulting supernatant by Ni^{2+} affinity chromatography. The purified protein was treated with TEV protease and again loaded onto a Ni^{2+} column to bind the cleaved His6-tag and the His6-tagged TEV protease. The flow-through containing Brd4-BD1 was concentrated and purified on a gel filtration column. NMR samples of Brd4-BD1 were prepared in sodium phosphate buffer containing 0.4–3 mM protein and 10% D_2O .

Comprehensive information concerning the overexpression of labeled H6-GB1 and Brd4-BD1 is given in the supporting information.

Protein NMR spectroscopy

NMR spectra of H6-GB1 and Brd4-BD1 were acquired at 298 K on a Bruker Avance 3 HD + 600 MHz spectrometer. The system was equipped with a RT 5 mm TXI probe. Spectra were processed using NMR Pipe (Delaglio et al. 1995) and analyzed using the Sparky software (Goddard and Kneller). Methodological details concerning the quantification of H6-GB1 labeling using compound **4**, as well as an evaluation of Trp- ^{13}C -labeling of Brd4-BD1 using compound **11** are reported in the supporting information.

Results and discussion

In order to investigate the uptake of anthranilic acid by the expressing *E. coli* strain, we prepared [^{15}N]anthranilic acid **4** in a two-step sequence starting from $^{15}\text{NH}_4\text{Cl}$ **1** or [^{15}N]urea **2** (Scheme 3) (Hijji and Benjamin 2003). An *E. coli* strain, grown in the presence of compound **4**, overexpressed the exclusively tryptophan- ^{15}N histidine-tagged G-binding protein domain (H6-GB1). The resulting ^{15}N -HSQC spectrum is devoid of any other signals, except the one caused by the only $^1\text{H}_\epsilon$ - $^{15}\text{N}_\epsilon$ Trp side chain spin system in the target protein (Fig. 1). [^{15}N]Anthranilic acid **4** was applied to correlate the precursor concentrations in the growth medium to the protein isotope incorporation rates. The corresponding plot in Fig. 1b displays normalized ^{15}N -HSQC signal intensities of H6-GB1 overexpressed in presence of different precursor **4** concentrations. Signal intensities were compared to the intensity found for the $^1\text{H}_\epsilon$ - $^{15}\text{N}_\epsilon$ Trp side chain signal in uniformly ^{15}N -labeled H6-GB1 (experimental details are given in the supporting information). A signal intensity of

Scheme 3 Synthesis of [^{15}N] anthranilic acid **4** and its application to selectively label Trp-residues. **a** Phthalic anhydride, 250 °C, DMAP, 66%; **b** phthalic anhydride, xylenes, 140 °C, 84%; **c** NaOH, Br $_2$, 53%; **d** *E. coli* protein overexpression, D-glucose, NH $_4$ Cl

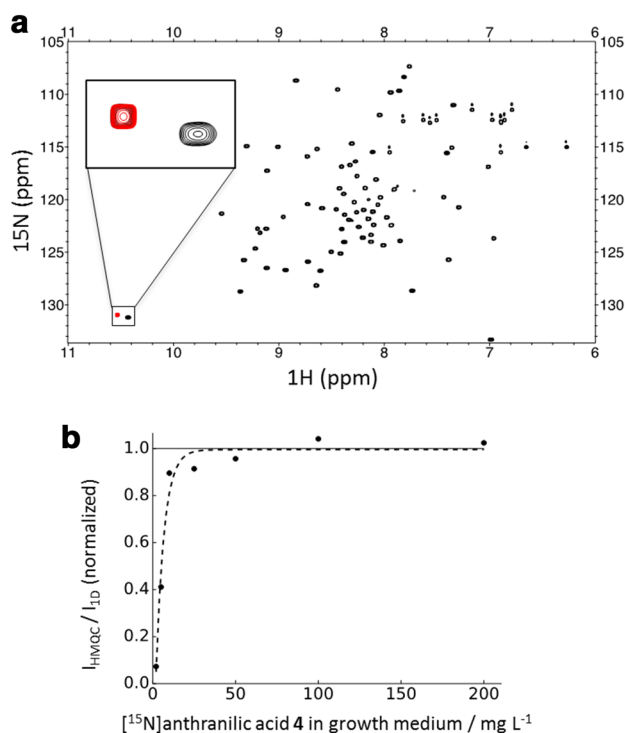
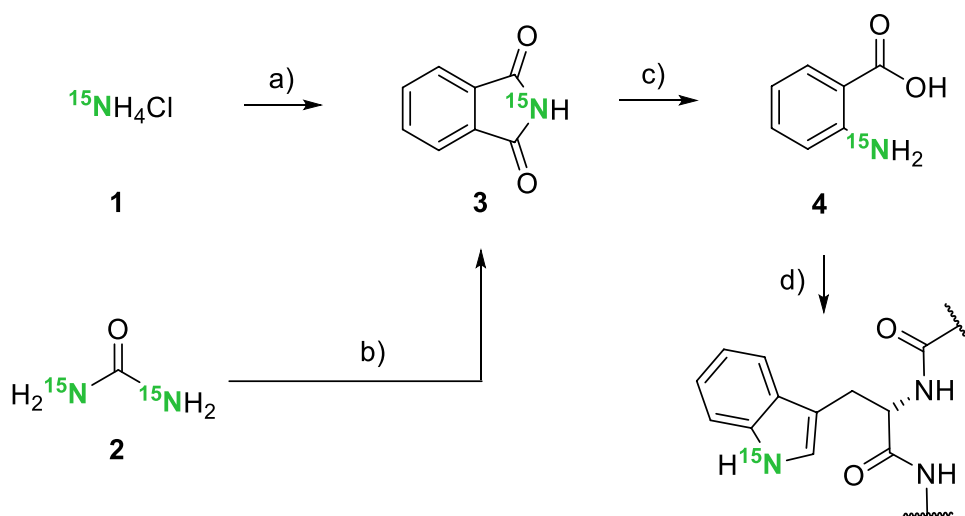


Fig. 1 **a** ^{15}N -HSQC spectra overlay of the uniformly ^{15}N -labeled (black) and tryptophan side chain ^{15}N labeled (red) H6-GB1 protein; **b** normalized signal intensity versus precursor concentration plot showing data for the overexpression of H6-GB1 in presence of 2, 5, 10, 25, 50, 100 and 200 mg/L of precursor **4**. The intensities have been normalized to a uniformly ^{15}N -labeled sample (1 on y-axis)

>90% at a concentration of 10 mg/L indicates the effective metabolic uptake and conversion of anthranilic acid by the overexpressing host organism.

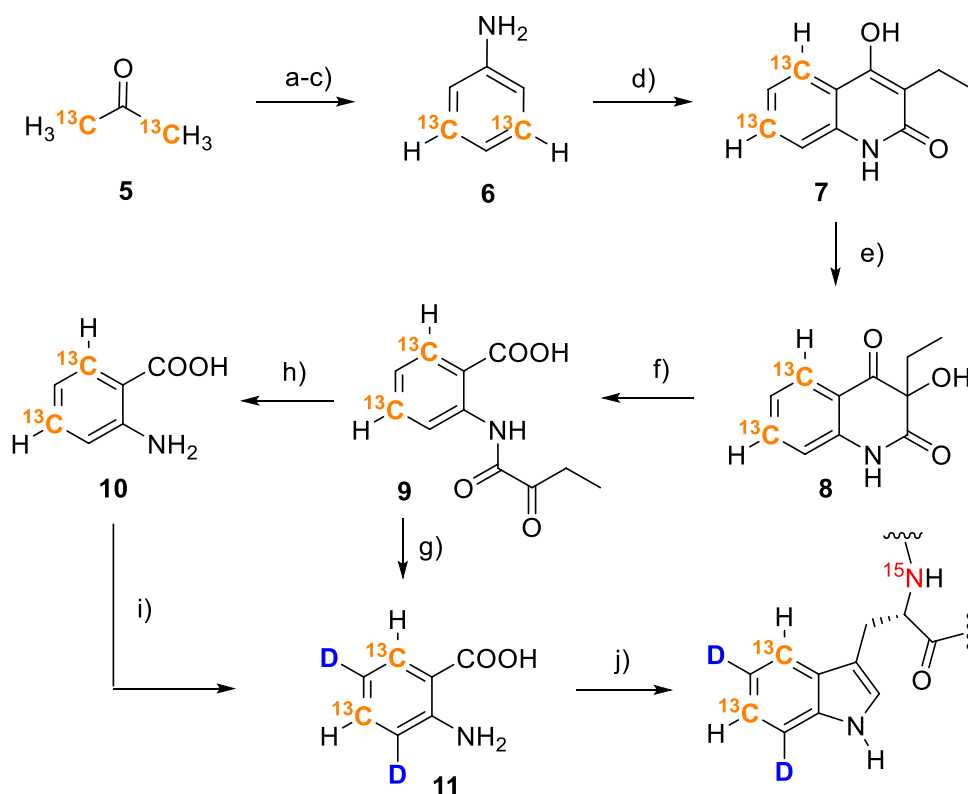
These results encouraged us to develop a synthetic route to anthranilic acid isotopologues possessing ^{13}C in the aromatic ring (Scheme 4). Starting from [1,3- $^{13}\text{C}_2$]acetone **5**,

we assembled the aromatic system through reaction with nitromalonaldehyde (Viswanatha and Hruby 1979). A two-step deoxygenation/reduction sequence yielded [3,5- $^{13}\text{C}_2$] aniline **6** (Musliner and Gates 1966), which was transformed into anthranilic acid following a literature protocol (Kafka et al. 2013). Hydrolysis of N-(α -ketoacyl)anthranilic acid **9** in DCl/D $_2$ O yielded 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11**. The ^2H -pattern can be installed using mild reaction conditions, since the electron donating amino substituent directs the reaction outcome to deuteration in the desired *ortho* and *para* positions.

Using a similar route, but starting from [2- ^{13}C]acetone **12** allows for the synthesis of the 4,6-dideuterio[5- ^{13}C] isotopologue of the target compound (**18**, Scheme 5). In this case ^2H is introduced by perdeuteration of [^{13}C]aminophenol in acidic D $_2$ O at 180 °C at prolonged reaction times (**13**→**14**). The labeling patterns of compounds **11** and **18** are well defined (Fig. 2) and feature isolated ^{13}C - ^1H spin systems, which can be effectively applied for the elucidation of tryptophan dynamics using ^{13}C transverse relaxation dispersion based methods.

3,5-Dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11** was supplemented to the *E. coli* growth medium to overexpress the recombinant human bromodomain 1 of Bromodomain containing protein 4 featuring an N-terminal TEV-cleavable His6-tag (H6-TEV-Brd4-BD1) (see “Materials and methods” section for details). BRD4 is linked to diverse human malignancies and has been shown to regulate the expression of *MYC* (Andrews et al. 2017). The *MYC* gene encodes a transcription factor, which plays a decisive part in cancer development and proliferation. The corresponding bromodomain was chosen as a model to demonstrate exclusive Trp labeling by our novel precursor compound, as well as give a first representative of its potential use. The resulting ^{13}C -HSQC spectra showed a total number of six signals, caused by the ^{13}C - $^1\text{H}_{\text{e}3}$ and ^{13}C - $^1\text{H}_{\text{e}2}$ couplings of the three

Scheme 4 Synthesis of 3,5-dideuterio[4,6- $^{13}\text{C}_2$] anthranilic acid **11** and its application to selectively label Trp-residues. **a** Sodium nitromalonaldehyde monohydrate, aqu. NaOH, then 6N HCl, 63%; **b** potassium *tert*-butoxide, dimethylformamide, 5-chloro-1-phenyl-1*H*-tetrazole, 74%; **c** Pd/C, methanol, toluene, H_2 , H-cube[®], 97%; **d** diethyl ethylmalonate, 250–270 °C, then toluene and aqu. NaOH, 77%; **e** peroxyacetic acid in acetic acid, aqu. NaOH; **f** paraperiodic acid, ethanol, 75% over steps (ef); **g** DCI/ D_2O , 94%; **h** HCl/ H_2O , 91%; **i** microwave irradiation, DCI/ D_2O , 92%; **j** *E. coli* protein overexpression, D-glucose, $^{15}\text{NH}_4\text{Cl}$



Scheme 5 Synthesis of 4,6-dideuterio[5- ^{13}C]anthranilic acid **18** and its application to selectively label Trp-residues. **a** Sodium nitromalonaldehyde monohydrate, aqu. NaOH, then 6N HCl, 60%; **b** Pd/C, H_2 , H-cube[®], methanol, toluene, 95%; **c** HCl/ D_2O , MW irradiation, 180 °C, 7 h, 90%; **d** potassium *tert*-butoxide, dimethylformamide, 5-chloro-1-phenyl-1*H*-tetrazole, 75%; **e** Pd/C 10%, MeOH, toluene, H_2 , H-cube[®], 95%; **f** diethyl ethylmalonate, heat, then toluene and aqu. NaOH, 69%; **g** peroxyacetic acid in acetic acid, aqu. NaOH; **h** paraperiodic acid, ethanol, 72% over steps (g, h); **i** HCl/ H_2O , 90%; **j** *E. coli* protein overexpression, D-glucose, $^{15}\text{NH}_4\text{Cl}$

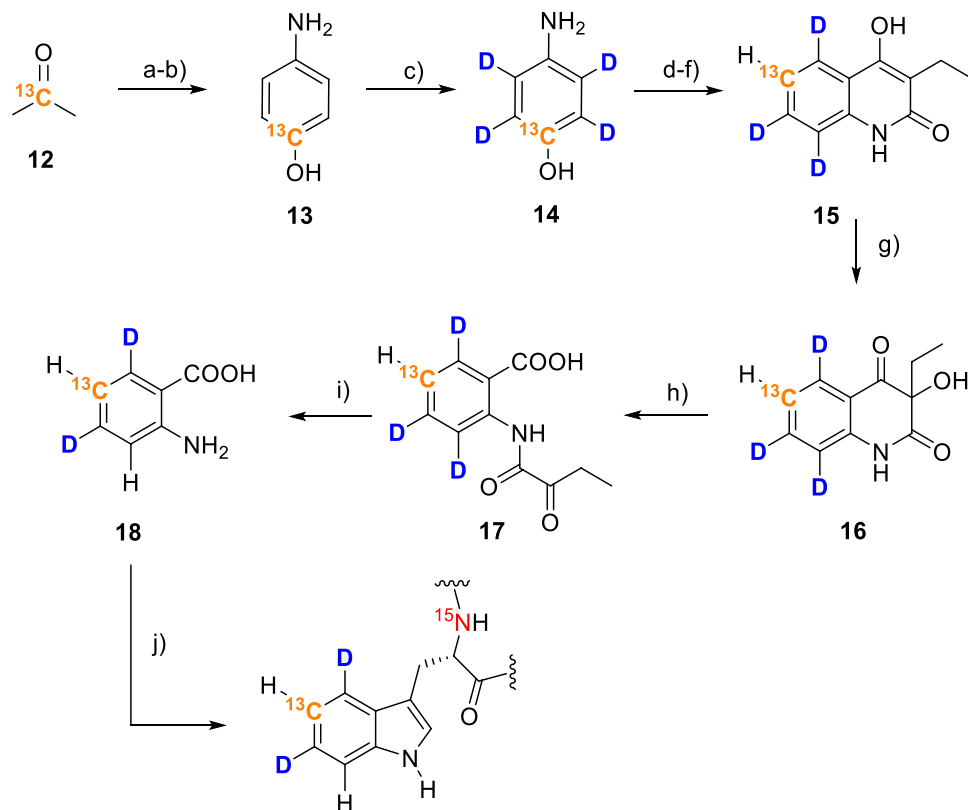
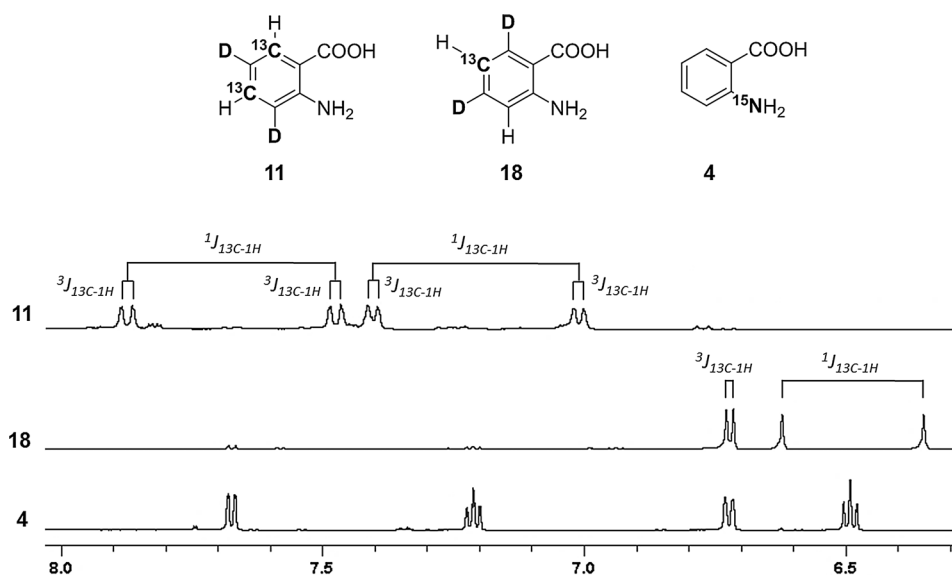


Fig. 2 ^1H -NMR spectral region of aromatic protons showing the corresponding ^{13}C - ^1H couplings of 3,5-dideuterio [4,6- $^{13}\text{C}_2$]anthranilic acid **11**, 4,6-dideuterio [5- ^{13}C]anthranilic acid **18** and [^{15}N]anthranilic acid **4**, respectively



tryptophan residues present in the corresponding sequence (Fig. 3a). Comparing this HSQC with spectra obtained from a uniformly ^{13}C labeled sample, which was overexpressed in presence of $^{13}\text{C}_6$ -glucose, reveals the value of our labeling approach in terms of well-defined signals of high intensity in absence of signal splitting or line shape distortion due to scalar couplings (Fig. 3b). The ^1H - ^{15}N HSQC of the Brd4-BD1 sample which was labeled using $^{15}\text{NH}_4\text{Cl}$ and compound **11** showed the complete absence of Trp- $\text{N}\epsilon_2$ resonances, given that the precursor 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11** carries an unlabeled amino-group (see supporting information).

This data confirmed our initial assumption that anthranilic acid serves as a selective tryptophan precursor with quantitative incorporation levels devoid of any cross-labeling to the other aromatic residues. Addition of a ligand (kindly provided by Boehringer Ingelheim) targeting the binding site of Brd4-BD1 for acetylated lysines in histones resulted in an upfield shift of the Trp81 signals (Fig. 3c). This residue is positioned in the bromodomain binding-cleft, thus playing a central role in protein–ligand interactions (Fig. 3d) (Zhang et al. 2015).

Conclusions

In summary, the *E. coli* overexpression of model proteins in presence of labeled anthranilic acid demonstrated both, high labeling selectivity for the target Trp-residues, as well as highly effective precursor uptake by the overexpressing host organism with quantitative isotope incorporation levels at low precursor concentrations (10–15 mg/L overexpression medium). Our novel labeling technique represents an advanced development of hitherto applied methods using

early metabolic intermediates as heavy isotope sources. Recently, isotopologues of erythrose have been introduced to direct ^{13}C labeling into defined atomic positions of the aromatic side-chains in phenylalanine, tyrosine and tryptophan (Kasinath et al. 2015; Weininger 2017). In the corresponding protocols, limited isotope incorporation levels in the range of 50–70% for tryptophan, as well as partial isotope scrambling to undesired atomic positions in the aromatic rings have been reported. On the contrary, anthranilic acid is metabolized to the target amino acid in a well-defined way without losing labeling selectivity at the major metabolic intersection points of the shikimate pathway. Our novel precursor completes our toolbox of isotope labeled compounds to transfer $^{13}\text{C}/^2\text{H}$ -patterns onto the aromatic side chains of phenylalanine and tyrosine (Lichtenecker et al. 2013b), tryptophan (Schörghuber et al. 2015), as well as histidine (Schörghuber et al. 2017), with every precursor targeting one specific type of residue. This residue selective isotope incorporation cannot be achieved using the early metabolic precursors mentioned above, but is of major importance in the case of high molecular weight protein samples, when signal overlap in the aromatic region becomes a serious issue.

Efficient metabolic *in vivo* conversion has also been observed for the selective tryptophan precursor indol, which we presented in a recent publication (Schörghuber et al. 2015). However, the synthetic routes to anthranilic acid isotopologues from the present study offer clear advantages in terms of yields and robustness and are thus preferred in cases where the γ and δ_1 positions are not targeted for ^{13}C labeling. The synthetic routes to 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11** and 4,6-dideuterio[5- ^{13}C]anthranilic acid **18** have been optimized to hold down costs for isotope sources and reagents, while at the same time minimizing synthetic efforts. Both precursors are accessible in gram

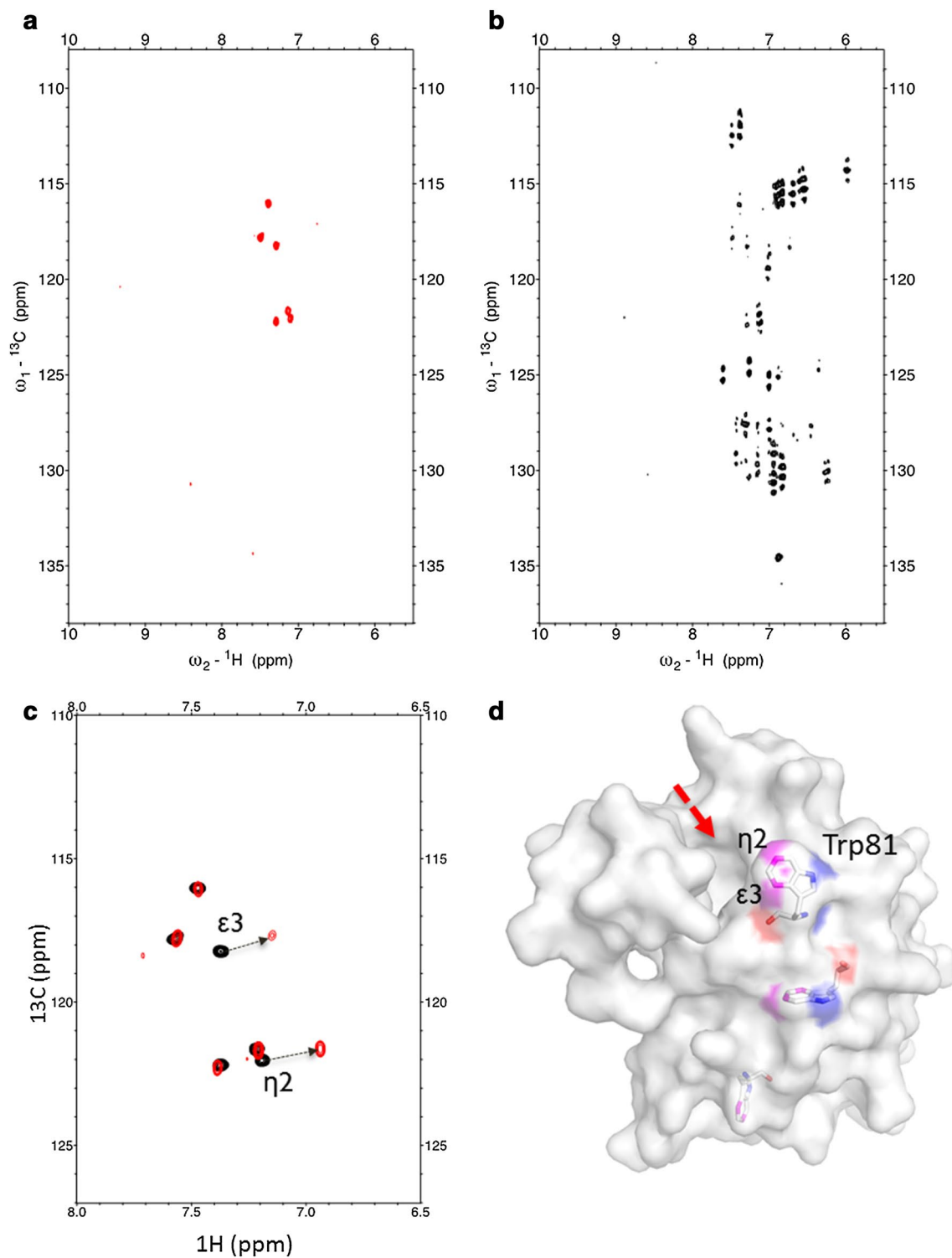


Fig. 3 **a** ^{13}C -HSQC of Brd4-BD1 overexpressed in the presence of 3,5-dideuterio[4,6- $^{13}\text{C}_2$]anthranilic acid **11**; **b** ^{13}C -HSQC of Brd4-BD1 overexpressed in the presence of [$^{13}\text{C}_6$]D-glucose; **c** ^{13}C -HSQC of selectively Trp- $^{13}\text{C}/^2\text{H}$ labeled Brd4-BD1 (*black*) and upon addition

of the Brd4-BD1 binding ligand (*red*); **d** X-ray structure of apo-Brd4-BD1 (4LYI) (Lucas et al. 2013) depicting the three tryptophan residues. Trp81 is part of the Brd4-BD1 binding pocket (*red arrow*)

scale and can be stored and handled as bench-stable solids. Considering the listed prizes for standard chemicals and ^{13}C -labeled acetone, material costs of €15–20 can be calculated to synthesize 10 mg of precursor **11** or **18**, which is the amount per liter minimal medium to achieve quantitative Trp-labeling. This is in sharp contrast to the current average prizes of commercial suppliers for isotope labeled early metabolic precursors like [1- ^{13}C]glucose (€240/g) [2- ^{13}C]glucose (€630/g), [4- ^{13}C]erythrose (€1100/g), [2- ^{13}C]pyruvate (€820/g), [3- ^{13}C]pyruvate (€1200/g) and [2- ^{13}C]glycerol (€520/g). Considering the high literature-reported concentrations required (1–3 g labeled compound/L medium) to achieve maximum incorporation levels, these cost issues severely affect their application for routine labeling of multiple protein samples (Teilum et al. 2006; Lundstrom et al. 2007; Kasinath et al. 2013; Weininger 2017; Milbradt et al. 2015).

To conclude, the novel tryptophan precursor anthranilic acid allows for heavy isotope incorporation into tryptophan side chains using *E. coli* protein overexpression with selectivities and incorporation levels, which were previously reserved only to cell-free protein synthesis. The economic advantages of our new labeling technique will prove beneficial for many future applications concerning the NMR-based studies of high molecular weight protein structure and dynamics.

Acknowledgements Open access funding provided by University of Vienna. We thank Boehringer Ingelheim for providing the expression plasmid for H6-TEV-Brd4-BD1, as well as the corresponding bromodomain ligand. This work was supported by a scholarship of the uni:docs program from the University of Vienna (J. Schörghuber). L. Geist was funded by the Christian Doppler Laboratory for High-Content Structural Biology and Biotechnology, Austria. The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ahlner A, Andresen C, Khan SN, Kay LE, Lundström P (2015) Fractional enrichment of proteins using [2- ^{13}C]-glycerol as the carbon source facilitates measurement of excited state $^{13}\text{C}\alpha$ chemical shifts with improved sensitivity. *J Biomol NMR* 62:341–351
- Andrews FH, Singh AR, Joshi S, Smith CA, Morales GA, Garlich JR, Durden DL, Kutateladze TG (2017) Dual-activity PI3K-BRD4 inhibitor for the orthogonal inhibition of MYC to block tumor growth and metastasis. *Proc Natl Acad Sci USA* 114:E1072–E1080
- Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. *Chem Rev* 106:3055–3079
- Campbell ID (2013) The evolution of protein NMR. *Biomed Spectrosc Imaging* 2:245–264
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–93
- Dougherty DA (1996) Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science* 271:163–168
- Goddard TD, Kneller DG, SPARKY 3. University of California, San Francisco
- Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10:585–592
- Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* 13:369–374
- Guo C, Geng C, Tugarinov V (2009) Selective backbone labeling of proteins using {1,2- $^{13}\text{C}_2$ }-pyruvate as carbon source. *J Biomol NMR* 44:167–173
- Hijji YM, Benjamin E (2003) Microwave assisted efficient conversion of anhydrides to cyclic imides and N-methoxy imides. 7th International Electronic Conference on Synthetic Organic Chemistry (ECSOC-7)
- Hoogstraten CG, Johnson JE Jr (2008) Metabolic labeling: taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concepts Magn Reson A* 32A:34–55
- Kafka S, Proisl K, Kašpárková V, Urankar D, Kimmel R, Košmrlj J (2013) Oxidative ring opening of 3-hydroxyquinoline-2,4(*H*,*3H*)-diones into *N*-(α -ketoacyl)anthranilic acids. *Tetrahedron* 69:10826–10835
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57
- Kasinath V, Valentine KG, Wand AJ (2013) A ^{13}C labeling strategy reveals a range of aromatic side chain motion in calmodulin. *J Am Chem Soc* 135:9560–9563
- Kasinath V, Fu Y, Sharp KA, Wand AJ (2015) A sharp thermal transition of fast aromatic-ring dynamics in ubiquitin. *Angew Chem Int Ed* 54:102–107
- Kelly MJS, Krieger C, Ball LJ, Yu YH, Richter G, Schmieder P, Bacher A, Oschkinat H (1999) Application of amino acid type-specific ^1H and ^{14}N -labeling in a ^2H -, ^{15}N -labeled background to a 47 kDa homodimer: potential for NMR structure determination of large proteins. *J Biomol NMR* 14:79–83
- Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J (2015) Methyl-specific isotope labeling: a molecular tool box for solution NMR studies of large proteins. *Curr Opin Struct Biol* 32:113–122
- LeMaster DM, Kushlan DM (1996) Dynamical mapping of *E. coli* thioredoxin via ^{13}C NMR relaxation analysis. *J Am Chem Soc* 118:9255–9264
- Lian L-Y, Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog Nucl Magn Reson Spectrosc* 39:171–190
- Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R (2004) Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing. *J Am Chem Soc* 126:5348–5349
- Lichtenecker RJ, Coudevylle N, Konrat R, Schmid W (2013a) Selective isotope labelling of leucine residues by using α -ketoacid precursor compounds. *ChemBioChem* 14:818–821
- Lichtenecker RJ, Weinhäupl K, Schmid W, Konrat R (2013b) α -Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J Biomol NMR* 57:327–331

- Liu Z, Yuan Q, Wang W (2009) Biosynthesis of [1-¹⁵N] L-tryptophan from ¹⁵N labeled anthranilic acid by fermentation of *Candida utilis* mutant. *Amino Acids* 36:71–73
- Lucas X, Wohlwend D, Hügler M, Schmidtkunz K, Gerhardt S, Schüle R, Jung M, Einsle O, Günther S (2013) 4-Acyl pyrroles: mimicking acetylated lysines in histone code reading. *Angew Chem Int Ed* 52:14055–14059
- Lundström P, Teilum K, Carstensen T, Bezsonova I, Wiesner S, Hansen DF, Religa TL, Akke M, Kay LE (2007) Fractional ¹³C enrichment of isolated carbons using [1-¹³C]- or [2-¹³C]-glucose facilitates the accurate measurement of dynamics at backbone C α and side-chain methyl positions in proteins. *J Biomol NMR* 38:199–212
- Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* 20:71–75
- Milbradt AG, Arthanari H, Takeuchi K, Boeszoermyeni A, Hagn F, Wagner G (2015) Increased resolution of aromatic cross peaks using alternate ¹³C labeling and TROSY. *J Biomol NMR* 62:291–301
- Miyanoiri Y, Takeda M, Jee J, Ono AM, Okuma K, Terauchi T, Kainosho M (2011) Alternative SAIL-Trp for robust aromatic signal assignment and determination of the χ^2 conformation by intra-residue NOEs. *J Biomol NMR* 51:425–435
- Moreira IS, Martins JM, Ramos RM, Fernandes PA, Ramos MJ (2013) Understanding the importance of the aromatic amino-acid residues as hot-spots. *Biochim Biophys Acta* 1834:404–414
- Musliner WJ, Gates JW Jr (1966) The replacement of phenolic hydroxyl groups by hydrogen. *J Am Chem Soc* 88:4271–4273
- Ohki S, Kainosho M (2008) Stable isotope labelling methods for protein NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* 53:208–226
- Rahman MM, Muhseen ZT, Junaid M, Zhang H (2015) The aromatic stacking interactions between proteins and their macromolecular ligands. *Curr Protein Pept Sci* 16:502–512
- Rajesh S, Nietlispach D, Nakayama H, Takio K, Laue ED, Shibata T, Ito Y (2003) A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp. *J Biomol NMR* 27:81–86
- Rodriguez-Mias R, 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
- Sathyamoorthy B, Singarapu KK, Garcia AE, Szyperski T (2013) Protein conformational space populated in solution probed with aromatic residual dipolar ¹³C–¹H couplings. *ChemBioChem* 14:684–688
- Schörghuber J, Sara T, Bisaccia M, Schmid W, Konrat R, Lichtenecker RJ (2015) Novel approaches in selective tryptophan isotope labeling by using *Escherichia coli* overexpression media. *ChemBioChem* 16:746–751
- Schörghuber J, Geist L, Platzer G, Konrat R, Lichtenecker RJ (2017) Highly selective stable isotope labeling of histidine residues by using a novel precursor in *E. coli*-based overexpression systems. *ChemBioChem* 18:1487–1491
- Staunton D, Schlickert R, Zanetti G, Colebrook SA, Campbell ID (2006) Cell-free expression and selective isotope labelling in protein NMR. *Magn Reson Chem* 44:S2–S9
- Takeda M, Ono AM, Terauchi T, Kainosho M (2010) Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. *J Biomol NMR* 46:45–49
- Teilum K, Brath U, Lundström P, Akke M (2006) Biosynthetic ¹³C labeling of aromatic side chains in proteins for NMR relaxation measurements. *J Am Chem Soc* 128:2506–2507
- Torizawa T, Ono AM, Terauchi T, Kainosho M (2005) NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. *J Am Chem Soc* 127:12620–12626
- Van Liemt WBS, Beijersbergen van Henegouwen WG, van Rijn A, Lugtenburg J (1996) Synthesis and spectroscopic characterization of ¹³C-labelled anthranilic acid. *Recl Trav Chim Pays Bas* 115:431–437
- Viswanatha V, Hruby VJ (1979) Synthesis of [3',5'-¹³C₂]tyrosine and its use in the synthesis of specifically labeled tyrosine analogs of oxytocin and arginine vasopressin and their 2-D-tyrosine diastereoisomers. *J Org Chem* 44:2892–2896
- Vuister GW, Kim SJ, Wu C, Bax A (1994) 2D and 3D NMR-study of phenylalanine residues in proteins by reverse isotope labeling. *J Am Chem Soc* 116:9206–9210
- Wand AJ, Bieber RJ, Urbauer JL, McEvoy RP, Gan ZH (1995) Carbon relaxation in randomly fractionally ¹³C-enriched proteins. *J Magn Reson Ser B* 108:173–175
- Wang H, Janowick DA, Schkeryantz JM, Liu XH, Fesik SW (1999) A method for assigning phenylalanines in proteins. *J Am Chem Soc* 121:1611–1612
- Weininger U (2017) Site-selective ¹³C labeling of proteins using erythrose. *J Biomol NMR* 67:191–200
- Weininger U, Diehl C, Akke M (2012a) ¹³C relaxation experiments for aromatic side chains employing longitudinal- and transverse-relaxation optimized NMR spectroscopy. *J Biomol NMR* 53:181–190
- Weininger U, Respondek M, Akke M (2012b) Conformational exchange of aromatic side chains characterized by L-optimized TROSY-selected ¹³C CPMG relaxation dispersion. *J Biomol NMR* 54:9–14
- Weininger U, Brath U, Modig K, Teilum K, Akke M (2014b) Off-resonance rotating-frame relaxation dispersion experiment for ¹³C in aromatic side chains using L-optimized TROSY selection. *J Biomol NMR* 59:23–29
- Weininger U, Modig K, Akke M (2014a) Rings flips revisited: ¹³C relaxation dispersion measurements of aromatic side chain dynamics and activation barriers in basic pancreatic trypsin inhibitor. *Biochemistry* 53:4519–4525
- Wüthrich K, Wagner G (1975) NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor. *FEBS Lett* 50:265–268
- Zhang H, van Ingen H (2016) Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Curr Opin Struct Biol* 38:75–82
- Zhang G, Smith SG, Zhou M-M (2015) Discovery of chemical inhibitors of human bromodomains. *Chem Rev* 115:11625–11668