J Biomol NMR (2012) 52:197–210
DOI 10.1007/s10858-012-9601-1

ARTICLE

Combinatorial triple-selective labeling as a tool to assist membrane protein backbone resonance assignment

Frank Löhr · Sina Reckel · Mikhail Karbyshev · Peter J. Connolly · Norzehan Abdul-Manan · Frank Bernhard · Jonathan M. Moore · Volker Dötsch

Received: 20 December 2011 / Accepted: 28 December 2011 / Published online: 18 January 2012

Abstract Obtaining NMR assignments for slowly tumbling molecules such as detergent-solubilized membrane proteins is often compromised by low sensitivity as well as spectral overlap. Both problems can be addressed by amino-acid specific isotope labeling in conjunction with ^{15}N–^{1}H correlation experiments. In this work an extended combinatorial selective in vitro labeling scheme is proposed that seeks to reduce the number of samples required for assignment. Including three different species of amino acids in each sample, ^{15}N, 1-^{13}C, and fully ^{13}C/^{15}N labeled, permits identification of more amino acid types and sequential pairs than would be possible with previously published combinatorial methods. The new protocol involves recording of up to five 2D triple-resonance experiments to distinguish the various isotopomeric dipeptide species. The pattern of backbone NH cross peaks in this series of spectra adds a new dimension to the combinatorial grid, which otherwise mostly relies on comparison of [^{15}N, ^{1}H]–HSQC and possibly 2D HN(CO) spectra of samples with different labeled amino acid compositions. Application to two α-helical membrane proteins shows that using no more than three samples information can be accumulated such that backbone assignments can be completed solely based on 3D HNCA/HN(CO)CA experiments. Alternatively, in the case of severe signal overlap in certain regions of the standard suite of triple-resonance spectra acquired on uniformly labeled protein, or missing signals due to a lack of efficiency of 3D experiments, the remaining gaps can be filled.

Keywords BEST–TROSY · Cell-free expression · Isotope labeling · Proteorhodopsin · Triple-resonance NMR · Voltage-sensor domain

F. Löhr · S. Reckel · M. Karbyshev · F. Bernhard · V. Dötsch (✉)
Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
e-mail: vdoetsch@em.uni-frankfurt.de

P. J. Connolly · N. Abdul-Manan · J. M. Moore
Vertex Pharmaceuticals Inc., Cambridge, MA 02139, USA

Introduction

Backbone resonance assignments form the basis of NMR structural investigations of proteins, especially for characterization of their interactions with binding partners and small molecules. For membrane proteins, this is a challenging task as the size contribution of the hydrophobic environment to the overall complex size leads to broad linewidths in the NMR spectrum, and consequently low sensitivity in experiments that are crucial for the assignment process (Sanders and Sönnichsen 2006; Tamm and Liang 2006; Kim et al. 2009). The choice of amenable hydrophobic environments is therefore restricted, and existing solution NMR structures have been solved only in detergent micelles (Nietlispach and Gautier 2011). Small isotropic lipid assemblies such as bicelles and nanodiscs are, however, gaining importance as the detergent micelle cannot always fulfil the folding requirements of the MP, leading to decreased structural stability and eventually loss of function (Raschle et al. 2010). These lipid assemblies have thus far exclusively been used as a reference medium to validate the structure in the detergent micelle (Raschle et al. 2009; Gautier et al. 2010). An additional aspect, in particular when working with α-helical membrane proteins, is the clustering of resonances in certain regions of the spectrum which impedes the identification of individual resonances. Slow tumbling and signal overlap thus complicate the assignment

of the membrane protein backbone to such an extent that ambiguities cannot be resolved based on uniform labeling strategies alone. For membrane proteins, some of the most informative triple-resonance experiments are often not sensitive enough, and deuteration, as standardly used in solution NMR, is often not sufficient to fully overcome these difficulties. Selective labeling strategies have therefore been suggested to assist the backbone assignment of large complexes and in particular membrane proteins (Trbovic et al. 2005; Reckel et al. 2008; Maslennikov et al. 2010; Sobhanifar et al. 2010).

Previous work by our group has shown that partial backbone assignments from 3D triple-resonance spectra can be completed using information from a combinatorial dual-selective labeling protocol (Trbovic et al. 2005). The latter involves simultaneous labeling of certain amino acid types with ^{15}N and 1-^{13}C and recording of [^{15}N, ^{1}H]–HSQC and 2D HN(CO) spectra for each of three samples containing different labeling patterns. It enables efficient amino acid-type identification in the same manner as combinatorial selective ^{15}N labeling (Wu et al. 2006) or unlabeling (Shortle 1994), but additionally permits selective detection of specific amino acid pairs based on inter-residual $^{1}J_{NC'}$ scalar interactions, as exploited in dual amino acid-selective 1-^{13}C/^{15}N labeling methods (Kainosho and Tsuji 1982; Ikura et al. 1990b; Yabuki et al. 1998; Weigelt et al. 2002). The most prominent feature of combinatorial selective labeling is the reduction of the number of samples required compared to selective ^{15}N labeling of single amino acid types or dual-selective labeling of specific ^{13}C–^{15}N pairs in individual samples. Conversely, increasing the number of samples and the number of labeled amino acids per sample may eventually lead to identification of all amino acid types (Wu et al. 2006) and unique sequential pairs (Parker et al. 2004; Hefke et al. 2010). Recently it was demonstrated that, of the ^{15}N–^{1}H cross peaks of a 185-residue membrane protein containing four transmembrane helices, 74% (22%) can be assigned amino acid type- (sequence-) specifically using an expanded seven-sample combinatorial dual-selective 1-^{13}C, ^{15}N labeling scheme (Maslennikov et al. 2010).

Here we pursue a different approach, termed combinatorial triple-selective labeling, to enhance the extractable information while employing only three samples. The advantages of combinatorial triple-selective labeling are two-fold: first, more amino acid types can be included for a given number of samples and second, sequential pairs involving an N-terminal fully labeled amino acid are directly identified for each sample independently. As also described for the dual-selective labeling scheme, the use of a cell-free expression system greatly facilitates this amino acid selective labeling strategy. The cell-free system offers reduced isotope scrambling, and only low amounts of isotopically labeled amino acids are required (Ozawa et al. 2006; Staunton et al. 2006; Sobhanifar et al. 2010). Additionally, the sample preparation is often faster as compared to in vivo methods. In this study, we demonstrate this approach for the assignment of two α-helical membrane proteins and evaluate the impact at different stages of the assignment process. In the case of the voltage-sensor domain (VSD) of the voltage-dependent K^+ channel from *Aeropyrum pernix* (KvAP), combinatorial triple-selective labeling was employed as the initial step to fully assign the backbone. The structure of this four-helix bundle is known from solution NMR (Butterwick and MacKinnon 2010; Shenkarev et al. 2010) and X-ray crystallographic (Jiang et al. 2003) studies and NMR resonance assignments are available in two different detergent systems. In contrast, de novo assignment of the bacterial retinal binding protein proteorhodopsin (PR) was completed with the help of combinatorial selective labeling which enabled the structure determination of this heptahelical membrane protein (Reckel et al. 2011).

Materials and methods

Sample preparations

The voltage sensor domain of KvAP and the green-absorbing variant proteorhodopsin were cloned into the pIVEX2.3d vector (Roche) optimized for cell-free expression. The KvAP VSD construct consisted of 159 amino acids (17.9 kDa) comprising residues 14–160 of full-length KvAP, a C-terminal His_{10}-tag and a two-residue linker N-terminal to the His-tag. PR was cloned without the signal peptide (amino acids 21–249) together with an N-terminal Met, a His_6-tag on the C-terminus and a five residue long linker N-terminal to the His-tag, resulting in a final size of 241 amino acids, 26 kDa. Both constructs were expressed in the continuous-exchange cell-free system based on an *E. coli* S30 extract (Schwarz et al. 2007). Per NMR sample, 3 ml reaction mixture were dialyzed against 52 ml feeding mixture yielding about 3–4 mg of purified membrane protein. For production of NMR samples stable-isotope labeled amino acids (Cambridge Isotope Laboratories) were included in the 4-mM amino acid stock solution used for the cell-free reaction, replacing unlabeled amino acids. The final amino acid concentration in the cell-free reaction was 0.6 mM for the reaction mixture and 0.8 mM for the feeding mixture. The amounts of labeled amino acids varied between 3 and 10 mg and isotope costs ranged between $50 and $120 per NMR sample depending on the type of isotopically labeled amino acids.

The expression of KvAP VSD was conducted in the absence of detergent, and thus the expressed protein was

present in the insoluble fraction of the cell-free reaction. After centrifugation of the reaction mixture (13,000 rpm, 10 min), the pellet was resuspended in 2% Fos-choline12 (DPC, Anatrace) and incubated at 310 K for 1 h to allow resolubilisation. The NiNTA purification buffer (*buffer A*) included 20 mM phosphate pH 7.5 and 400 mM NaCl. Binding to the NiNTA resin (Qiagen) in the presence of 0.1% Fos-choline12 was allowed for 2 h at room temperature. A washing step was conducted prior to elution (*washing buffer*: buffer A + 25 mM imidazole, 0.1% Fos-choline12, *elution buffer*: buffer A + 375 mM imidazole, 0.1% Fos-choline12). *N*, *N*-Dimethyldodecylamine *N*-oxide (LDAO, Sigma) was added to a final concentration of 0.03% (w/v) to the elution fraction. Buffer exchange for NMR conditions was achieved on gravity flow PD-10 columns (GE Healthcare, *NMR buffer*: 25 mM NaOAc pH 5, 2 mM DTT, 0.1% Fos-choline12, 0.03% LDAO). The detergent added up during concentration to approximately 5% Fos-choline12/1.7% LDAO. For the preparation of PR NMR samples, the detergent mode of the cell-free expression system was used containing 0.4% digitonin (Sigma) mixed with diheptanoyl-phosphatidylcholine (diC7PC, Avanti Polar Lipids) in a 4:1 molar ratio. 0.6 mM *all-trans* retinal (Sigma) was included directly in the reaction mixture. Subsequent Ni-affinity purification removed impurities from the extract and allowed detergent exchange. In essence, the procedure described for KvAP VSD was used with buffers containing 0.05% n-dodecyl β-D-maltoside (DDM) in the loading and washing buffer. After the washing step, the detergent exchange was accomplished prior to elution (*exchange buffer*: buffer A + 0.1% diC7PC, *elution buffer*: buffer A + 375 mM imidazole, 0.1% diC7PC). Buffer exchange for NMR conditions was achieved on gravity flow PD-10 columns (GE Healthcare, *exchange buffer*: 25 mM NaOAc pH 5, 2 mM DTT, 0.1% diC7PC). The detergent added up during concentration to approximately 2%. Final protein concentrations of all samples were 0.35 mM in a volume of 0.3 ml.

NMR spectroscopy

All experiments were conducted at a sample temperature of 323 K on a Bruker Avance 600 MHz spectrometer equipped with a cryogenic $^{1}H\{^{13}C/^{15}N\}$ triple-resonance probe. Two-dimensional ^{15}N–^{1}H correlations resulted from HSQC (Bodenhausen and Ruben 1980) as well as HNCO (Ikura et al. 1990a), HNCA (Ikura et al. 1990a), HNCOCA (Bax and Ikura 1991), COHNCA (Szyperski et al. 1995), and double-quantum (DQ) HNCA (Nietlispach et al. 2002) triple-resonance pulse schemes, where ^{13}C evolution periods are replaced by a fixed 3-μs delay. As originally proposed (Bax and Ikura 1991), the $^{13}C'$–$^{13}C^{\alpha}$ out-and-back magnetization transfer in the HNCOCA experiment was implemented in an HMQC-like rather than the more common INEPT-based version, yielding higher sensitivity owing to the reduced number of RF pulses and shorter delay periods (Bayrhuber and Riek 2011). All pulse sequences were of the [^{15}N, ^{1}H]-TROSY (Pervushin et al. 1997; Salzmann et al. 1998) type and employed sensitivity-enhanced gradient echo/antiecho coherence selection (Kay et al. 1992; Czisch and Boelens 1998; Pervushin et al. 1998; Weigelt 1998). Acceleration of longitudinal ^{1}H relaxation between scans was achieved by the Band-selective Excitation Short-Transient (BEST) (Schanda et al. 2006; Farjon et al. 2009) technique, exclusively using shaped proton pulses centered at 8.3 ppm. In experiments on the KvAP VSD, ^{1}H acquisition was immediately followed by a ^{15}N 180° pulse in order to constructively add in the subsequent scan polarization from proton magnetization longitudinally relaxed during the pulse sequence and then transferred to ^{15}N by the single-transition-to–single-transition (ST2-PT) element (Favier and Brutscher 2011). No such ^{15}N pulse was applied in experiments on PR as the data were collected before publication of this method for sensitivity improvement.

Two-dimensional H(N)CA, H(N)CACB (Wittekind and Mueller 1993), and H(NCA)CO (Clubb et al. 1992) experiments on KvAP VSD were recorded with acquisition times (complex data points) of 49 ms (384) in the ^{1}H domain and 9.2 ms (56), 8.9 ms (86), and 20 ms (42), respectively, along the ^{13}C dimensions. Each spectrum was acquired within 8 h using a 400-ms repetition delay and 512, 352, and 640 scans per FID, respectively. Spectral widths in 3D BEST–[^{15}N, ^{1}H]–TROSY–HNCA and BEST–[^{15}N, ^{1}H]–TROSY–HN(CO)CA experiments were 28 ppm (^{13}C), 26 ppm (^{15}N), and 13 ppm (^{1}H). Time-domain data consisted of 38 × 96 × 384 points, corresponding to acquisition times of 8.9 ms (^{13}C), 60 ms (^{15}N), and 49 ms (^{1}H). Each FID was the sum of 48 transients. The repetition delay was linearly decreased from 400 to 250 ms along with t_1 (Macura 2009), yielding a reduction of the measurement time from 112 to 93 h in both experiments.

Spectral widths in 2D ^{15}N–^{1}H correlation experiments on KvAP VSD were 13 and 50 ppm in the ^{1}H and ^{15}N dimension, respectively. Acquisition times in BEST–[^{15}N, ^{1}H]–TROSY–HSQC spectra were 65 ms (t_2) and 63 ms (t_1), corresponding to 512 × 192 complex data points. The recycle delay was set to either 0.5 or 0.7 s. Semi-constant time (Logan et al. 1992; Grzesiek and Bax 1993) ^{15}N evolution was employed in BEST–[^{15}N, ^{1}H]–TROSY triple-resonance pulse sequences. Compared to the HSQC spectra those experiments were acquired with lower resolution in both dimensions (ranging from 256 to 384 data points in t_2 and from 60 to 160 data points in t_1) according to the expected number of correlations, with scan repetition delays

of 0.3 s. Typically, 16, 64, 128, 224, 288, and 1,024 transients were accumulated for each FID in (2D) HSQC, HN(CO), HN(CA), HN(COCA), (CO)HN(CA), and DQ-HN(CA) experiments, respectively, giving rise to measurement times between 1.2 and 27 h for individual spectra.

Experiments on PR were recorded with spectral widths of 11 (^{1}H) and 30 (^{15}N) ppm and acquisition times were adjusted to 27.2–45.5 ms (180–300 complex data points) along t_2 and 23.0–54.8 ms (42–100 complex data points) along t_1, depending on the required spectral resolution and sensitivity. The number of scans per FID were 32, 384, and 1,024 for HSQC, HN(CO), and HN(COCA) spectra of both samples and 768/1,024, 1,792/1,472, and 2,048/1,152 for HN(CA), (CO)HN(CA), and DQ-HN(CA) spectra of samples A/B. Using an inter-scan delay of 1 s, individual measurement times ranged from 2 h for HSQC to up to 48 h for triple-resonance experiments, except for the DQ-HN(CA) experiment on sample A which required 73 h.

Spectra processing and analysis was performed with TopSpin 3.0 software (Bruker). Where appropriate, ^{15}N time domain data were extended by 20–30% using linear prediction in order to enhance resolution. A cosine-squared window function was applied for apodization in all dimensions. Spectra were referenced with respect to internal DSS using consensus Ξ values for ^{13}C and ^{15}N (Wishart et al. 1995). For preparation of all spectral plots shown here contour levels were drawn on an exponential scale using a factor of $2^{1/2}$.

Results and discussion

Providing direct correlations between backbone amide groups and α-carbons of the same and the preceding residue in ^{13}C/^{15}N labeled proteins (Ikura et al. 1990a), the HNCA is one of the most sensitive triple-resonance experiments in solution NMR. In our experience, it can often be applied successfully to moderately slow tumbling molecules such as membrane proteins in detergent micelles even without sample deuteration. On the other hand, extensive sequential backbone assignment cannot be obtained based on the HNCA alone due to chemical shift degeneracy of $^{13}C^{\alpha}$ resonances as well as their broad line widths caused by efficient transverse relaxation. Further independent pathways are indispensible for tracing sequential connectivities. Crucial experiments such as HNCACB (Wittekind and Mueller 1993) or HN(CA)CO (Clubb et al. 1992), however, suffer from low sensitivity due to comparatively inefficient magnetization transfers. This is demonstrated in Fig. 1, where the performance of the three experiments is compared for the detergent solubilized KvAP VSD, which has an apparent MW of ≈50 kDa (Butterwick and MacKinnon 2010; Shenkarev et al. 2010). It is readily apparent that only relatively small subsets of correlations presumably corresponding to mobile regions of the protein can be observed in 2D H(N)CACB and H(NCA)CO spectra. Therefore, only very sparse sequential connectivities based on $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts would be obtainable. In the H(N)CA, however, reasonable signal intensities are obtained over a wide range of $^{1}H^{N}$ chemical shifts, suggesting that the assignment of well-structured parts is feasible, provided that additional information is available to overcome ambiguities inherent to the $^{13}C^{\alpha}$ pathway.

Pulse sequences have been devised that are suitable to distinguish amino acids based on the spin topologies of their side chains. This allows two-dimensional amino-acid type selective ^{15}N–^{1}H correlation spectra to be obtained for uniformly ^{13}C/^{15}N labeled proteins (Schubert et al. 1999). In a related approach, seven classes of amino acid topologies can be differentiated from the signs of cross peaks in a set of eight Hadamard encoded 2D triple-resonance spectra (Lescop et al. 2008; Feuerstein et al. 2011). Due to their lengths and number of coherence transfer steps such experimental schemes are most favorably applied to proteins featuring relatively long transverse relaxation times. This restriction does not apply to amino-acid specific isotope labeling methods (Ozawa et al. 2006; Staunton et al. 2006). Here we show that, despite restricting the number of samples to three, combinatorial triple-selective labeling is a rich source of anchor points to complement 3D triple-resonance experiments applied to uniformly labeled proteins.

Concept of combinatorial triple-selective labeling

The approach introduced here extends combinatorial dual-selective labeling schemes (Shi et al. 2004; Parker et al. 2004; Trbovic et al. 2005; Staunton et al. 2006; Maslennikov et al. 2010) by employing three different labeling patterns in each sample. One fully ^{13}C/^{15}N labeled amino acid type is included per sample in addition to four ^{15}N-labeled and two 1-^{13}C labeled types in a three-sample scheme. This gives rise to six possible combinations of either ^{15}N- or ^{13}C/^{15}N labeled residues at the C-terminus of a dipeptide (position i) with either unlabeled, 1-^{13}C-, or ^{13}C/^{15}N labeled residues at the N-terminus (position $i-1$), each of which can be detected in a [^{15}N, ^{1}H]-HSQC spectrum. As outlined in Fig. 2, these combinations can be distinguished by the presence/absence in five 2D HN(C)-type triple resonance spectra recorded for each sample. No distinction can be made as to whether the observed NH is preceded by an unlabeled or a ^{15}N labeled residue. Contrary to ^{15}N or 1-^{13}C labeled moieties, correlations can involve the ^{13}C/^{15}N labeled amino acid type either in position i or $i-1$. Importantly, the latter can be differentiated from the two singly labeled amino acid types by spectroscopic means owing to the additional ^{13}C label at the 2-(C^{α}) position. As a

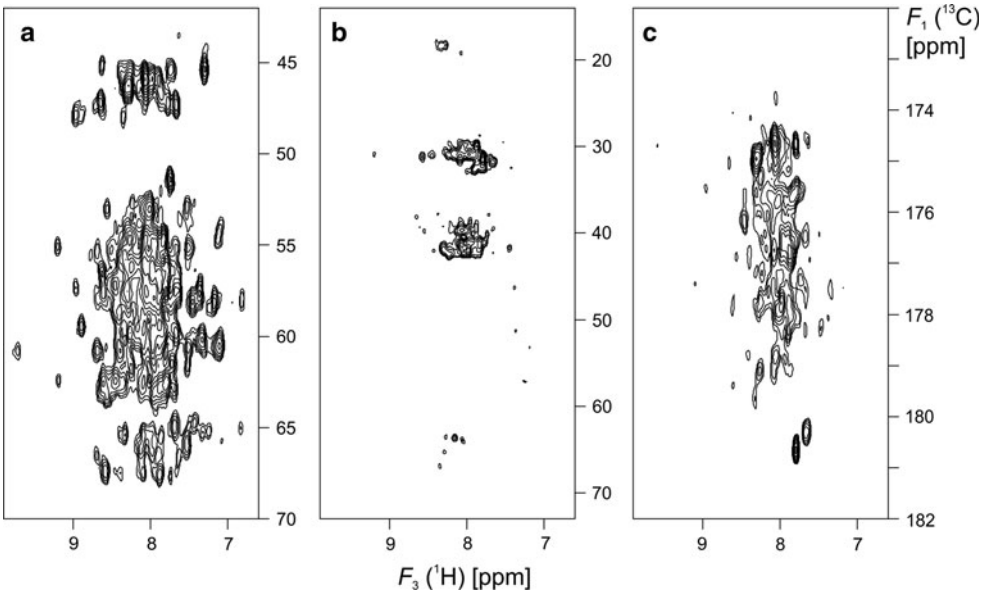


Fig. 1 Performance of standard triple-resonance experiments for sequential backbone assignment on [U–^{13}C; ^{15}N] KvAP VSD. Shown are 2D BEST–[^{15}N, ^{1}H]–TROSY–H(N)CA (**a**), –H(N)CACB (**b**), and –H(NCA)CO (**c**) spectra each recorded within 8 h on a 600 MHz spectrometer. In the H(N)CACB experiment the C^{α}–C^{β} transfer delay was adjusted to $(4\ ^1J_{C\alpha C\beta})^{-1}$ resulting in approximately equal signal intensities of H^N–C^{α} and H^N–C^{β} cross peaks, however only negative contours corresponding to the latter are plotted

consequence, more amino acid types and more sequential pairs can be identified than in previous combinatorial labeling procedures. Using combinatorial selective ^{15}N labeling (Wu et al. 2006) a maximum of 7 ($= 2^n - 1$) amino acid types are discerned from the presence/absence of cross peaks in [^{15}N, ^{1}H]–HSQC spectra for n = 3 samples: One amino acid type is present in all samples, three are unique in each sample and three are missing in one sample, but present in the other two. Application of triple-resonance pulse sequences that involve $^{13}C^{\alpha}$ nuclei allow selective detection of ^{15}N–^{1}H cross peaks of a supplementary fully ^{13}C/^{15}N labeled amino acid, thereby increasing the number of amino acid types to 10 ($= 2^n - 1 + n$). Once residue type assignments have been obtained, dipeptide sequences with the ^{13}C/^{15}N labeled amino acid on the N-terminal side are immediately identified by means of interresidual correlation between ^{15}N and $^{13}C^{\alpha}$ spins. This, in turn, leaves only two possible dipeptide sequences for the remaining peaks observed in the HN(CO) spectrum of each sample, some of which can be narrowed down to a single solution if one of the combinations occurs in the HN(CO) spectra of the respective other samples.

In order to unambiguously recognize the fully ^{13}C/^{15}N labeled amino acid type in either the N- or the C-terminal side of a dipeptide, 2D (^{15}N–^{1}H) versions of HNCA (Ikura et al. 1990a), HN(CO)CA (Bax and Ikura 1991), and DQ-HNCA (Nietlispach et al. 2002) experiments are recorded in addition to HSQC and HN(CO). Optionally, a (CO)HN(CA) (Szyperski et al. 1995) spectrum, which contains the subset of correlations simultaneously present in HN(CO) and HN(CA), can be acquired. Although this information is redundant it may be useful to confirm cross peaks that are arguable in the latter two spectra due to overlap. It should be noted that detection in the HN(CA) is a necessary but not a sufficient condition for the identification of ^{13}C/^{15}N doubly labeled residues, because the N–C^{α} magnetization transfer can arise through one-bond and two-bond scalar couplings, thus making sequentially following ^{15}N labeled residues observable in the HN(CA), too. These are, however, sorted out by the presence in HN(COCA) spectra and the absence in DQ-HN(CA) spectra, which exclusively show ^{15}N–^{1}H correlations of amides coupled to both the intraresidual and the preceding $^{13}C^{\alpha}$ spin. As an example, a ^{13}C/^{15}N labeled residue preceded by a 1-^{13}C labeled one will give rise to cross peaks in HSQC, HN(CO), HN(CA), and (CO)HN(CA) spectra. In contrast, two ^{13}C/^{15}N labeled residues sequentially following each other are recognized by the appearance of cross peaks in all spectra (Fig. 2). Since only one residue type is fully ^{13}C/^{15}N labeled in each sample this immediately identifies sequential pairs of the corresponding species. Note that pairs of like amino acid types are "invisible" in combinatorial dual selective ^{15}N/1–^{13}C labeling schemes, while in 3D HNCA spectra of uniformly labeled proteins they often cause ambiguities due to $^{13}C^{\alpha}$ chemical shift degeneracy.

Conceptually, combinatorial triple-selective labeling is related to an approach employed for deconvolution of ^{15}N–^{1}H correlation spectra of ternary mixtures consisting of uniformly ^{15}N labeled, uniformly ^{13}C/^{15}N labeled and 1-^{13}C/^{15}N or 2-^{13}C/^{15}N labeled protein (Masterson et al. 2008; Tonelli et al. 2009), where signals from either or both ^{13}C labeled species are selectively suppressed by virtue of ^{15}N–^{13}C spin-echo filtering. Also, ^{13}C/^{15}N doubly labeled amino acids have been combined with ^{15}N singly labeled ones to derive amino acid types and sequential pairs from ^{13}C chemical shifts obtained in 2D H(N)CACB and CBCA(CON)H experiments (Shi et al. 2004). This information was used to supplement a set of 3D triple resonance spectra of uniformly ^{13}C/^{15}N- and ^{2}H/^{13}C/^{15}N-enriched

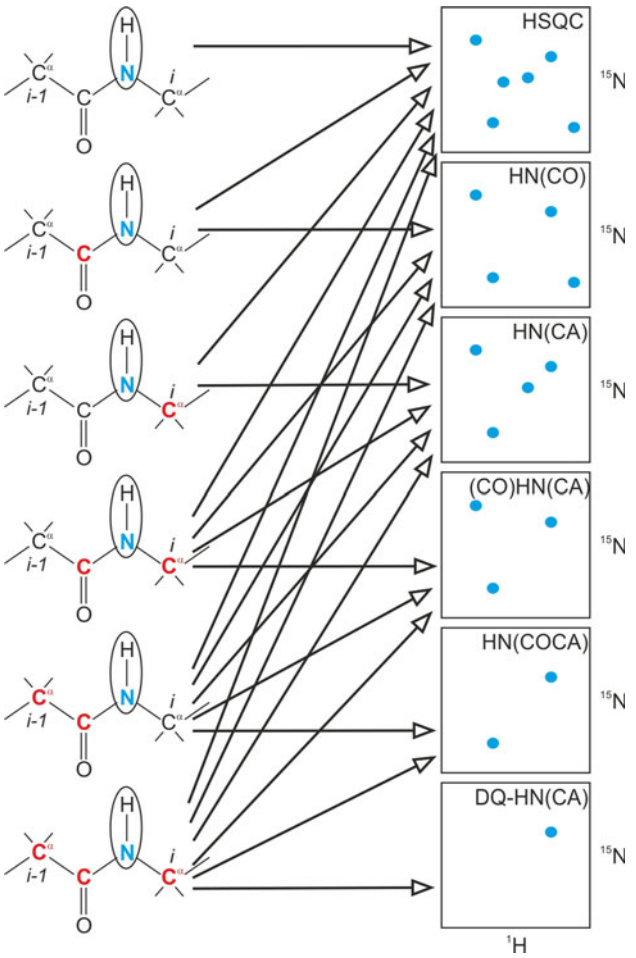


Fig. 2 Correlation between the six different combinations of labeled or unlabeled amino acids in a dipeptide and the six types of ^{15}N–^{1}H correlation experiments recorded in this study. The ^{15}N labeled amide nitrogen of the detected amide group is drawn in *blue* and ^{13}C labeled carbonyl or α-carbons are marked in *red*. The remaining carbons are in natural abundance. Each of the six isotopomers is represented schematically by a cross peak in the HSQC spectrum. Its identitiy is revealed by the pattern of occurrence in the various triple resonance spectra

samples. In a similar vein, [^{15}N, ^{1}H]–HSQC, 2D HN(CO) and 2D HN(CA) experiments on samples with different combinations of either a ^{15}N or $^{13}C/^{15}N$ doubly labeled amino acid with two 1-^{13}C or 2-^{13}C labeled amino acid type were employed to resolve ambiguities in 3D spectra of uniformly labeled samples (Butterwick and MacKinnon 2010). To our knowledge the systematic combinatorial use of three isotopic species in the backbone of the same protein as presented here has not been described so far.

Assignment of KvAP VSD with no prior information

To assess the feasibility of obtaining complete backbone assignments for a membrane protein using combinatorial triple-selective labeling supplemented with 3D HNCA data, the 149-residue KvAP voltage-sensor domain in Fos-choline12/LDAO mixed micelles was chosen as a test case. Resonance assignments for the uniformly $^{2}H/^{13}C/^{15}N$ labeled *E.coli*-expressed protein in the same detergent mixture (Shenkarev et al. 2010) and in diC7PC (Butterwick and MacKinnon 2010) have been published recently, allowing a validation of our approach. The labeling scheme devised for the three selectively labeled samples is depicted in Table 1. Each contains one uniformly $^{13}C/^{15}N$ labeled, four ^{15}N labeled and two 1-^{13}C labeled amino acid types.

The choice of amino acid types to be labeled for a particular protein is mostly governed by two aspects, spectral dispersion and liability to isotopic scrambling. Selecting the most abundant amino acid types for ^{15}N and $^{13}C/^{15}N$ labeling maximizes the information that can be obtained but may cause ambiguities due to signal overlap. Cell-free synthesis of selectively labeled proteins results in reduced scrambling of the ^{15}N label as compared to in vivo expression but requires optimization (Morita et al. 2004; Ozawa et al. 2004; Klammt et al. 2007; Tonelli et al. 2011). Depending on the expression system, interconversion between glutamate and aspartate, glutamate and asparagine, and serine and glycine is often not completely eliminated unless enzymes such as transaminases and glutamine synthetases are inhibited. Cross-labeling is less critical for carbonyl carbons (Takeuchi et al. 2007) and, as in combinatorial dual-selective labeling, the choice of the 1-^{13}C labeled amino acid types therefore primarily requires inspection of the amino acid sequence to maximize the number of unique 1-$^{13}C_{i-1}$–$^{15}N_i$ combinations while avoiding those that do not occur in the target protein. Computational methods have been developed to optimize combinatorial selective labeling schemes for a given protein

Table 1 Combinatorial triple-selective labeling scheme employed for the KvAP voltage-sensor domain

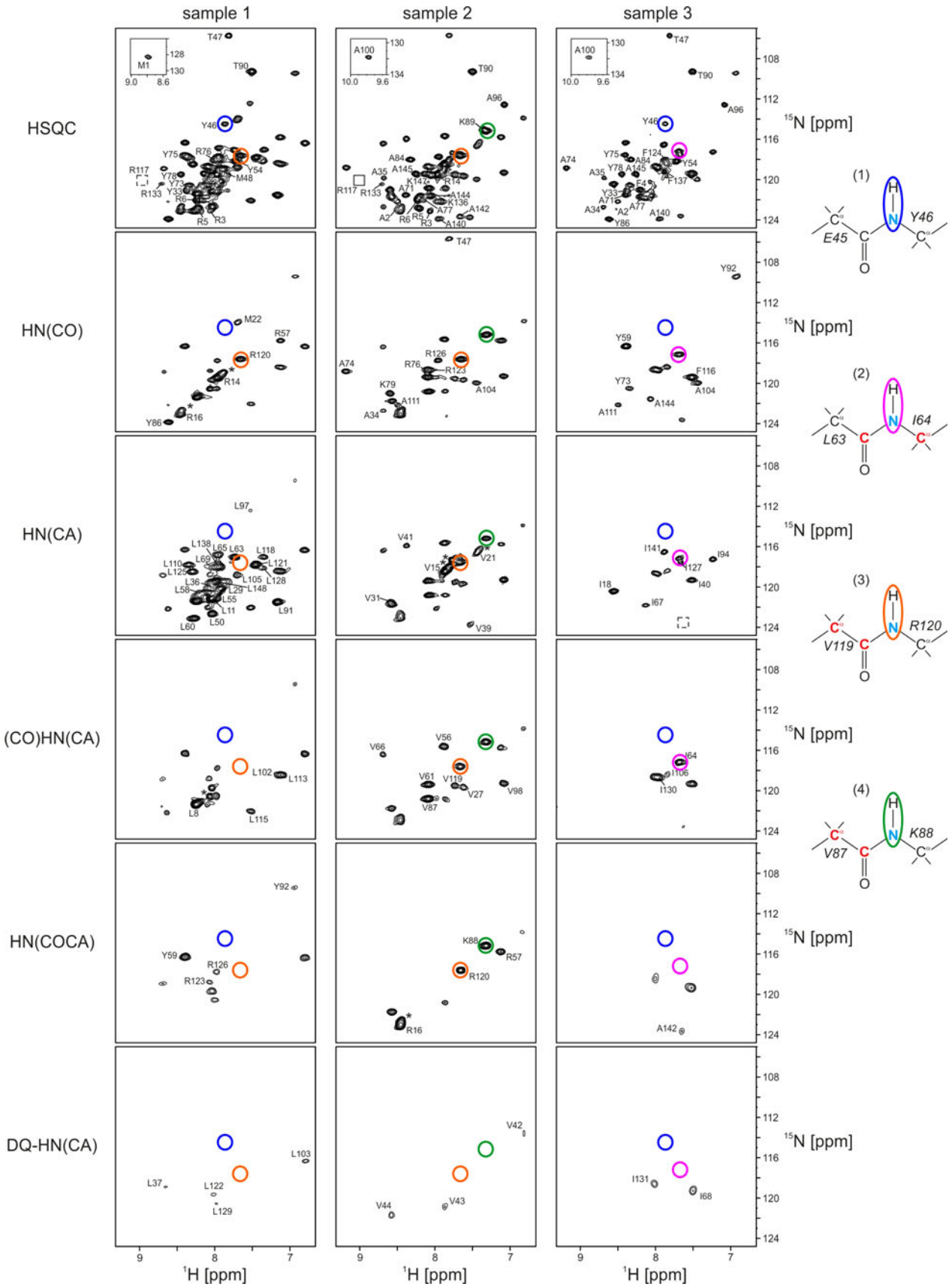
Amino acid type	Samples		
	1	2	3
Leucine	$^{13}C/^{15}N$	1-^{13}C	1-^{13}C
Valine	1-^{13}C	$^{13}C/^{15}N$	
Isoleucine			$^{13}C/^{15}N$
Methionine	^{15}N		
Lysine		^{15}N	
Phenylalanine			^{15}N
Arginine	^{15}N	^{15}N	
Tyrosine	^{15}N	1-^{13}C	^{15}N
Alanine		^{15}N	^{15}N
Threonine	^{15}N	^{15}N	^{15}N
Glycine	1-^{13}C		
Aspartate			1-^{13}C

sequence (Trbovic et al. 2005; Hefke et al. 2010; Maslennikov et al. 2010), but no such software was employed here.

In the present application, leucine, valine and isoleucine which have a high abundance in KvAP VSD were selected for $^{13}C/^{15}N$-labeling in order to maximize the number of correlations to be expected in triple resonance experiments. Of the remaining 16 non-proline amino acid types glycine, serine, asparagine, aspartic acid, glutamine and glutamic acid, which showed signs of scrambling in the cell-free expression system used in this study, were disregarded for ^{15}N labeling. Furthermore, KvAP VSD does not contain cysteines and only one tryptophan, and histidine was considered unsuitable due to the presence of a His_{10}-tag in the construct used here. This left alanine, methionine, threonine, arginine, lysine, phenylalanine and tyrosine as the seven amino acid types to be ^{15}N labeled in a three-sample combinatorial scheme. The decision for leucine, valine, tyrosine, aspartate and glycine as 1-^{13}C labeled residues (Table 1) was guided by an attempt to maximize the sequential connectivities detected in the HN(CO) experiment for the given set of ^{15}N labeled amino acids in each sample. Identification of specific sequential pairs requires the simultaneous presence of a certain ^{13}C labeled amino acid type (either 1-^{13}C or uniformly $^{13}C/^{15}N$) and a different ^{15}N labeled amino acid type in at least two samples. For instance, samples 2 and 3 contain 1-^{13}C labeled leucine and ^{15}N labeled alanine (Table 1). The dipeptide Leu–Ala gives rise to cross peaks in HN(CO) spectra of both samples. In contrast, dipeptides Tyr–Ala and Asp–Ala result in HN(CO) cross peaks exclusively in samples 2 and 3, respectively. Hence, the three dipeptides are unambiguously identified. In a similar manner, arginine preceded by glycine, leucine, valine or tyrosine can be spotted by inspection of spectra of samples 1 and 2. Note that the pairs Leu–Arg and Val–Arg can only be differentiated due to the fact that leucine and valine are fully labeled in one sample and 1-^{13}C labeled in the other such that cross peaks appear in HN(CO) and HN(COCA) spectra in the first case, but only in HN(CO) spectra in the second case. No differentiation would be possible between the dipeptides Gly–Met and Val–Met, both observed in HSQC and HN(CO) spectra of sample 1 exclusively. However, since only Val–Met occurs in the sequence of KvAP VSD, in fact only once, the ^{1}H and ^{15}N resonance frequencies of the corresponding methionine are readily assigned.

A total of 18 2D spectra were recorded on the three samples of KvAP VSD. As illustrated in Fig. 3, a virtually complete set of the expected correlations was observed. The only exceptions are residues Lys147 (sample 2) for which no HSQC cross peak could be identified, Ile40 (sample 3) which is presumably hidden by a stronger peak belonging to Ile68 (Butterwick and MacKinnon 2010; Shenkarev et al. 2010), and Leu26 (sample 1) whose amide proton resonates at 10.61 ppm, outside the bandselectively excited region. The latter signal was, however, detected in HSQC and HN(CA) spectra acquired with a more downfield shifted ^{1}H offset (not shown). Some of the spectra contain spurious broad peaks in the region around 8.0 (^{1}H) and 120 ppm (^{15}N) that are due to a small fraction of unfolded protein in our preparation. In addition residues Leu8, Arg14, Val15, Arg16, and Val21 exhibit a second (and a third in the case of Val15) set of signals, upfield-shifted in both dimensions by a varying amount. Most of these residues belong to the amphiphatic α-helix S0 which has been suggested to interact with the interfacial region of the micelle (Butterwick and MacKinnon 2010) and is not seen in the KvAP crystal structure due to a lack of electron density (Jiang et al. 2003). These signals could be sequence-specifically assigned using the combinatorial spectra shown here in conjunction with a 3D HNCA, but the origin of this heterogeneity in our samples remains not fully understood.

In the following the extraction of information from this matrix of spectra is described using four representative examples (Fig. 3): (1) Cross peaks for residue Tyr46 are observed in the HSQC spectra of samples 1 and 3 (at $^{1}H/^{15}N$ chemical shifts 7.84/114.4 ppm, blue circles), which immediately identifies its residue type. As there are no peaks at this position in any of the triple resonance spectra, no specific hint about the sequentially preceding residue can be gained, except that it cannot be preceded by amino acid types that are ^{13}C labeled in samples 1 or 3, i.e. leucine, valine, isoleucine, glycine or aspartate. Similar information would be obtained via combinatorial dual-selective labeling (Parker et al. 2004; Trbovic et al. 2005; Maslennikov et al. 2010). (2) Residue Ile64 gives rise to cross peaks (7.68/117.2 ppm, magenta circles) in spectra of sample 3 exclusively. The fact that this amide group is observed in the HN(CA) but not in the HN(COCA) spectrum proves that it belongs to a $^{13}C/^{15}N$ labeled isoleucine and not to a ^{15}N phenylalanine, which is also present in sample 3 only and might be preceded by an isoleucine. Since the peak also shows up in the HN(CO) and, hence, in the (CO)HN(CA), it must have either of the two 1-^{13}C labeled residue types on its N-terminal side. The sequence Asp-Ile does not occur in KvAP VSD, leaving Leu-Ile as the only possibility in this case. (3) The amino acid type of Arg120 (7.66/117.6 ppm, orange circles) follows from the presence of a cross peak in the HSQCs of samples 1 and 2. Signals are also observed in the corresponding HN(CO) spectra, compatible with the dipeptides Leu–Arg and Val–Arg which both occur in the protein sequence. These two possibilities can be easily distinguished by the observation of cross peaks in the HN(CA), (CO)HN(CA) and HN(COCA) spectra of sample 2 expected for Val–Arg but not for Leu–Arg. Finally, (4) Lys88 at 7.31/115.1 ppm


sample 1
sample 2
sample 3
HSQC
HN(CO)
HN(CA)
(CO)HN(CA)
HN(COCA)
DQ-HN(CA)
^{15}N [ppm]
^{1}H [ppm]
(1)
(2)
(3)
(4)
E45
Y46
L63
I64
V119
R120
V87
K88

◀ **Fig. 3** Experimental ^{15}N–^{1}H correlation spectra obtained for the three combinatorial triple-selectively labeled samples of KvAP VSD. For clarity, peak assignments are indicated only once within each series of spectra for the individual samples, and are displayed in the last spectrum where the corresponding residue is detectable. For instance, only those peaks are annotated in the HSQC that are not present in any of the triple-resonance spectra. The *insets* in the *upper left corner* of the HSQCs show cross peaks located outside the plotted region. *Empty dashed boxes* indicate the positions of peaks observed at lower contour levels. *Asterisks* mark additional signals belonging to residues Leu8, Arg14, Val15, Arg16, and Val21. Resonance positions of Tyr46, Ile64, Lys88, and Arg120 discussed in the text, are highlighted by *colored circles*. The corresponding dipeptide labeling patterns are shown on the *right-hand side*. Note that the cross peaks of Ile64 and Lys88 partially overlap with those of Ile127 (HSQC and HN(CA)) and Lys89 (HSQC), respectively

(green circles) is detected only in sample 2 which is in accordance with either ^{15}N labeled lysine or fully ^{13}C/^{15}N labeled valine. Because the cross peak is present in all triple resonance spectra of this sample except for the DQ-HN(CA) it must be due to the ^{15}N labeled amino acid preceded by a fully ^{13}C/^{15}N labeled one rather than the fully ^{13}C/^{15}N labeled amino acid itself. The dipeptide Val–Lys is unique in the sequence of KvAP VSD resulting in the unambiguous sequential assignment of Lys88.

The overall amount of information provided by combinatorial triple-selective labeling of KvAP VSD is summarized in Fig. 4a. Of the 145 non-proline residues (excluding the His_{10} tag) the amino acid type could be determined for 103 residues (71%). Considering that glycines can often be identified already from their highfield ^{15}N chemical shift in a [^{15}N, ^{1}H]-HSQC and, more reliably, from their highfield $^{13}C^{\alpha}$ chemical shift in a 3D HNCA, the number of residues for which the amino acid type remains undetermined reduces to 28 (19%). Sequential information is obtained for 45 sites (31%), where 38 pairs are determined unambiguously, while 7 pairs are two-fold degenerate because Leu–Val and Tyr–Val cannot be distinguished with the labeling scheme chosen here. Among the non-degenerate pairs 12 (8%) are unique in the sequence of KvAP VSD, meaning that the C-terminal residue of each dipeptide is assigned sequence-specifically. These are then, as the next step, used as anchor points during analysis of 3D HNCA and 3D HN(CO)CA spectra.

Connecting anchor points in KvAP VSD

An example for the combinatorial-labeling assisted sequential assignment of KvAP VSD is given in Fig. 4b. It shows strips from the 3D HNCA spectrum for residues Trp70 through Asp82. In some cases (Trp70, Asp72, Tyr75, Tyr78, Lys79) the interresidual correlation was not observed, requiring the combined use of HNCA and HN(CO)CA spectra. In principle

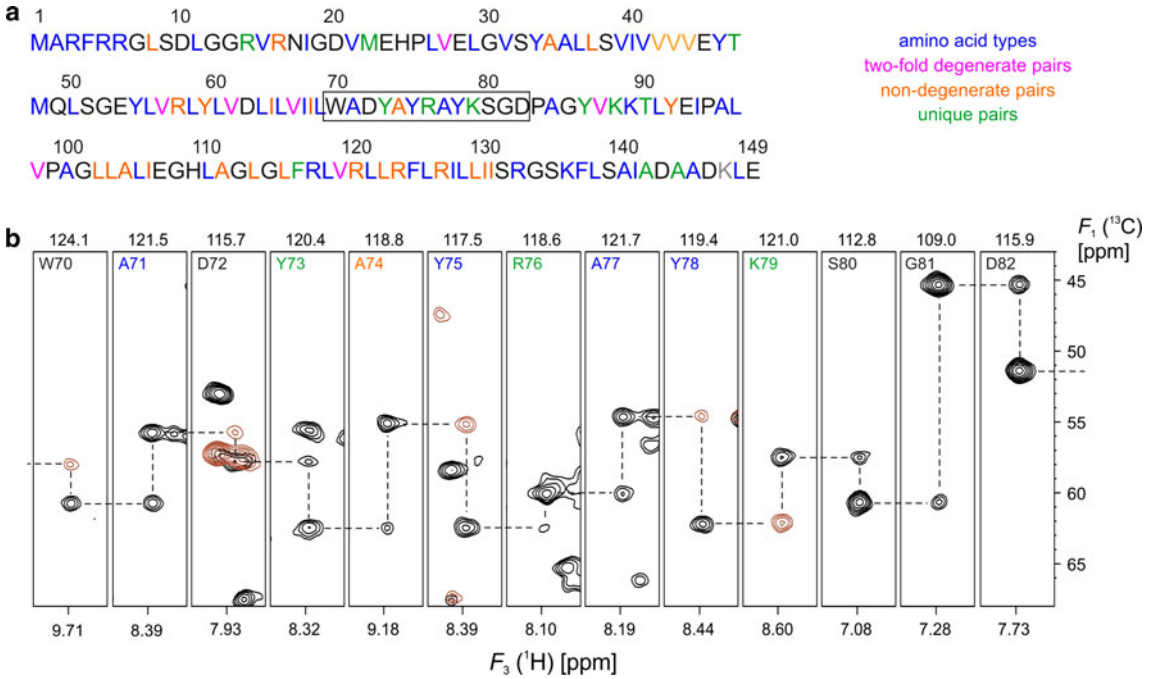


Fig. 4 Sequential assignment of KvAP VSD. **a** The amino acid sequence is shown with those residues highlighted for which information is obtained from combinatorial triple-selective labeling. The *color coding* is indicated adjacent to the sequence. The stretch of residues selected in (**b**) is highlighted in a *box*. Residues that were not observed experimentally are printed in *grey*. **b** F_1 ($^{13}C^{\alpha}$)–F_3 ($^{1}H^{N}$) strips from a 3D HNCA spectrum (*black contours*) taken at F_2 (^{15}N) chemical shifts indicated above each strip and centered (0.3 ppm width) at the ^{1}H chemical shifts given at the bottom. The *dashed line* indicates the sequential walk through residues Trp70 to Asp82. For residues where the sequential correlation in the HNCA is too weak, the corresponding strip from a 3D HN(CO)CA (*brown contours*) is superimposed

a complete "sequential walk" through this segment would be possible, indicating that the spectral quality obtained for a non-deuterated 50-kDa protein-detergent complex is sufficient. However, a closer look at some of the strips (Tyr73, Ala74, Tyr75, Tyr78), which all contain pairs of cross peaks at very similar position, reveals potential problems. Typical for an α-helical membrane protein, there are many more instances in the sequence of KvAP VSD where $^{13}C^{\alpha}$ chemical shift degeneracies make it very difficult—if not impossible—to find the correct sequential connectivities without prior knowledge of at least individual residue types or preferentially residue types of sequential pairs. Among the tyrosine residues of the stretch shown in Fig. 4b, Tyr73 could already be unambiguously placed into the sequence based on the combinatorial labeling results whereas the residue type was known for Tyr75 and Tyr78. The latter two could be distinguished based on a small difference of their $^{13}C^{\alpha}$ chemical shifts which were known from the interresidual cross peaks of Arg76 and Lys79, both sequentially assigned by combinatorial labeling. Similarly, it was known that the correlation at 9.18/118.8 ppm belongs to an alanine preceded by a tyrosine. This dipeptide occurs twice in KvAP VSD, but the alternative assignment could be ruled out by a mismatch of the $^{13}C^{\alpha}(i-1)$ chemical shift with the $^{13}C^{\alpha}(i)$ chemical shift of the Tyr73 strip. No information from combinatorial labeling was available for the three rightmost strips (Ser80, Gly81, Asp82), but they could be readily linked together and appended to Lys79. It is apparent from Fig. 4a that, together with Ser51 to Glu53 and Glu107 to His109 this represents the largest stretch lacking residue type or sequential information from combinatorial triple-selective labeling.

Overall, $^{1}H^{N}$, ^{15}N, and $^{13}C^{\alpha}$ chemical shifts obtained from combinatorial triple-selective labeling and 3D HNCA/HN(CO)CA spectra on uniformly labeled KvAP VSD are in reasonable agreement with published values at 42°C and at pH 4.7 (Shenkarev et al. 2010). Proton and nitrogen chemical shifts usually match within 0.1 and 0.6 ppm, respectively, and $^{13}C^{\alpha}$ chemical shifts differ systematically by about 1 ppm due to $^{2}H/^{1}H$ isotope shifts since the previous assignments were obtained for a per-deuterated protein. Significantly different resonance frequencies were only found for residues Ile141 through Asp146, presumably because in our construct the two C-terminal residues (Arg148, Ser149) were substituted and a His_{10}-tag attached. Signals for residues Met1, Ala2, and Arg3, previously unassigned, were identified here with the help of combinatorial triple-selective labeling.

Resolving signal overlap in proteorhodopsin

In the implementation described above amino-acid specific labeling was employed as the initial step in the sequential assignment process. Information attained about the majority of the residues then enabled signal assignment of 3D HNCA/HN(CO)CA spectra of the uniformly labeled variant of the protein. However, the reverse scenario is also conceivable. For α-helical membrane proteins with a limited chemical shift dispersion the conventional 3D triple-resonance approach often yields incomplete backbone assignment even with deuterated protein samples, where sequential connectivities via $^{13}C^{\beta}$ and ^{13}CO spins are available. Reasons for this may be signal overlap, fast transverse relaxation of nuclear spins in certain regions of the protein and incomplete back-exchange of deuterated amide groups. For such cases dual amino-acid specific labeling in combination with [^{15}N, ^{1}H]–HSQC and 2D triple-resonance experiments has been shown to provide valuable information to promote the degree of assignment and resolve ambiguities (Shi et al. 2004; Trbovic et al. 2005; Butterwick and MacKinnon 2010). In the following, we outline how triple amino-acid specifically labeled samples contributed to the near complete backbone resonance assignment of the 241-residue retinal-binding integral membrane protein proteorhodopsin, which was essential for its solution structure determination (Reckel et al. 2011). Using [^{15}N, ^{1}H]–TRACT (Lee et al. 2006) a rotational correlation time of 24 ns at 50°C was determined for this seven-transmembrane-helix protein in diC7PC micelles, corresponding to a 70-kDa complex. Standard triple-resonance and ^{15}N-separated NOESY experiments on [U–^{13}C; ^{15}N] and [U–^{2}H; ^{13}C; ^{15}N] labeled samples yielded unambiguous assignments only for about 80% of its residues as the analysis was complicated either due to overlap in the ^{15}N–^{1}H plane or because of missing correlations in the 3D spectra. Both problems were compounded by conformational exchange line broadening in particular for residues in the retinal-binding pocket, entailing a decrease of signal intensities with increasing B_0 field (Reckel et al. 2011). On the other hand, the utility of 2D triple-resonance spectra recorded for amino-acid specifically labeled samples as proposed here does not critically depend on high static magnetic field strengths. For this purpose, two samples were prepared with labeled amino-acid type compositions that allowed identification of as many of the unassigned amide signals as possible. Sample A contained $^{13}C/^{15}N$ Ile, ^{15}N Tyr, Trp, and 1-^{13}C Gly, Leu while sample B contained $^{13}C/^{15}N$ Phe, ^{15}N Ile, Lys, Tyr, and 1-^{13}C Ala, Met.

The spectral simplification achieved by amino-acid specific labeling becomes apparent in Fig. 5 (top) where the most congested region of a [^{15}N, ^{1}H]-HSQC spectrum of uniformly labeled PR is superimposed with those of samples A and B. Using only two samples the potential of combinatorial triple-selective labeling cannot be fully exploited. Nevertheless, as illustrated in the following for residues Ile112 and Lys125, the labeling scheme allows one to derive information from the intra-sample as well as inter-sample pattern of cross peak occurrence. Assignment

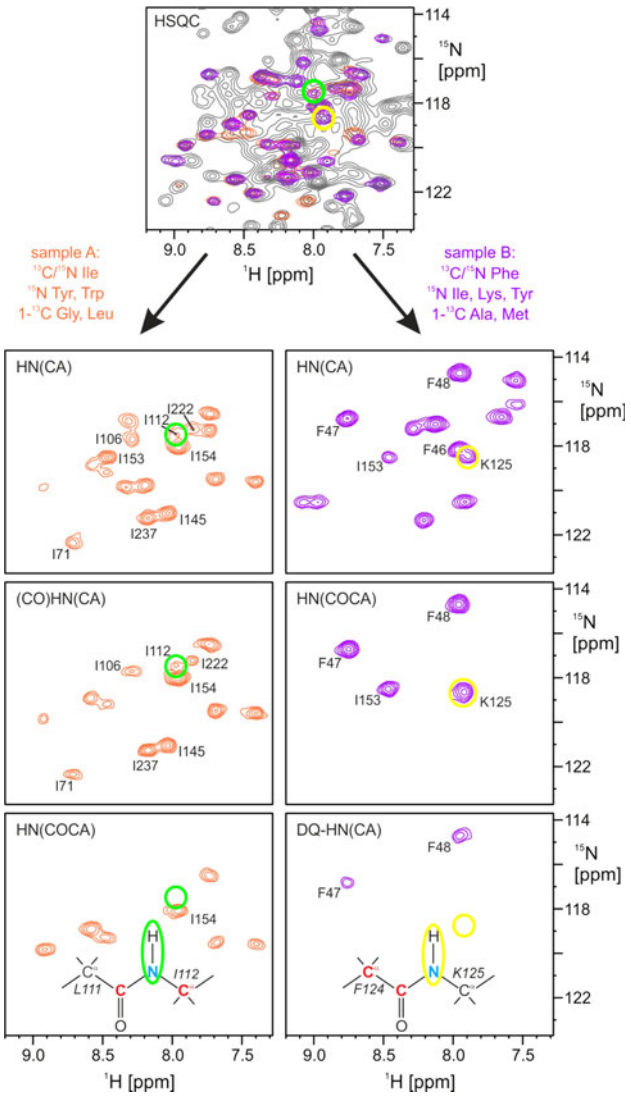


Fig. 5 Assignment of the NH resonances of Ile112 and Lys125 of proteorhodopsin with the help of two triple-selectively labeled samples. The top panel is an overlay of [^{15}N, ^{1}H]–TROSY–HSQC spectra of [U–^{15}N] labeled (*grey contours*) protein and amino-acid specific labeled protein samples A (*orange contours*) and B (*purple contours*). Remaining panels on the *left* and *right* are expansions from 2D triple-resonance spectra of the amino-acid specific labeled samples showing the same spectral region. Assignments are indicated only for residues mentioned in the text. Cross peaks of residues Ile112 and Lys125 are marked with *colored circles*. Labeling patterns of the respective dipeptide fragments are shown in the *bottom panels*

of the two residues in the uniformly labeled protein was hampered by near degeneracy of their $^{1}H^{N}$ and ^{15}N chemical shifts with those of Glu85 (for Ile112) and with Asn176 (for Lys125), and partial overlap of both signals with Phe46, Ile154, and Glu245. Low intensities or even absence of cross peaks in 3D triple resonance spectra, in particular for Ile112, posed additional problems. Cross peaks for Ile112 were however observed in [^{15}N, ^{1}H]-HSQC of both amino-acid specifically labeled samples, which would be expected for either a $^{13}C/^{15}N$ labeled isoleucine or a ^{15}N labeled tyrosine residue. Presence of a cross peak in the HNCA of sample A would also be in agreement with an isoleucine or a tyrosine preceded by an isoleucine but the absence in the HN(COCA) rules out the second option. Since it is additionally present in the HN(CO) (not shown) and the (CO)HN(CA), this isoleucine residue must be preceded either by a glycine or a leucine which are both 1-^{13}C labeled in sample A. The same pattern is observed for isoleucine residues 71, 106, 145, 222 and 237, but these were already assigned from 3D experiments applied to uniformy labeled PR, leaving Ile112 as the only possibility. In the second example, residue Lys125 is part of a unique dipeptide and could therefore be assigned exclusively based on the 2D spectra of the amino-acid specifically labeled samples. It is observed in the HSQC as well as HN(CO) (not shown), HN(CA), (CO)HN(CA) (not shown), and HN(COCA) spectra of sample B, which is also the case for Phe47, Phe48, and Ile153. The two phenylalanine signals are immediately identified owing to their presence in the DQ-HN(CA) spectrum, where only the NH of the second residue in Phe–Phe pairs is detected, and Ile153 can be ruled out because the signal also occurs in spectra of sample A. Hence, the assignment of the fourth peak in the HN(COCA) as Lys125 was established unequivocally. By avoiding extensive overlap encountered in spectra of uniformly labeled PR, the use of amino-acid specifically labeled samples in combination with 2D triple-resonance experiments allowed to raise the amount of backbone amide assignments to 96% and to confirm several tentative assignments from conventional 3D spectra.

Sensitivity considerations

The combinatorial triple-selective labeling method proposed here relies on the recording of several 2D $^{1}H/^{15}N/^{13}C$ triple-resonance experiments to assign amino acid types and sequential pairs to ^{15}N–^{1}H cross peaks. Compared to their 3D counterparts, these experiments have a clear sensitivity advantage arising from three major sources: (1) Dispensing with the ^{13}C dimension, as opposed to the echo/antiecho type ^{15}N dimension, yields a gain of a factor of $2^{1/2}$. (2) Transverse ^{13}C relaxation which leads to a significant loss of magnetization in larger proteins is avoided. (3) Pulse sequences can be simplified, as decoupling pulses during carbon evolution and possibly ^{13}C 180° pulses required to refocus chemical shift evolution during the decoupling pulses in the first ^{13}C increment of the 3D sequences are not needed. In particular, we found that omission of the pair of ^{1}H 180° pulses usually applied in the ^{13}C evolution period of TROSY-type experiments enhances sensitivity by 10–20% on our spectrometer. Taken together, the sensitivity of, e.g., 2D HN(CA)

experiments on combinatorial labeled samples was sufficient to observe all expected interresidual correlations between ^{15}N-labeled and fully ^{13}C/^{15}N labeled amino acids, despite the small size of the active $^{2}J_{NC\alpha}$ coupling typically found in α-helical conformations (Delaglio et al. 1991; Wirmer and Schwalbe 2002; Schmidt et al. 2010). In contrast, several sequential cross peaks were not detected in the 3D HNCA applied to uniformly labeled KvAP VSD at the same concentration although the measurement time was significantly longer. It should however be noted that in uniformly labeled proteins the efficiency of the magnetization transfer to α-carbons of the preceding residue is reduced by passive $^{1}J_{NC\alpha}$ couplings.

All pulse sequences employed in the current study take advantage of the BEST (Schanda et al. 2006; Farjon et al. 2009; Favier and Brutscher 2011) technique to enhance longitudinal relaxation rates of amide protons, allowing the use of short scan repetition times. Differing from its original intention of recording "fast" experiments, BEST was applied here to increase signal-to-noise ratios by enabling more scans in the same time. In all experiments, including the [^{15}N, ^{1}H]–TROSY–HSQCs, experimental times were determined by sensitivity rather than by sampling issues. This implies that under more favorable conditions (narrow linewidths, high sample concentrations, higher magnetic field) measurement time demands of combinatorial triple-selective labeling are moderate. For KvAP VSD the recording time for the 18 2D spectra totaled 6 ¼ days. Most of this time was spent in acquiring the two least sensitive experiment types, DQ-HN(CA) (just under 3 days) and (CO)HN(CA) (1 ¼ days). Taking into account that information from the DQ-HN(CA) experiment is only required when there are two or more consecutive residues of the ^{13}C/^{15}N labeled amino acid type in the protein sequence, and recalling that the (CO)HN(CA) is redundant if peaks are clearly identified in both HN(CO) and HN(CA) spectra, a reduced set consisting of HSQC, HN(CO), HN(CA), and HN(COCA) experiments can be obtained in about 2 days even under sensitivity-limited conditions.

It should however be noted that, although not a problem in the applications described here, peaks observed in the HSQC might escape detection in some or all of the 2D triple-resonance experiments because of short T_2 and/or severe conformational exchange leading to false negatives. This problem would be alleviated in related methods such as IDIS–HSQC/IDIS–TROSY (Golovanov et al. 2007) or CCLS–HSQC (Tonelli et al. 2007) where reference ^{15}N–^{1}H correlation spectra are recorded with constant-time type pulse sequences. Peaks detected here should also be observable in the ^{13}C edited/filtered subspectra because they use identical periods of transverse magnetization.

Conclusions

We have introduced a new combinatorial labeling scheme which either provides anchor points for or completes 3D triple-resonance based protein backbone resonance assignment while minimizing sample preparation efforts and expenses. It involves a four-step procedure to retrieve residue type and sequential information: (1) identification of the fully ^{13}C/^{15}N labeled amino-acid type within each sample from the pattern of occurrence in 2D triple-resonance spectra, (2) amino-acid type determination of the ^{15}N labeled residues by alignment of the remaining signals in [^{15}N, ^{1}H]–HSQC spectra of all samples, (3) discovery of sequential pairs with the ^{13}C/^{15}N labeled residue type in position $i-1$ and either a ^{15}N or the same ^{13}C/^{15}N labeled amino acid type in position i from their appearance in HN(COCA) and DQ-HN(CA) spectra, respectively, and (4) analysis of HN(CO) spectra to identify further sequential pairs involving 1-^{13}C labeled amino acid types in position $i-1$. Steps (2) and (4) are analogous to the combinatorial dual-selective labeling protocol whereas steps (1) and (3) are unique to the combinatorial triple-selective labeling method presented here and constitute the source of additional information. An alternative strategy—increasing the number of samples in combinatorial dual-selective labeling—not only requires the production of more isotopically labeled protein but also means that more residue types per sample have to be enriched with ^{15}N, which gradually reintroduces the overlap problem. On the other hand, a drawback of combinatorial triple-selective labeling is that more NMR spectra have to be recorded per sample, resulting in increased spectrometer time demands.

It has been demonstrated that using only three amino-acid specifically labeled samples and 2D spectra recorded at moderate polarizing field, a sufficient number of amino acid types and sequential pairs can be tagged to achieve complete backbone assignments with the help of complementary 3D HNCA and 3D HN(CO)CA experiments on [U–^{13}C; ^{15}N] labeled protein. Successful application to integral membrane proteins with apparent molecular weights of approximately 50 and 70 kDa indicates that the approach is feasible in cases where otherwise perdeuteration would be required. The method is in principle not limited to membrane proteins and may also be useful for assignment of other demanding targets such as larger protein complexes or intrinsically disordered proteins.

Acknowledgments This work was supported by the Deutsche Forschungsgemeinschaft (DO545/7-1 and SFB 807), NIH (U54 GM094608) and the Cluster of Excellence Frankfurt (Macromolecular Complexes).

References

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. J Biomol NMR 1:99–104

Bayrhuber M, Riek R (2011) Very simple combination of TROSY, CRINEPT and multiple quantum coherence for signal enhancement in an HN(CO)CA experiment for large proteins. J Magn Reson 209:310–314

Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem Phys Lett 69:185–189

Butterwick JA, MacKinnon R (2010) Solution structure and phospholipid interactions of the isolated voltage-sensor domain from KvAP. J Mol Biol 403:591–606

Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N–13C-labelled proteins. J Magn Reson 97:213–217

Czisch M, Boelens R (1998) Sensitivity enhancement in the TROSY experiment. J Magn Reson 134:158–160

Delaglio F, Torchia DA, Bax A (1991) Measurement of 15N–13C J couplings in staphylococcal nuclease. J Biomol NMR 1:439–446

Farjon J, Boisbouvier J, Schanda P, Pardi A, Simorre JP, Brutscher B (2009) Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. J Am Chem Soc 131:8571–8577

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Feuerstein S, Plevin MJ, Willbold D, Brutscher B (2011) iHADAMAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. J Magn Reson. doi:10.1016/j.jmr2011.10.019

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

Golovanov AP, Blankley RT, Avis JM, Bermel W (2007) Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions in vitro. J Am Chem Soc 129:6528–6535

Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins. J Biomol NMR 3:185–204

Hefke F, Bagaria A, Reckel S, Ullrich SJ, Dötsch V, Glaubitz C, Güntert P (2010) Optimization of amino acid type-specific ^{13}C and ^{15}N labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. J Biomol NMR 49:75–84

Ikura M, Kay LE, Bax A (1990a) A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 29:4659–4667

Ikura M, Krinks M, Torchia DA, Bax A (1990b) An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling. FEBS Lett 266:155–158

Jiang Y, Ruta V, Chen J, Lee A, MacKinnon R (2003) The principle of gating charge movement in a voltage-dependent K + channel. Nature 423:42–48

Kainosho M, Tsuji T (1982) Assignment of the three methionyl carbonyl carbon resonances in Streptomyces subtilisin inhibitor by a carbon-13 and nitrogen-15 double-labeling technique. A new strategy for structural studies of proteins in solution. Biochemistry 21:6273–6279

Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665

Kim HJ, Howell SC, Van Horn WD, Jeon YH, Sanders CR (2009) Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. Prog Nucl Magn Reson Spectrosc 55:335–360

Klammt C, Schwarz D, Dötsch V, Bernhard F (2007) Cell-free production of integral membrane proteins on a preparative scale. Methods Mol Biol 375:57–78

Lee D, Hilty C, Wider G, Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson 178:72–76

Lescop E, Rasia R, Brutscher B (2008) Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. J Am Chem Soc 130:5014–5015

Logan TM, Olejniczak ET, Xu RX, Fesik SW (1992) Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. FEBS Lett 314:413–418

Macura S (2009) Accelerated multidimensional NMR data acquisition by varying the pulse sequence repetition time. J Am Chem Soc 131:9606–9607

Maslennikov I, Klammt C, Hwang E, Kefala G, Okamura M, Esquivies L, Mors K, Glaubitz C, Kwiatkowski W, Jeon YH, Choe S (2010) Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. Proc Natl Acad Sci USA 107:10902–10907

Masterson LR, Tonelli M, Markley JL, Veglia G (2008) Simultaneous detection and deconvolution of congested NMR spectra containing three isotopically labeled species. J Am Chem Soc 130:7818–7819

Morita EH, Shimizu M, Ogasawara T, Endo Y, Tanaka R, Kohno T (2004) A novel way of amino acid-specific assignment in 1H–15N HSQC spectra with a wheat germ cell-free protein synthesis system. J Biomol NMR 30:37–45

Nietlispach D, Gautier A (2011) Solution NMR studies of polytopic alpha-helical membrane proteins. Curr Opin Struct Biol 21:497–508

Nietlispach D, Ito Y, Laue ED (2002) A novel approach for the sequential backbone assignment of larger proteins: selective intra-HNCA and DQ-HNCA. J Am Chem Soc 124:11199–11207

Ozawa K, Headlam MJ, Schaeffer PM, Henderson BR, Dixon NE, Otting G (2004) Optimization of an *Escherichia coli* system for cell-free synthesis of selectively N-labelled proteins for rapid analysis by NMR spectroscopy. Eur J Biochem 271:4084–4093

Ozawa K, Wu PS, Dixon NE, Otting G (2006) ^{15}N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. FEBS J 273:4154–4159

Parker MJ, Aulton-Jones M, Hounslow AM, Craven CJ (2004) A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. J Am Chem Soc 126:5020–5021

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

Pervushin KV, Wider G, Wuthrich K (1998) Single transition-to-single transition polarization transfer (ST2-PT) in [^{15}N,^{1}H]-TROSY. J Biomol NMR 12:345–348

Raschle T, Hiller S, Yu TY, Rice AJ, Walz T, Wagner G (2009) Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. J Am Chem Soc 131:17777–17779

Raschle T, Hiller S, Etzkorn M, Wagner G (2010) Nonmicellar systems for solution NMR spectroscopy of membrane proteins. Curr Opin Struct Biol 20:471–479

Reckel S, Sobhanifar S, Schneider B, Junge F, Schwarz D, Durst F, Löhr F, Güntert P, Bernhard F, Dötsch V (2008) Transmembrane segment enhanced labeling as a tool for the backbone assignment of alpha-helical membrane proteins. Proc Natl Acad Sci USA 105:8262–8267

Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen MK, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F, Schwalbe H, Güntert P, Dötsch V (2011) Solution NMR structure of proteorhodopsin. Angew Chem Int Ed Engl 50:11942–11946

Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc Natl Acad Sci USA 95:13585–13590

Sanders CR, Sönnichsen F (2006) Solution NMR of membrane proteins: practice and challenges. Magn Reson Chem 44:S24–S40

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

Schmidt JM, Hua Y, Löhr F (2010) Correlation of ^{2}J couplings with protein secondary structure. Proteins 78:1544–1562

Schubert M, Smalla M, Schmieder P, Oschkinat H (1999) MUSIC in triple-resonance experiments: amino acid type-selective 1H–15N correlations. J Magn Reson 141:34–43

Schwarz D, Junge F, Durst F, Frölich N, Schneider B, Reckel S, Sobhanifar S, Dötsch V, Bernhard F (2007) Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. Nat Protoc 2:2945–2957

Shenkarev ZO, Paramonov AS, Lyukmanova EN, Shingarova LN, Yakimov SA, Dubinnyi MA, Chupin VV, Kirpichnikov MP, Blommers MJ, Arseniev AS (2010) NMR structural and dynamical investigation of the isolated voltage-sensing domain of the potassium channel KvAP: implications for voltage gating. J Am Chem Soc 132:5630–5637

Shi J, Pelton JG, Cho HS, Wemmer DE (2004) Protein signal assignments using specific labeling and cell-free synthesis. J Biomol NMR 28:235–247

Shortle D (1994) Assignment of amino acid type in 1H–15N correlation spectra by labeling with 14N-amino acids. J Magn Reson B 105:88–90

Sobhanifar S, Reckel S, Junge F, Schwarz D, Kai L, Karbyshev M, Löhr F, Bernhard F, Dötsch V (2010) Cell-free expression and stable isotope labelling strategies for membrane proteins. J Biomol NMR 46:33–43

Staunton D, Schlinkert R, Zanetti G, Colebrook SA, Campbell ID (2006) Cell-free expression and selective isotope labelling in protein NMR. Magn Reson Chem 44:S2–S9

Szyperski T, Braun D, Fernandez C, Bartels C, Wüthrich K (1995) A novel reduced-dimensionality triple-resonance experiment for efficient polypeptide backbone assignment, 3D CO HN N CA. J Magn Reson B 108:197–203

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

Tamm LK, Liang B (2006) NMR of membrane proteins in solution. Prog Nucl Magn Reson Spectrosc 48:201–210

Tonelli M, Masterson LR, Hallenga K, Veglia G, Markley JL (2007) Carbonyl carbon label selective (CCLS) 1H–15N HSQC experiment for improved detection of backbone 13C–15N cross peaks in larger proteins. J Biomol NMR 39:177–185

Tonelli M, Masterson LR, Cornilescu G, Markley JL, Veglia G (2009) One-sample approach to determine the relative orientations of proteins in ternary and binary complexes from residual dipolar coupling measurements. J Am Chem Soc 131:14138–14139

Tonelli M, Singarapu KK, Makino S, Sahu SC, Matsubara Y, Endo Y, Kainosho M, Markley JL (2011) Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. J Biomol NMR 51:467–476

Trbovic N, Klammt C, Koglin A, Löhr F, Bernhard F, Dötsch V (2005) Efficient strategy for the rapid backbone assignment of membrane proteins. J Am Chem Soc 127:13504–13505

Weigelt J (1998) Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR Experiments. J Am Chem Soc 120:10778–10779

Weigelt J, van Dongen M, Uppenberg J, Schultz J, Wikström M (2002) Site-selective screening by NMR spectroscopy with labeled amino acid pairs. J Am Chem Soc 124:2446–2447

Wirmer J, Schwalbe H (2002) Angular dependence of 1J(Ni, Cαi) and 2J(Ni, Cα(i − 1)) coupling constants measured in J-modulated HSQCs. J Biomol NMR 23:47–55

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140

Wittekind M, Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. J Magn Reson B 101:201–205

Wu PS, Ozawa K, Jergic S, Su XC, Dixon NE, Otting G (2006) Amino-acid type identification in ^{15}N-HSQC spectra by combinatorial selective ^{15}N-labelling. J Biomol NMR 34:13–21

Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue ED, Cooper JA, Kainosho M, Yokoyama S (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. J Biomol NMR 11:295–306