J Biomol NMR (2009) 44:245–260
DOI 10.1007/s10858-009-9338-7

ARTICLE

Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively ^{13}C-labelled proteins

Victoria A. Higman · Jeremy Flinders · Matthias Hiller · Stefan Jehle · Stefan Markovic · Sebastian Fiedler · Barth-Jan van Rossum · Hartmut Oschkinat

Received: 1 April 2009 / Accepted: 22 June 2009 / Published online: 17 July 2009
© Springer Science+Business Media B.V. 2009

Abstract In recent years, solid-state magic-angle spinning nuclear magnetic resonance spectroscopy (MAS NMR) has been growing into an important technique to study the structure of membrane proteins, amyloid fibrils and other protein preparations which do not form crystals or are insoluble. Currently, a key bottleneck is the assignment process due to the absence of the resolving power of proton chemical shifts. Particularly for large proteins (approximately >150 residues) it is difficult to obtain a full set of resonance assignments. In order to address this problem, we present an assignment method based upon samples prepared using [1,3-^{13}C]- and [2-^{13}C]-glycerol as the sole carbon source in the bacterial growth medium (so-called selectively and extensively labelled protein). Such samples give rise to higher quality spectra than uniformly [^{13}C]-labelled protein samples, and have previously been used to obtain long-range restraints for use in structure calculations. Our method exploits the characteristic cross-peak patterns observed for the different amino acid types in ^{13}C-^{13}C correlation and 3D NCACX and NCOCX spectra. An in-depth analysis of the patterns and how they can be used to aid assignment is presented, using spectra of the chicken α-spectrin SH3 domain (62 residues), αB-crystallin (175 residues) and outer membrane protein G (OmpG, 281 residues) as examples. Using this procedure, over 90% of the Cα, Cβ, C′ and N resonances in the core domain of αB-crystallin and around 73% in the flanking domains could be assigned (excluding 24 residues at the extreme termini of the protein).

Keywords Solid-state NMR · Membrane proteins · Protein structure determination · Isotopic labelling · Resonance assignment

Introduction

Structure determination of membrane proteins, amyloid fibrils, complexes involving filament proteins or polydisperse protein complexes in a near-native environment is a challenging undertaking. Recently, there has been great interest in using solid-state NMR spectroscopy (ssNMR) as a tool to study such proteins and protein complexes. While structure determination of small proteins, among them the α-spectrin SH3 domain (Castellani et al. 2002), ubiquitin (Zech et al. 2005), GB1 (Zhou et al. 2007b; Franks et al. 2008) and Crh (Loquet et al. 2008), testifies to the potential of the technique, further methodological advances are required to extend the reach of ssNMR to the many large systems of interest (e.g. G-protein coupled receptors). Recent developments in labelling strategies, pulse sequence technologies and structure determination protocols suggest that the study of such large systems is feasible. However, a key problem is achieving sufficiently complete resonance assignment. In this context, a suite of assignment strategies is needed to obtain resonance-specific assignments for proteins larger than ~150 amino acids.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-009-9338-7) contains supplementary material, which is available to authorized users.

V. A. Higman · J. Flinders · M. Hiller · S. Jehle · S. Markovic · S. Fiedler · B.-J. van Rossum · H. Oschkinat (✉)
Leibniz-Institut für Molekulare Pharmakologie,
Robert-Rössle-Str. 10, 13125 Berlin, Germany
e-mail: oschkinat@fmp-berlin.de

Present Address:
J. Flinders
Structural Biology Department, Genentech, Inc.,
1 DNA Way, South San Francisco, CA 94080, USA

In recent years an ever-increasing number of solid-state assignments have been reported (Pauli et al. 2001; Böckmann et al. 2003; Igumenova et al. 2004a; b; Marulanda et al. 2004; Franks et al. 2005; Lange et al. 2005; Goldbourt et al. 2007; Li et al. 2008; Schneider et al. 2008; Jehle et al. 2009), mostly for small proteins of up to around 100 residues in length. The use of uniformly labelled protein in conjunction with ^{13}C-^{13}C correlation experiments and 2D or 3D experiments of the NCACX, NCOCX or CANCO type has been sufficient to achieve high levels of assignment for these proteins. However, with increasing complexity of the spectra or increasing line width these conventional assignment strategies may not yield sufficiently high degrees of assignment. In contrast to solution NMR, these experiments rely on correlations between carbons and nitrogens without being able to make use of the resolving power of amide proton frequencies. Although the use of protons in solid-state MAS NMR spectra is an emerging field of study (Zhou et al. 2007a; b; Agarwal and Reif 2008; Linser et al. 2008), there are still several technical hurdles which have to be overcome before this can be applied more widely.

Here we present an assignment procedure which addresses both the issue of increased overlap and that of lower spectral quality by exploiting extensive and selective labelling, i.e. by making use of samples prepared with [1,3-^{13}C] and [2-^{13}C]-glycerol in the bacterial growth medium (Hong and Jakes 1999; Castellani et al. 2002). Such labelling has previously been used to obtain an increased number of long-range cross peaks which can be translated into structural restraints for structure calculations (Castellani et al. 2002, 2003; Zech et al. 2005; Ferguson et al. 2006). In addition to the reduction of strong dipolar interactions, the benefits of extensive and selective labelling are reduced spectral overlap due to the reduction of the number of signals and a marked decrease in line width (and thus increase in resolution) due to the removal of most one-bond ^{13}C-^{13}C scalar and dipolar couplings, and the appearance of characteristic cross-peak patterns for specific amino-acid types. This assignment procedure is also applicable to small proteins with broad line widths or in situations where only limited amounts of sample are available. The idea is thus to introduce the [U-^{15}N, 1,3-^{13}C]-glycerol and [U-^{15}N, 2-^{13}C]-glycerol samples at the resonance assignment stage and not only later for restraint generation. This basic concept was already introduced in our group's study of C150.WW2 fibrils (Becker et al. 2008), but here we provide a complete and more detailed strategy for this approach.

We discuss primarily the cross-peak pattern in 3D-NCACX, 3D-NCOCX (Castellani et al. 2003) and ^{13}C-^{13}C correlation spectra recorded with different mixing times. The latter spectra are either proton-driven spin diffusion (PDSD) (Bloembergen 1949) or dipolar-assisted rotational resonance (DARR) (Takegoshi et al. 2001; Takegoshi et al. 2003) spectra. The cross peak patterns observed for these two types of spectrum are identical. For DARR spectra slightly shorter mixing times can be employed than for PDSD spectra on account of the more efficient magnetisation transfer. However, the order of magnitude remains similar and the mixing times mentioned in this paper are only intended as a rough guide. ^{13}C-^{13}C correlation and NCACX spectra recorded with short mixing times help to identify spin systems, and experiments recorded with longer mixing times as well as the NCOCX spectrum then provide sequential correlations which enable sequence-specific resonance assignment. Other spectra such as radio frequency-driven-dipolar-recoupling (RFDR; Bennett et al. 1992, 1998), C7 (Lee et al. 1995), POST-C7 (Hohwy et al. 1998), PAR (De Paëpe et al. 2008), REDOR (Gullion and Schaefer 1989), CANCO (Li et al. 2007) or PAIN (Lewandowski et al. 2007) may be used alternatively but are not discussed here.

Our assignment approach is shown and discussed for several proteins: the 62-residue α-spectrin SH3 domain (microcrystalline), the 175-residue protein αB-crystallin (PEG precipitated) and the 281-residue membrane protein, outer membrane protein G (OmpG, lipid-embedded, 2D-crystalline, the detailed assignment of this protein will be presented elsewhere). 1D slices taken from spectra of these proteins illustrating typical linewidths, signal-to-noise ratios and experimental parameters are shown in the supplementary information (Fig. S1–3). The SH3 domain yields beautifully resolved spectra, already allowing for nearly complete resonance assignment by conventional methods. The application of our new procedure to the considerably larger αB-crystallin, however, was essential for achieving near-complete assignment of the protein. The previously reported assignment of the core domain of αB-crystallin (Jehle et al. 2009) has been extended to include 73% of the Cα, Cβ, C′ and N resonances in the N- and C-terminal extensions (excluding 24 residues at the extreme termini). A list of chemical shifts is provided in the supplementary information. The even more challenging case of OmpG has called for an extension of the method in which samples with specifically labelled amino acids are employed. By discussing the evaluation of the data sets of all three proteins, we thus show how this methodology can be applied to a variety of systems and how it can be adapted to suit the particular requirements of different protein samples using other labelling schemes and pulse sequences.

Results

Labelling

The assignment strategy presented here relies on the specific labelling pattern observed in samples produced using

[1,3-^{13}C]- and [2-^{13}C]-glycerol (as well as [^{15}N]-NH_4Cl) in the bacterial growth medium. In this case, the amino acids which are derived from the glycolytic and pentose phosphate pathways (alanine, cysteine, glycine, histidine, leucine, phenylalanine, serine, tryptophan, tyrosine and valine) have all sites either ^{13}C or ^{12}C labelled in virtually 100% of the cases ('all-or-nothing' labelling, Fig. 1a, c, d). The remainder of the amino acids are synthesised from precursors which are involved in the citric acid pathway and give rise to non-random mixtures of isotopomers ('mixed' labelling), as shown in Fig. 1b, d, e. In this case, several sub-groups of amino acids show similar labelling patterns (Fig. 1b). These labelling schemes have been verified by LeMaster and Kushlan (LeMaster and Kushlan 1996) and by solution-state NMR studies of the SH3 domain samples (Castellani et al. 2002). Based on these NMR results, the relative populations of the different isotopomers were determined. Figure S4 in the supplementary information illustrates the isotopomers and their populations for the ten different amino acid types with 'mixed' labelling in SH3. For αB-crystallin and OmpG slightly different populations of isotopomers were observed, most likely due to differences in the *E.coli* amino acid metabolism arising from different amino acid distributions in the proteins. More importantly, weak peaks arising from unexpected isotopomers (LeMaster and Kushlan 1996; Castellani et al. 2002) were observed for some of the amino acids arising from the glycolytic/pentose phosphate pathway. Although it is necessary to be aware that such scrambling can occur, it does not interfere with the assignment procedure outlined here. The convention throughout this paper is to refer to [2-^{13}C]-glycerol labelling in red and [1,3-^{13}C]-glycerol labelling in blue (previously the colour scheme used for these samples was red/green, but on account of the relatively high incidence of red/green colour blindness we use red/blue here in order to ensure clarity in a paper designed as a general guide to solid-state MAS NMR assignment using the glycerol-based labelling schemes). In addition to the two glycerol-based samples, a uniformly [^{13}C]-labelled sample is employed.

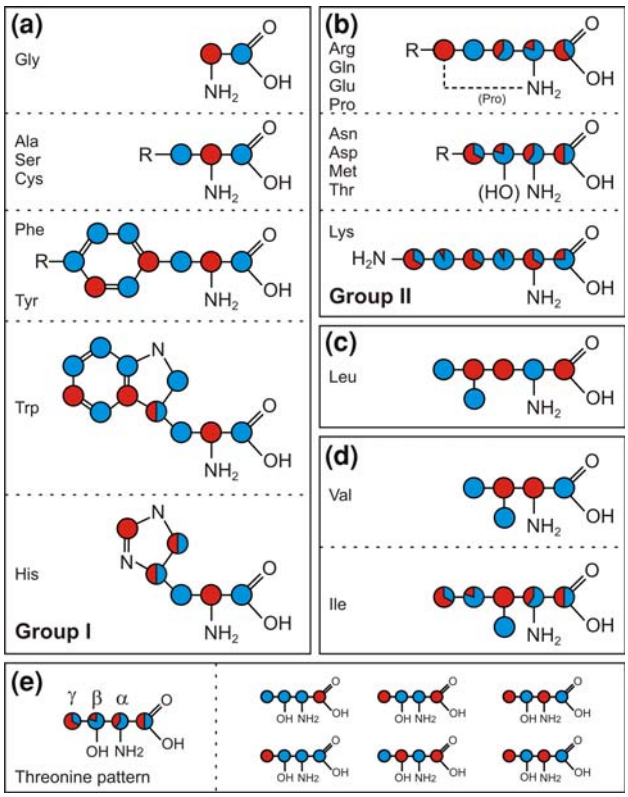


Fig. 1 Schematic representation of the effective ^{13}C enrichment of amino acids in the chicken α-spectrin SH3 domain expressed in *E. coli* BL21 (DE3). Sites which are [^{13}C]-labelled by growth on [1,3-^{13}C]-glycerol are shown in blue; sites labelled by growth on [2-^{13}C]-glycerol are shown in red. For residues with mixed labelling, the percentage of labelling is represented using relative blue/red colouring, except for the tryptophan Cγ and histidine Cδ and Cε sites, for which the percentages could not be determined and a half/half distribution is displayed. **a** Group I residues have [^{13}C]-labelled C′ and Cβ sites in the [1,3-^{13}C]-glycerol sample and [^{13}C]-labelled Cα sites in the [2-^{13}C]-glycerol sample. **b** Group II residues have mixed labelling patterns arising from different isotopomers as shown for threonine in (**e**). **c** Leucine is the only residue with a backbone labelling pattern exactly opposite to that of the Group I residues. **d** Valine and isoleucine residues present exceptions from Groups I and II, respectively, as their Cβ sites are only [^{13}C]-labelled in the [2-^{13}C]-glycerol sample. **e** Illustration of the fractional labelling pattern seen for Group II residues and isoleucine, using threonine as an example. A full list of the isotopomers is provided in the supplementary information

Assignment strategy

The first step comprises the identification of as many spin systems as possible and their assignment to an amino-acid type by identifying the specific labelling patterns in the [1,3-^{13}C]-glycerol and [2-^{13}C]-glycerol samples in addition to their chemical shifts. The spectra utilised in this process are ^{13}C-^{13}C correlation and NCACX spectra with short mixing times. It was found convenient to define an 'entry point' into the resonance assignment for each amino acid type which usually consists of cross peaks that are unique based on their chemical shifts and/or labelling patterns. For example, the proline Cδ-Cβ cross peaks are well resolved in the spectra of the [2-^{13}C]-glycerol samples (at ~50 and ~32 ppm). Likewise, all peaks present at ~27 ppm in the carbonyl region of the [2-^{13}C]-glycerol spectrum (with a mixing time on the order of 50 ms) can be attributed to Leucine C′-Cγ cross peaks.

In the second step, the identified spin systems are linked sequentially. The diagram in the lower panel of Fig. 2a shows how magnetisation transfers in NCACX and NCOCX spectra can be used to obtain sequential links in an amino-acid triplet with a central leucine residue. By

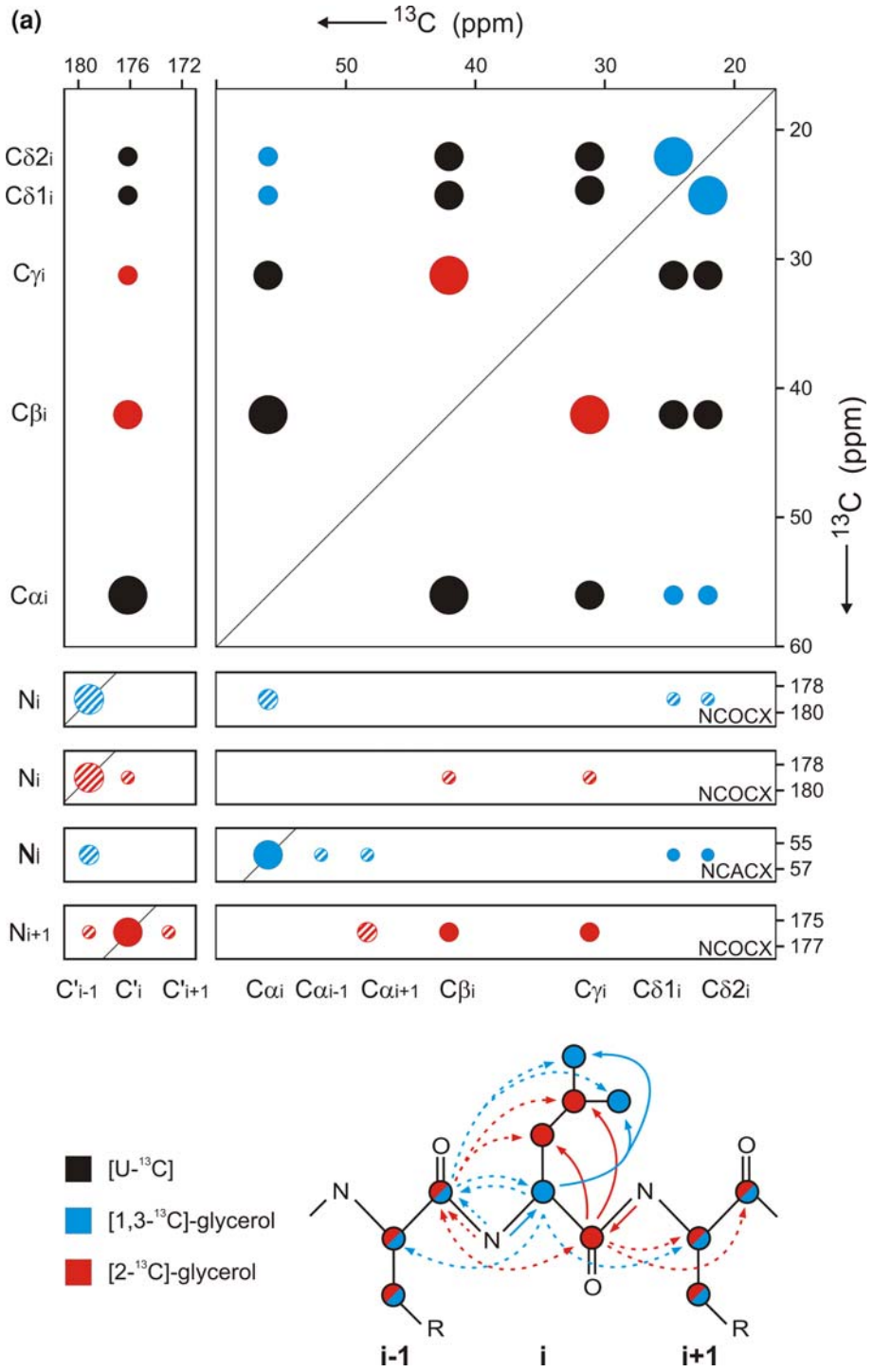


Fig. 2 **a** Schematic drawing of the cross peaks expected for a leucine residue in ^{13}C-^{13}C correlation spectra, NCACX and NCOCX spectra of the U-[^{13}C,^{15}N] (*black*), [1,3-^{13}C]-glycerol (*blue*) and [2-^{13}C]-glycerol (*red*) samples. Cross peaks visible in the latter two samples are also visible in the U-[^{13}C,^{15}N] sample. A schematic drawing shows the magnetisation transfers giving rise to the cross peaks in the NCACX and NCOCX spectra. Cross peaks/transfers which depend on the labelling of neighbouring residues are shown *hatched*. **b** Spectra illustrating cross peaks arising from the Thr32-Leu33-Leu34 motif in the chicken α-spectrin SH3 domain. PDSD spectra are shown for the U-[^{13}C,^{15}N] (*black*, 100 ms mixing time), [1,3-^{13}C]-glycerol (*blue*, 50 ms mixing time) and [2-^{13}C]-glycerol (*red*, 50 ms mixing time) samples. Four strips taken from NCACX and NCOCX spectra of the [1,3-^{13}C]-glycerol (*blue*) and [2-^{13}C]-glycerol (*red*) samples (the mixing time in ms is indicated on the *right*). A schematic drawing shows the magnetisation transfers giving rise to the cross peaks observed in the NCACX and NCOCX spectra. For clarity, only positive contours are drawn

analogy with solution NMR, there are some 'quasi-through-bond' links, such as those provided by an NCOCX spectrum which directly links N_{i+1} with C'_i. Alternatively, links can be established 'through-space' via short-range PDSD/DARR transfers, either in 2D ^{13}C-^{13}C correlation spectra or 3D NCACX and NCOCX spectra. Based on the labelling patterns, we suggest in the following sections which 3D 'strips' best provide such linking information for each amino acid type, focussing on the cross peak patterns observed in strips from NCACX and NCOCX spectra taken at the frequencies of N_i and N_{i+1}, respectively. An instructive example is shown in the upper panel of Fig. 2a by means of schematic diagrams simulating ^{13}C-^{13}C correlation spectra and strips from 3D NCACX and NCOCX spectra for leucine. It contains predictions of cross-peak locations on the basis of average chemical shifts (from the

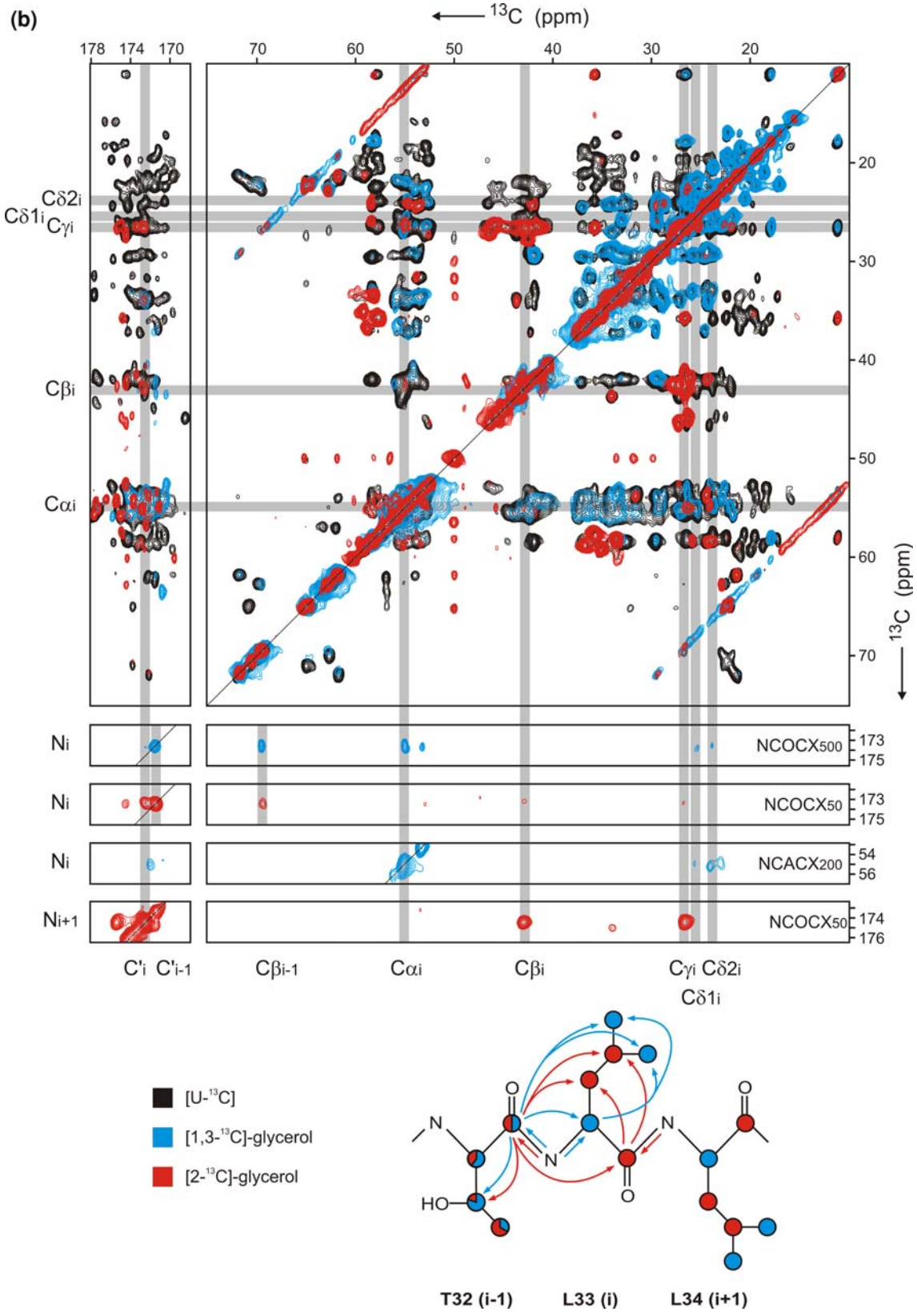


Fig. 2 continued

Biological Magnetic Resonance Data Bank, http://www.bmrb.wisc.edu) and the labelling patterns determined for the SH3 domain. Schematic representations of regions from ^{13}C-^{13}C correlation spectra and strips from 3D NCACX and NCOCX spectra have been generated for each of the twenty amino acid types and are available as supplementary information. Figure 2b shows regions from SH3 domain spectra illustrating cross peaks arising from the Thr32-Leu33-Leu34 motif which corresponds to the situation in Fig. 2a. It can be seen that in real cases some predicted cross peaks are absent. This may be a result of dynamics, low signal-to-noise ratios, or variations in the labelling pattern.

The way in which sequential links can be made varies depending on the amino acid types involved. To simplify the discussions below, the amino acids are divided into three groups based on the labelling observed at the C′ and Cα sites in the [1,3-^{13}C]- and [2-^{13}C]-glycerol labelled samples: Group I contains the amino acids with 100% labelling except leucine and valine (alanine, cysteine, glycine, histidine, phenylalanine, serine, tryptophan and tyrosine, Fig. 1a); Group II contains the amino acids with mixed isotopomers except isoleucine (arginine, asparagine, aspartic acid, glutamine, glutamic acid, methionine, lysine, proline and threonine, Fig. 1b); finally leucine is considered separately, since its C′ and Cα labelling is unique (Fig. 1c). Although, according to their C′ and Cα labelling patterns, valine and isoleucine belong to Group I and Group II, respectively, their Cβ and Cγ labelling differs significantly from the other group members (Fig. 1d). Therefore, they need to be treated slightly differently and do not entirely fit the schemes shown in Fig. S5 in the supplementary information. This figure summarises the different transfers which are expected for pairs of amino acids from different groups in NCACX and NCOCX spectra of the [1,3-^{13}C]- and [2-^{13}C]-glycerol labelled samples. Transfers shown in bold represent strong cross peaks which are almost always observed for all amino acids except where strong dynamic effects are present. Transfers indicated by dashed arrows represent weak cross peaks and are therefore not always observed, or are only observed for certain amino acid types within a particular group. This arises from the subtle differences in the labelling patterns of the Group II amino acids whereby, for instance, the Cγ atom in arginine, glutamine, glutamic acid and proline is exclusively labelled in the [1,3-^{13}C]-glycerol sample but in methionine and lysine is predominantly labelled in the [2-^{13}C]-glycerol sample (Fig. S4).

Some special, recurring patterns are particularly useful when sequentially linking spin systems. They depend not only on the labelling of residue i, but also on the labelling of residues i ± 1. Examples are the N_i-$C\alpha_i$-C'_{i-1} cross peaks in an NCACX spectrum and the N_i-C'_{i-1}-$C\alpha_i$ cross peaks in an NCOCX spectrum (Fig. S5). Provided that one carbon atom is fully labelled and the other at least partially, these cross peaks are usually fairly strong. Furthermore, they are, in effect, symmetrical and therefore reinforce one another. $N_iC'_{i-1}$ strips are especially useful because they include N_i-C'_{i-1}-CX_{i-1} and N_i-C'_{i-1}-CX_i cross peaks and thus provide information on at least two resonances from each residue. Further recurring patterns are observed when inspecting NCACX and NCOCX spectra recorded with longer mixing times (~300–500 ms). It should, in principle, always be possible to observe N_i-$C\alpha_i$-$C\alpha_{i\pm1}$ cross peaks in NCACX strips if $C\alpha_i$ is labelled as well as $C\alpha_{i\pm1}$ in the sample under consideration (Ferguson et al. 2006; Becker et al. 2008). In a similar manner, N_i-C'_i-$C'_{i\pm1}$ cross peaks are, in principle, visible in any NCOCX strip, provided both C'_i and $C'_{i\pm1}$ are labelled and the signal-to-noise ratio and the resolution are high enough. These general patterns which depend on the labelling of both the central amino acid and its neighbours will largely be ignored in the discussions for the individual amino acid types below.

Leucine

A good starting point is leucine because its carbonyl site is 100% labelled in the [2-^{13}C]-glycerol sample, while the Cα site is 100% labelled in the [1,3-^{13}C]-glycerol sample. This is in sharp contrast to the remaining 9 all-or-nothing amino acids. Cross peaks involving the Cβ and Cγ atoms are very dominant in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample and provide an obvious starting point. The carbonyl carbons can then be identified from C′-Cβ and C′-Cγ cross peaks in the same spectrum. Cross peaks of the type Cδ1-Cδ2 and Cα-Cδ1/δ2 can be spotted in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample at all mixing times. The peaks from the two glycerol samples can be linked using the spectra of the [U-^{13}C]-labelled sample. The backbone nitrogen chemical shift can be determined using the NCACX spectrum. If the [U-^{13}C] sample is too crowded for individual leucine cross peaks to be resolved, then N_i-$C\alpha_i$-$C\delta1_i$/$C\delta2_i$ cross peaks in the [1,3-^{13}C]-glycerol sample can be used. Alternatively, since these can be overlapped with argenine N_i-$C\alpha_i$-$C\gamma_i$ signals, the relatively strong N_i-C'_{i-1}-$C\alpha_i$ peaks in the NCOCX of the [1,3-^{13}C]-glycerol sample can also be used to aid the identification of the backbone nitrogen chemical shift (Fig. 3c).

Sequential correlations are contained in the NCOCX spectra of both glycerol-based samples. In most cases, the $N_iC'_{i-1}$ strips from the [1,3-^{13}C]-glycerol sample provide links to the preceding residue and the $N_{i+1}C'_i$ strips from the [2-^{13}C]-glycerol spectrum to the following residue (Figs. S5e–h; 3c, e). The $N_{i+1}C'_i$ strips from the NCOCX spectrum of the [2-^{13}C]-glycerol sample are particularly helpful, since they contain strong cross peaks involving the

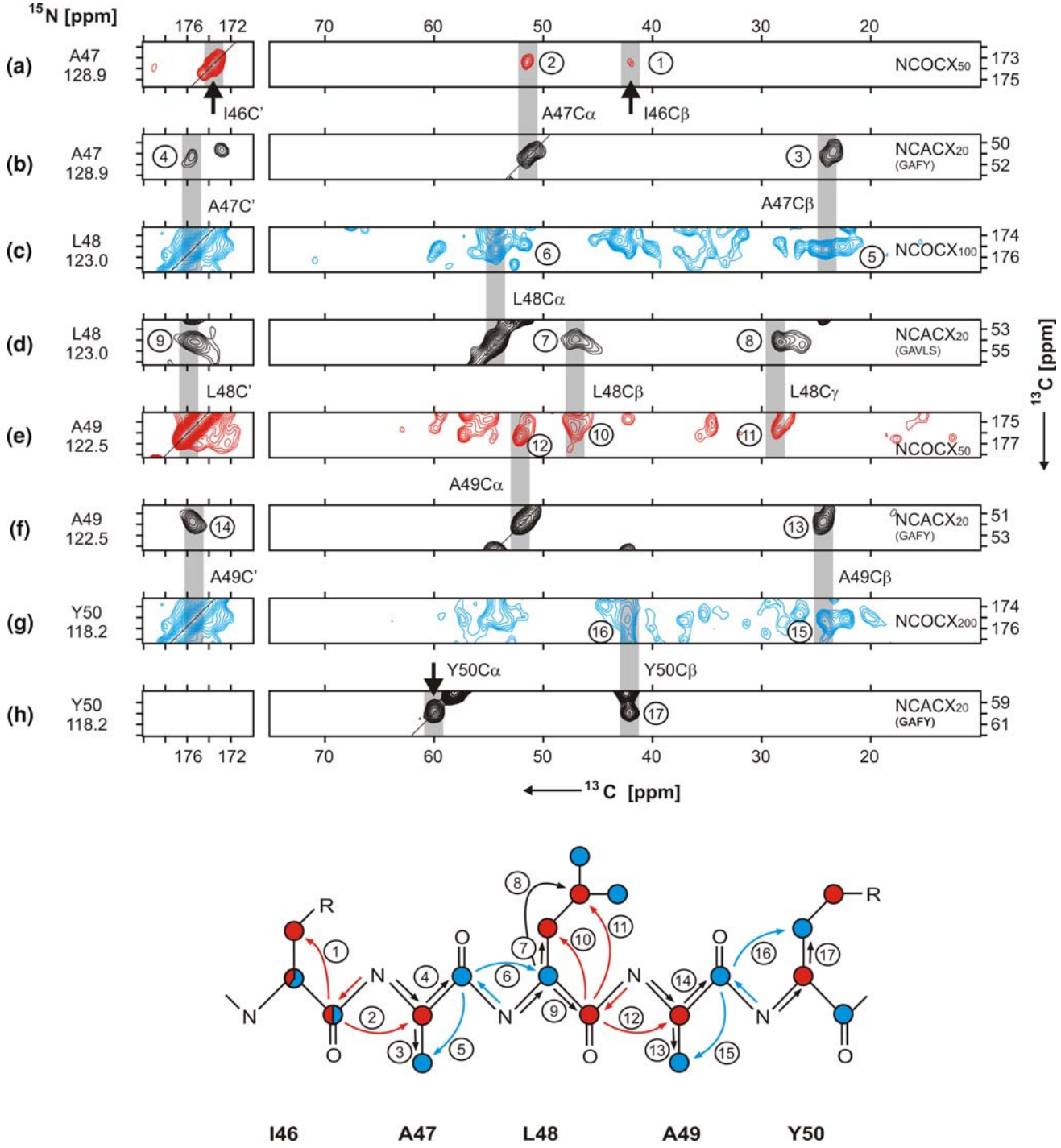


Fig. 3 Strips taken from 3D NCACX and NCOCX spectra of the Ile46-Ala47-Leu48-Ala49-Tyr50 motif in OmpG recorded at 400 MHz and 270 K. Mixing times in ms are indicated on the *right*. Spectra are coloured *blue* ([1,3-^{13}C]-glycerol sample), *red* ([2-^{13}C]-glycerol sample) and *black* (OmpG-GAFY and OmpG-GAVLS samples containing [U-^{13}C, U-^{15}N] and [2,3-^{13}C, U-^{15}N]-labelled amino acids). A schematic diagram (*lower panel*) shows the labelling of the residues in the glycerol-based samples (colour coding as in Fig. 1) and the magnetisation transfers giving rise to the cross peaks observed in the spectra. For clarity, only positive contours are drawn. For discussion of the figure see text

leucine Cβ and Cγ resonances. This means that the signal pattern of this strip can generally be used to identify a particular leucine spin system unambiguously. In most cases the Cα_{i+1} can been seen, too, unless the following residue is also a leucine residue in which case, of course, it is not labelled. Figure 3 shows the example of the

Ile46-Ala47-Leu48-Ala49-Tyr50 motif in OmpG. The $N_{i+1}C'_i$ strip in the NCOCX of the [2-^{13}C]-glycerol sample contains cross peaks involving Leu48Cβ, Leu48Cγ and Ala49Cα (Fig. 3e, peaks 10–12, respectively). Figures 2 and 3 also show how the $N_iC'_{i-1}$ strips of an NCOCX spectrum can be used to establish sequential links, dependent upon the labelling at C'_{i-1}. For the case of the Ala47-Leu48 motif in OmpG, the Ala47C$'$ site is ^{13}C-labelled in the [1,3-^{13}C]-glcyerol sample and thus the NCOCX spectrum of this sample contains cross peaks involving Ala47Cβ and Leu48Cα (Fig. 3c, peaks 5 and 6, respectively). The $N_iC\alpha_i$ strip of the [1,3-^{13}C]-glycerol NCACX spectrum may also provide sequential correlations. Dependent upon the labelling of the neighbouring residues, links to neighbouring $C\alpha_{i\pm1}$ or C'_{i-1} carbons may be observed, or, in the reverse direction, cross peaks involving the leucine Cα may be observed in the $N_iC\alpha_i$ strips of the neighbouring residues (Fig. S5e, g).

Valine

Due to their all-or-nothing labelling pattern and their chemical shift distribution, valine residues are also ideal starting points for resonance assignment. In ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol samples, strong Cα-Cβ cross peaks with chemical shifts of $\sim$60 and 35 ppm are expected at all mixing times. Large cross peaks in the 2D spectra of the [1,3-^{13}C]-glycerol samples seen at all mixing times near the diagonal at $\sim$20 ppm can be assigned as valine Cγ1-Cγ2 cross peaks. The two sets of peaks appearing in the spectra of the [1,3-^{13}C]- and [2-^{13}C]-glycerol samples can be linked via the Cα-Cγ1/Cγ2 and the Cβ-Cγ1/Cγ2 cross peaks which appear in the spectra of the [U-^{13}C] sample. Cross peaks involving carbonyl atoms are present in spectra of the [U-^{13}C] sample. C$'$-Cγ1/Cγ2 cross peaks expected in the spectra of the [1,3-^{13}C]-glycerol samples are often only visible after $\sim$200 ms mixing time. Overlap in the 2D spectra can best be resolved by means of the 3D NCACX spectrum of the [2-^{13}C]-glycerol sample which gives rise to strong N_i-Cα_i-Cβ_i cross peaks (Fig. S6b in the Supplementary Information, peak 4). These cross peaks also enable the extraction of the backbone nitrogen chemical shift.

At the sequential assignment stage, the 3D NCOCX spectrum of the [1,3-^{13}C]-glycerol sample provides useful information, since the carbonyl site is 100% ^{13}C-labelled. For example, in the $N_{i+1}C'_i$ strip of this spectrum a diagonal peak and the cross peaks involving Cγ1/Cγ2 are observed, when the mixing time is sufficiently long (Fig. S6c, peak 7). Dependent upon the identity of the neighbouring amino acids, sequential cross peaks involving their carbonyl or side-chain carbons may also be visible (Fig. S5a, c, e). In the case of SH3 Val23, a cross peak with the chemical shift of the following Thr24 carbonyl carbon is observed, as well as cross peaks involving the Thr24 Cα and Cβ (Fig. S6c, peaks 8–10). $N_iC'_{i-1}$ strips from the NCOCX spectrum will yield correlations to the preceding amino acid, with the labelling of the C'_{i-1} carbon determining whether peaks will be present either in the [1,3-^{13}C] or [2-^{13}C]-glycerol sample or both. The $N_iC\alpha_i$ strip from the [2-^{13}C]-glycerol NCACX spectrum is expected to yield not only the N_i-Cα_i-Cβ_i cross peak, but also possible cross peaks to the preceding carbonyl carbon (C'_{i-1}) and to neighbouring Cα carbons ($C\alpha_{i\pm1}$), if these are labelled (Fig. S5b, d, f). In the case of SH3 Val23, the expected intra-residue cross peaks are observed, as well as several sequential cross peaks to Glu22 and Thr24 (Fig. S6b, peaks 5 and 6). Similarly, the Val23 Cα and Cβ carbons can be observed in [2-^{13}C]-glycerol strips belonging to the neighbouring residues.

Glycine

Depending on the glycerol precursor used, either the Cα or the C$'$ site is labelled in glycine. Therefore, an intra-residue cross peak (C$'$-Cα) is only observed in a ^{13}C-^{13}C correlation spectrum of the [U-^{13}C] labelled sample. The backbone nitrogen chemical shift is easily determined using the N_i-Cα_i-C'_i cross peaks observed in the NCACX spectrum of the [U-^{13}C] sample.

As glycine consists of a small spin-system with only one large intra-residue dipolar coupling, many sequential and long-range links via weak dipolar couplings are seen for this amino acid in spectra of the [U-^{13}C]-labelled sample. Links to the carbonyl, Cα and Cβ atoms of neighbouring amino acids are even more pronounced in the $N_iC\alpha_i$ strips of NCACX spectra of the [2-^{13}C]-glycerol sample (though this is of course dependent upon the labelling of the neighbouring residue, Fig. S5b, d, e). Similarly, the $N_{i+1}C'_i$ strips from the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample can provide links to $C'_{i\pm1}$ or $C\alpha_{i+1}$ atoms dependent upon the labelling of the neighbouring residues (Fig. S5a, c, e). Due to the special chemical shift of the glycine Cα resonance, well resolved inter-residue cross peaks involving other Cα resonances can be resolved in 2D ^{13}C-^{13}C correlation spectra. These provide an additional useful source of information for assignment.

Alanine, serine and cysteine

These three amino acids are identically labelled and thus give rise to similar spectral patterns. Alanine, serine and cysteine can be distinguished from one another by their different Cβ chemical shifts ($\sim$18 ppm for alanine, $\sim$63 ppm for serine, $\sim$28 ppm for reduced cysteine and $\sim$41 ppm for oxidised cysteine). For these residues, only

the C′-Cβ cross peak appears in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample and no intra-residue cross peaks are present for the [2-^{13}C]-glycerol sample. Based on the metabolic pathways of *E.coli*, serine should experience more scrambling than alanine or cysteine, leading to the observation of Cα-Cβ cross peaks in the glycerol spectra. Cα-Cβ cross peaks were indeed observed for serine in the [2-^{13}C]-glycerol spectra of all three proteins investigated here, but these were generally only weak. More noticeable were Cα-Cβ cross peaks for serine and alanine in spectra of the [1,3-^{13}C]-glycerol sample of OmpG (but not the other two proteins). The backbone nitrogen resonance can only be identified from N_i-Cα_i-Cβ_i or N_i-Cα_i-C′$_i$ cross peaks in the NCACX spectrum of the [U-^{13}C]-labelled sample.

The sequential links which can be obtained for alanine, serine and cysteine in the 3D spectra are in fact characteristic for all the remaining Group I residues. The N_{i+1}C′$_i$ strips from the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample contain strong N_{i+1}-C′$_i$-Cβ_i cross peaks and, in addition, cross peaks of the type N_{i+1}-C′$_i$-Cα_{i+1} or N_{i+1}-C′$_i$-Cβ_{i+1} (which of these peaks is visible depends on the labelling of residue i+1). This means that there are usually two resonances by which to identify each of the two connected amino-acid spin systems. Figure 3g shows such a strip linking the carbonyl and Cβ of Ala49 in OmpG to the nitrogen and Cβ of Tyr50 (peaks 15 and 16). The occurrence of sequential links using N_iC′$_{i-1}$ strips from the NCOCX spectrum depends on the nature of the preceding amino acid. In the case of OmpG Ala47 where the preceding amino acid is an isoleucine (whose C′ is ^{13}C labelled in both glycerol-based samples), cross peaks involving both the alanine Cα and the isoleucine Cβ can be seen in the N_iC′$_{i-1}$ strip from the NCOCX spectrum of the [2-^{13}C]-glycerol sample (Fig. 3a, peaks 1 and 2). Sequential links to neighbouring Cα atoms can be observed in the NCACX spectra of the [2-^{13}C]-glycerol sample recorded with long mixing time in the form of N_i-Cα_i-C$\alpha_{i\pm1}$ cross peaks, when the C$\alpha_{i\pm1}$ sites are also labelled.

Phenylalanine, tyrosine, tryptophan and histidine

Identical to alanine, the aromatic amino acids all have carbonyl and Cβ sites which are 100% ^{13}C enriched in the [1,3-^{13}C]-glycerol sample, and Cα sites which are 100% ^{13}C enriched in the [2-^{13}C]-glycerol sample, resulting in a similar spectral pattern. Thus the C′-Cβ cross peak appears in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample and the only intra-residue cross peaks present in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample involve at least one aromatic carbon resonance. The general labelling pattern of the aromatic rings is shown in Fig. 1a, but appears to vary somewhat from protein to protein. Unfortunately, peaks arising from intra-ring correlations in ^{13}C-^{13}C correlation spectra tended to be weak for all three proteins studied here, most likely due to dynamics effects. But cross peaks linking aliphatic and aromatic resonances are usually visible and can significantly aid spin system and amino acid type identification.

Phenylalanine and tyrosine both show intense Cα-Cγ cross peaks in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample. A convenient way to resolve overlap in 2D, identify the backbone nitrogen chemical shift and distinguish between phenylalanine and tyrosine is to use the N_i-Cα_i-Cγ_i peaks in the NCACX spectrum of the [2-^{13}C]-glycerol sample.

Cα-Cγ and Cβ-Cγ cross peaks in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol and the [1,3-^{13}C]-glycerol samples, respectively, provide a convenient entry point into the assignment of the tryptophan residues. Tryptophan nitrogen chemical shifts are most easily extracted from N_i-Cα_i-Cγ_i/C$\delta2_i$/$\zeta3_i$ cross peaks the NCACX spectrum of the [2-^{13}C]-glycerol sample.

Histidine residues give rise to strong Cβ-Cγ and weaker Cβ-Cδ2 cross peaks in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample. The Cα-Cε1 cross peak predicted for the [2-^{13}C]-glycerol sample is generally too weak to be observed. The spin system identification can be completed using additional cross peaks in the uniformly labelled sample.

Sequential links can be found in analogy to alanine using the [1,3-^{13}C]-glycerol NCOCX N_{i+1}C′$_i$ strips and [2-^{13}C]-glycerol NCACX N_iCα_i strips. Likewise, beginning with the nitrogen of the aromatic residue, N_iC′$_{i-1}$ strips from the NCOCX spectrum of either the [1,3-^{13}C]-glycerol or the [2-^{13}C]-glycerol sample may be inspected, dependent upon the labelling of the C′$_{i-1}$ atom. In many cases peaks involving the Cγ or Cδ ring carbons are visible, especially in spectra of the [2-^{13}C]-glycerol sample, and these provide a useful additional means of unambiguously identifying the aromatic spin system.

Proline

Proline is one of the ten ‘mixed’ amino acids and is composed of a mixture of five different isotopomers (Fig. 1b). The resulting spectra are therefore more complicated to analyse than those for the ‘all-or-nothing’ amino acids discussed thus far. Although arginine and glutamine/glutamic acid have the same labelling pattern as proline, their chemical shifts differ significantly, and so they will be discussed separately.

In the 2D ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol and the [U-^{13}C] labelled samples, cross peaks of the type Cα-Cδ and Cβ-Cδ are good starting points due to the unique chemical shifts of the Cδ atoms. From here, it is

possible to locate the Cα-Cβ and Cα-Cγ cross peaks in the spectra of the [1,3-^{13}C]-glycerol and the [U-^{13}C] samples. While for the SH3 domain, these peaks only appear after ~ 300 ms mixing time in the spectrum of the [1,3-^{13}C]-glycerol sample, they are already very dominant at 50 ms mixing time in the OmpG spectra. Assignment of the Cβ-Cγ cross peaks, which lie near the diagonal, can be accomplished using either the [1,3-^{13}C]-glycerol or [U-^{13}C] labelled samples, unless hampered by overlap with other residue types. The carbonyl cross peaks can be assigned by overlaying all three spectra.

Since prolines lack an amide proton, transfer of magnetization from hydrogen onto the backbone nitrogen is not always very efficient and can lead to low signal intensities in the NCACX and NCOCX spectra. However, since the proline Cδ atom has a chemical shift similar to that of other Cα atoms and is directly bonded to the nitrogen, it can give rise to N_i-$C\delta_i$-CX_i peaks in the NCACX spectrum and these can be used in order to identify the backbone nitrogen chemical shift. For OmpG it was also possible to obtain some weak N_i-$C\alpha_i$-CX_i cross peaks in the [1,3-^{13}C]-glycerol NCACX spectrum.

Sequential links to prolines can be identified via N_{i+1}-C_i-CX_i cross peaks in the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample. When preceded by leucine (whose C′ is fully labelled in the [2-^{13}C]-glycerol sample) it was also possible to identify N_i-C'_{i-1}-CX_{i-1} cross peaks in spectra of OmpG. Furthermore, 2D ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample with long mixing times often give rise to well resolved cross peaks involving the Cδ chemical shift.

Arginine

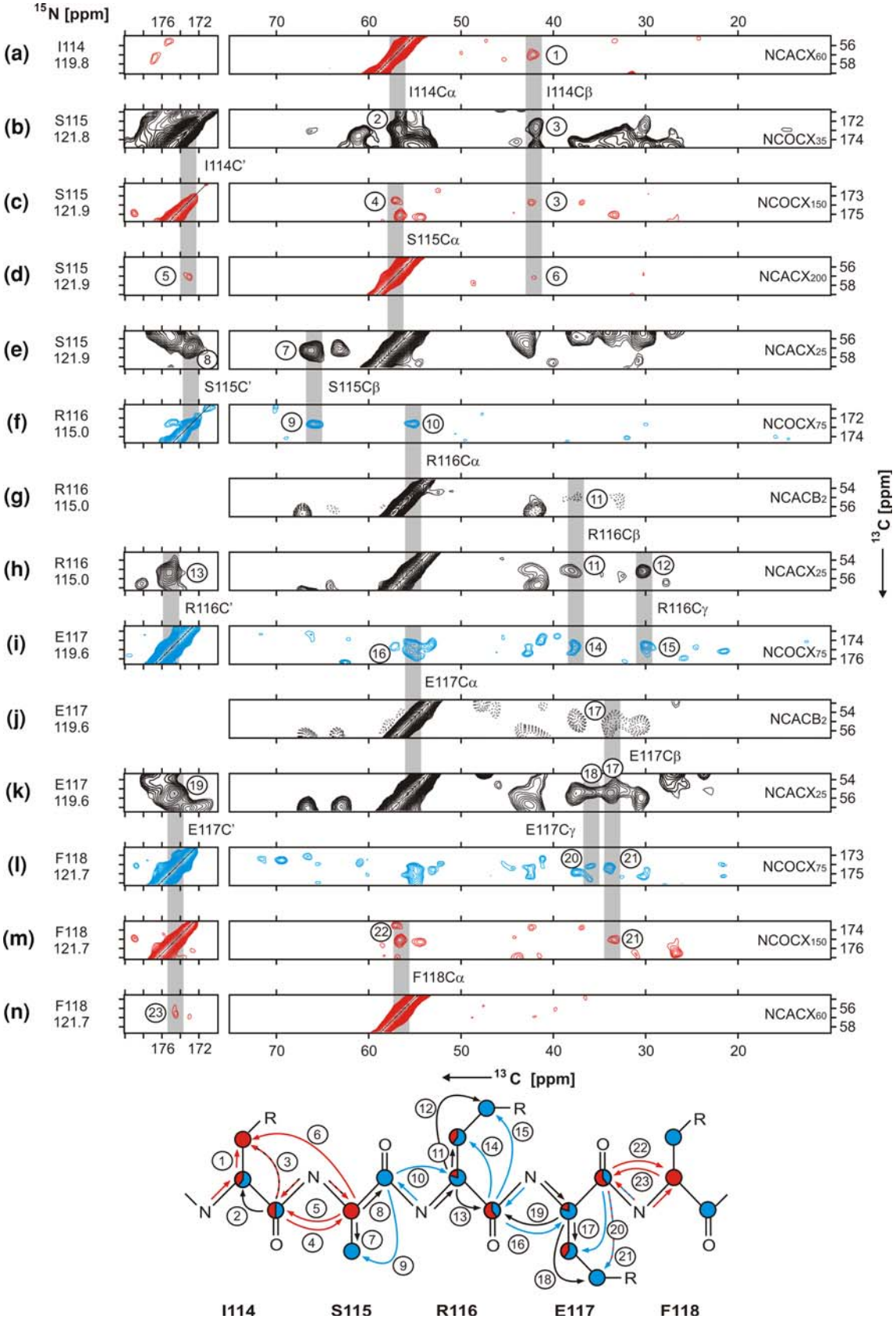
The identification of arginine cross peaks can be difficult since many are overlapped with signals arising from other amino acid types. If there are only a small number of arginine residues in the protein then identification may be possible via the guanidinium group (as was the case for the α-spectrin SH3 domain). This is advantageous, since the cross peaks involving Cζ lie at ~160 ppm and are therefore well separated from other signals. Other cross peaks which can be used to identify arginine residues include the intense Cα-Cγ signals at ~55–59/~30 ppm in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample in the vicinity of leucine and lysine signals, and the Cβ-Cδ cross peaks at ~30/~43 ppm in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C-glcyerol] sample. The NCACX spectrum of the [1,3-^{13}C]-glycerol sample contains strong N_i-$C\alpha_i$-$C\beta_i$ and N_i-$C\alpha_i$-$C\gamma_i$ cross peaks which can be very helpful in identifying arginine spin systems. Unfortunately, these cross peaks may sometimes be overlapped with leucine N_i-$C\alpha_i$-$C\delta_i$ cross peaks or glutamine/glutamic acid N_i-$C\alpha_i$-$C\beta_i$/$C\gamma_i$ cross peaks.

Fig. 4 Strips taken from 3D NCACX and NCOCX spectra of the Ile114-Ser115-Arg116-Glu117-Phe118 motif in αB-crystallin recorded at 700 MHz and 270 K. Mixing times in ms are indicated on the *right*. Spectra are coloured *blue* ([1,3-^{13}C]-glycerol sample), *red* ([2-^{13}C]-glycerol sample) and *black* ([U-^{13}C] sample). A schematic diagram (*lower panel*) shows the labelling of the residues in the glycerol-based samples (colour coding as in Fig. 1) and the magnetisation transfers giving rise to the cross peaks visible in the spectra. For clarity, only positive contours are drawn, except for the NCACB spectrum in which negative contours are shown using dashed lines. For discussion of the figure see text

The sequential assignment of arginines can be achieved using strips from the NCOCX spectra of the [2-^{13}C]-glycerol sample, where cross peaks usually only involve the Cβ site, e.g. N_{i+1}-C'_i-$C\beta_i$ cross peaks (or N_i-C'_{i-1}-$C\beta_i$ cross peaks if the arginine residue is preceded by a leucine). Alternatively, the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample may be used, evaluating cross peaks involving the Cα, Cβ or Cγ resonances. This is shown in Fig. 4i which includes N_{i+1}-C'_i-$C\beta_i$/$C\gamma_i$ cross peaks (peaks 14 and 15). Figure 4f shows the use of N_i-C'_{i-1}-$C\alpha_i$ cross peaks in an NCOCX spectrum of the [1,3-^{13}C]-glycerol sample for linking arginine to a preceding Group I residue (i.e. an amino acid with 100% ^{13}C′-labelling).

Glutamine/glutamic acid

The labelling pattern of glutamine and glutamic acid resembles that of proline and arginine, but with Cβ and Cγ chemical shifts predominantly at ~28–30 ppm and ~33–36 ppm, respectively. Several cross peaks in the aliphatic region of the [1,3-^{13}C]-glycerol spectra are observed, but unfortunately most are in regions of heavy overlap. Cross peaks involving the Cδ resonance appear downfield from the backbone carbonyl signals, at around 184 ppm, but need to be treated with caution: in the spectra of αB-crystallin many of these cross peaks were found to be inter-residue cross peaks and were thus not suitable for spin system identification. N_i-$C\alpha_i$-$C\gamma_i$ peaks in the NCACX spectrum of the [1,3-^{13}C]-glycerol sample can sometimes help to resolve overlap problems and provide the nitrogen backbone chemical shift. However, they can form agglomerates with N_i-$C\alpha_i$-$C\beta_i$ cross peaks for larger proteins which contain many glutamines and glutamic acid residues. During the assignment of αB-crystallin, the use of the NCACB experiment (Pauli et al. 2001), employing DREAM for mixing (Verel et al. 1998), enabled identification of N_i-$C\alpha_i$-$C\beta_i$ cross peaks and thus the distinction between the Glu/Gln Cβ and Cγ resonances (Fig. 4g, j, peaks 11 and 17).


15N [ppm]
176 172
70 60 50 40 30 20
(a) I114 119.8
(b) S115 121.8
(c) S115 121.9
(d) S115 121.9
(e) S115 121.9
(f) R116 115.0
(g) R116 115.0
(h) R116 115.0
(i) E117 119.6
(j) E117 119.6
(k) E117 119.6
(l) F118 121.7
(m) F118 121.7
(n) F118 121.7
NCACX60
NCOCX35
NCOCX150
NCACX200
NCACX25
NCOCX75
NCACB2
NCACX25
NCOCX75
NCOCX150
NCACX60
I114Cα
I114Cβ
I114C'
S115Cα
S115C'
S115Cβ
R116Cα
R116Cβ
R116C'
R116Cγ
E117Cα
E117Cβ
E117C'
E117Cγ
13C [ppm]
I114
S115
R116
E117
F118

Sequential links can be obtained in a similar manner as for arginine due the identical labelling pattern. An instructive example is shown in Fig. 4i–m.

Threonine

The unique chemical shift of the Cβ resonance at $\sim$70 ppm offers a good starting point for the assignment of threonine residues, leading to a distinctive pattern involving the C′, Cα, Cβ and Cγ resonances, visible in the 2D ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol samples. The Cα-Cγ2 cross peaks occur predominantly in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample. The NCACX spectrum of the [1,3-^{13}C]-glycerol sample usually gives rise to both N_i-Cα_i-Cβ_i and N_i-Cα_i-C$\gamma 2_i$ cross peaks, allowing for easy identification of the backbone nitrogen chemical shifts. Alternatively, the $N_iC\alpha_i$ strip taken from the NCACX spectrum of the [2-^{13}C]-glycerol sample generally also contains N_i-Cα_i-C$\gamma 2_i$ peaks (Fig. S6d, peak 13).

Sequential assignment is best carried out using the $N_{i+1}C'_i$ strip in the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample. N_{i+1}-C′$_i$-Cβ_i cross peaks are relatively strong and in many cases (dependent upon the labelling of residue i+1) N_{i+1}-C′$_i$-Cβ_{i+1} cross peaks can also be identified (Fig. S6c, f). The uniqueness of the Cβ chemical shift means that 2D ^{13}C-^{13}C correlation spectra with longer mixing times of all three samples can be used to identify inter-residue correlations involving the threonine Cβ.

Asparagine/aspartic acid

Asparagine and aspartic acid have the same labelling pattern as threonine, but different chemical shifts. The Cγ resonances of asparagine and aspartic acid residues are slightly downfield shifted from the backbone carbonyl resonances. In the 2D ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol sample the Cβ-Cγ cross peaks are usually the only signals with chemical shifts of $\sim$39 ppm and $\sim$177 ppm (asparagine) or $\sim$180 ppm (aspartic acid). These cross peaks offer an obvious starting point for the assignment of these amino acids. In the ^{13}C-^{13}C correlation spectra of [1,3-^{13}C]-glycerol samples the Cα-Cβ cross peaks appear in a region that is not particularly crowded ($\sim$54 and $\sim$39 ppm), unless weak leucine Cα-Cβ cross peaks are visible (contrary to the expected labelling scheme) as was the case for OmpG. In the [2-^{13}C]-glycerol spectra, the Cα-Cγ cross peaks can sometimes be observed. The NCACX spectrum of the [1,3-^{13}C]-glycerol sample contains strong N_i-Cα_i-Cβ_i peaks which resolve asparagine or aspartic acid residues which are overlapped in the 2D spectra and offer a means of identifying the backbone nitrogen chemical shift.

Sequential links in the 3D spectra are similar to the threonine case (Fig. S5c–d, g–j), where the $N_{i+1}C'_i$ and N_i-C′$_{i-1}$ strips in the NCOCX spectrum of the [1,3-^{13}C]-glcyerol sample are particularly useful (unless preceded by a leucine).

Methionine

The methionine cross peak pattern in the 2D ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol samples resembles that of threonine and asparagine/aspartic acid (Fig. S4). A useful starting point is the intense Cα-Cγ cross peak in the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol samples. The Cε chemical shift can be obtained from the Cγ-Cε cross peaks in the ^{13}C-^{13}C correlation spectra of the [U-^{13}C] or [1,3-^{13}C]-glycerol samples. These spectra also contain the Cε-Cβ cross peaks which enable identification of the Cβ resonance. The nitrogen backbone resonance can be determined via the NCACX spectrum of the [2-^{13}C]-glycerol sample in which N_i-Cα_i-Cγ_i cross peaks are visible at $\sim$32 ppm (Fig. S6g, peak 19).

As for other Group II amino acids, sequential links can be found in all types of NCOCX spectra making use of both the $N_{i+1}C'_i$ and $N_iC'_{i-1}$ strips (Fig. S5c–d, g–j). Figure S6e–j shows how the methionine is linked to its neighbouring residues for the Glu22-Val23-Thr24-Met25-Lys26 motif in the α-spectrin SH3 domain.

Isoleucine

The unique chemical shift of the Cδ resonance of isoleucine residues ($\sim$13.5 ppm) provides a starting point for spin system identification. In the ^{13}C-^{13}C correlation of the [2-^{13}C]-glycerol sample spectra, the Cδ resonance gives rise to cross peaks involving the Cα, and Cβ resonances. In the spectra of the [1,3-^{13}C]-glycerol samples, correlations involving the Cγ1 and Cγ2 chemical shifts are observed. In spectra of OmpG, the intensities of cross peaks involving Cδ chemical shifts were low. In this case, the Cα-Cβ cross peaks in the spectra of the [2-^{13}C]-glycerol sample proved to be easily identifiable, together with other correlations to the Cβ resonances. Further cross peaks involving the Cγ2 signals occur in the spectra of the [1,3-^{13}C]-glycerol sample. Nitrogen backbone chemical shifts can be determined from N_i-Cα_i-Cβ_i cross peaks in the NCACX spectrum of the [2-^{13}C]-glycerol sample (Fig. 4a, peak 1) or from N_i-Cα_i-C$\gamma 2_i$ cross peaks in the spectra of the [1,3-^{13}C]-glycerol sample.

Sequential links can often be obtained via the [2-^{13}C]-glycerol spectra in which strong cross peaks involving the Cβ resonance are observed. For αB-crystallin, for example, the link between Ile114 and Ser115 is established via the NCOCX spectrum of the [2-^{13}C]-glycerol sample which

shows N_{i+1}-C'_i-$C\beta_i$ and N_{i+1}-C'_i-$C\alpha_{i+1}$ cross peaks (Fig. 4c, peaks 3 and 4). Similarly, in many cases cross peaks involving the isoleucine Cβ can be observed in the NCACX spectra of the [2-^{13}C]-glycerol sample, as was the case for the αB-crystallin Ile114-Ser115 motif (Fig. 4a, d; peaks 1, 6).

Lysine

The Cα-Cγ cross peak serves as a good starting point for lysine residues, due to its appearance in an uncrowded region of the ^{13}C-^{13}C correlation spectra of the [2-^{13}C]-glycerol sample. The Cγ-Cε peaks can then be located in the same spectrum and the Cγ-Cδ, Cβ-Cδ, Cβ-Cγ and Cβ-Cε peaks in the ^{13}C-^{13}C correlation spectra of the [1,3-^{13}C]-glycerol or [U-^{13}C] samples. The nitrogen backbone chemical shift can be identified from a weak N_i-$C\alpha_i$-$C\gamma_i$ peak in the NCACX spectrum of the [2-^{13}C]-glycerol sample.

Sequential links are once again most easily obtained using the NCOCX spectra of both the [1,3-^{13}C] and [2-^{13}C]-glycerol samples via identification of N_{i+1}-C'_i-$C\beta_i$ and N_{i+1}-C'_i-$C\alpha_{i+1}$ cross peaks, respectively (Fig. S5c–d, g–j). Links to the preceding residue depend on the labelling of C'_{i-1}. Figure S6i illustrates an example for the Glu22-Val23-Thr24-Met25-Lys26 motif in the α-spectrin SH3 domain.

Discussion

The assignment procedure presented here uses characteristic amino acid type specific patterns observed in spectra of [1,3-^{13}C]-glycerol, [2-^{13}C]-glycerol and [U-^{13}C] labelled samples to identify spin system types, link them sequentially and thus sequence specifically assign them. In addition, the [1,3-^{13}C]- and [2-^{13}C]-glycerol labelling helps to reduce signal overlap as well as improve spectral line widths. This method has been successfully applied to the assignment of the 175-residue protein αB-crystallin (Jehle et al. 2009). It is also being applied to the even larger, 281-residue protein, OmpG.

As for solution NMR assignment strategies, there are certain amino acid types which are easily assigned due to their special chemical shifts. These include glycine, alanine, threonine, serine and proline. On the basis of their glycerol labelling pattern, three further amino acids, valine, leucine and isoleucine, become easy to assign in the strategy presented here. Leucine is the only amino acid with a high degree of labelling at the C′ position in the [2-^{13}C]-glycerol sample, and thus yields strong and unambiguous correlations in the NCOCX spectrum of the [2-^{13}C]-glycerol sample. Isoleucine benefits not only from its downfield shifted Cα and upfield shifted Cδ1 chemical shifts, but more importantly, from the 100% ^{13}C-labelling of the Cβ atom in the [2-^{13}C]-glycerol sample. This results in very strong cross peaks in less crowded regions of the spectra of the [2-^{13}C]-glycerol sample. In the αB-crystallin spectra isoleucine was therefore very clearly visible and straightforward to assign. It is worth mentioning that only one carbon atom is labelled in glycine residues in the glycerol samples, therefore giving rise to many inter-residue cross peaks via relatively weak dipolar couplings. Even in the uniformly labelled sample many inter-residue correlations involving glycine are observed.

In contrast to these amino acids which are easily assigned, there are several others which present significant difficulties using only a uniformly labelled sample, mainly because their chemical shifts are not very distinctive and their cross peaks appear in crowded areas of the spectra. In this regard, glutamine, glutamic acid, argenine, lysine and methionine are rather difficult to assign. However, lysine and methionine, in particular, benefit from their labelling patterns in the spectra of the glycerol samples.

Not only does the amino acid composition have a significant effect on how easy the assignment process of a particular protein is likely to be, but the pairings of the amino acids is an equally important issue. A sequential row of Group I residues (alanine, cysteine, glycine, histidine, phenylalanine, serine, tryptophan and tyrosine) and valine can be readily linked sequentially using the NCOCX spectrum of the [1,3-^{13}C]-glycerol sample using N_i-C'_{i-1}-$C\beta_i$ and N_i-C'_{i-1}-$C\beta_{i-1}$ cross peaks. Strings of Group II residues, on the other hand, tend to be more difficult to deal with using only spectra of the glycerol-based samples. If the probability of the Cα or C′ sites being labelled is only about 50% in two neighbouring residues, then the likelihood of Cα and C′ atoms being labelled in both residues simultaneously is only around 25%. Thus the intensity of cross peaks between two Group II residues will be dramatically reduced compared to those between two Group I residues. Threonine and leucine can be assigned well almost regardless of the nature of their neighbouring amino acids: threonine because many assignments can be made from cross peaks involving the Cβ resonances in 2D ^{13}C-^{13}C correlation spectra which can easily be recorded with a high signal-to-noise ratio; leucine because it gives rise to excellent cross peaks in the [2-^{13}C]-glycerol NCOCX, very often with cross peaks to the Cα or Cβ of the following amino acid.

The assignment procedure outlined here is, of course, flexible and may be extended. For αB-crystallin, for example, the NCACB experiment (Pauli et al. 2001) was included into the suite of pulse sequences. This was found to be particularly helpful for identifying glutamine and glutamic acid spin systems. In the case of OmpG, the

assignment procedure was extended to include additional samples. The [U-^{13}C]-labelled sample of OmpG could only be used to a limited extent, because of the severe overlap. However, a reference sample containing uniformly labelled Group I amino acids is necessary in order to match up the signals from the two glycerol samples into complete spin systems. Therefore, several additional samples were produced in which only a small number of amino acids were fully or [2,3-^{13}C]-labelled, e.g. by mixing [U-^{13}C]-glycine and alanine with [2,3-^{13}C]-phenylalanine and tyrosine (OmpG-GAFY; Hiller et al. 2008) or using [U-^{13}C]-glycine, alanine, valine, leucine and serine (OmpG-GAVLS). Furthermore, this assignment procedure can be extended to samples which suffer from broad lines or low signal intensity (if the amount of sample is limited, for instance).

It has already been shown that the assignment of CA150 amyloid fibril samples was significantly aided by the use of [1,3-^{13}C]- and [2-^{13}C]-glycerol samples (Becker et al. 2008). The spectra of CA150 amyloid fibrils suffer from broad lines, especially in the nitrogen dimension, and the advantage offered by the glycerol samples also included an improvement of the line width. Similar results have been seen for neurotoxin II from Asian cobra snake venom bound to the nicotinic acetylcholine receptor (Krabben et al. 2009). The main limitation in this case was the amount of sample available, since much of the rotor was taken up with lipids and the receptor, and only around 0.4 mg of labelled, bound neurotoxin was present in each sample.

Other labelling schemes used to obtain resonance assignments in the solid state include the use of [1,4-], [2,3-] and [1,2,3,4-^{13}C] succinic acid based samples (van Gammeren et al. 2004; van Gammeren et al. 2005) and reverse labelling which yielded samples in which key hydrophobic residues were left unlabelled (Etzkorn et al. 2007; Schneider et al. 2008). The reverse labelling approach has been applied to two membrane proteins (Etzkorn et al. 2007; Schneider et al. 2008) and has the advantage of removing a large number of overlapping resonances from the hydrophobic transmembrane helices. Thus the overlap problem is addressed and extensive assignment is made possible. However, in order to address the hydrophobic transmembrane stretches the assignment protocol presented here could be applicable. Other labelling schemes, such as fractional [U-^{13}C]-glucose labelling (Neri et al. 1989; Schubert et al. 2006) or labelling with [1-^{13}C]-glucose (Hong 1999) or [2-^{13}C]-glucose (Lundström et al. 2007) could potentially form alternatives to labelling with [1,3-^{13}C] and [2-^{13}C]-glycerol, and may be well suited to proteins with particular amino acid type compositions.

It is worth noting that in our experience the exact isotopomer distribution for the [1,3-^{13}C] and [2-^{13}C]-glycerol samples appears to vary slightly from protein to protein. The isotopomer distributions discussed here are those found for SH3 as verified by solution NMR. But for OmpG, for instance, additional valine and leucine isotopomers were observed compared to those found for SH3. We attribute this to different amino acid compositions in the proteins leading to differences in the *E.coli* amino acid catabolism and metabolism.

Assignment following the method outlined here would benefit from special features in the assignment software used. Many software packages will typically offer possible assignment options for a selected peak. When using the [1,3-^{13}C]- and [2-^{13}C]-glycerol samples, however, not all assignment options of a [U-^{13}C] sample remain possible, and for the amino acids with mixed labelling the assignment possibilities become quite involved. For instance, a threonine Cα-Cβ cross peak in a [1,3-^{13}C]-glycerol spectrum could be both intra- or inter-residue, but in a [2-^{13}C]-glycerol spectrum it can only be an inter-residue cross peak. Or a peak at 60 ppm, which typically might belong to a valine Cα, cannot do so in the [1,3-^{13}C] sample. Such modules were recently incorporated into the CCPNmr Analysis software package (Vranken et al. 2005).

In the longer term, larger proteins will benefit from solution-like assignment strategies using deuterated samples (Zhou et al. 2007a; Linser et al. 2008). So far, such methods have only been applied to microcrystalline protein samples of GB1 (Zhou et al. 2007a) and SH3 (Linser et al. 2008) and further developments may be necessary to be able to extend this methodology to large proteins or those which give rise to lower quality spectra. There is no doubt that this would be a great advantage for the assignment of large proteins by solid-state MAS NMR in the future. It would enable both the amide group nitrogen and hydrogen chemical shifts to be used as an 'anchor' for each spin system, as is the case in solution NMR. This is currently one of the key elements missing for ssNMR assignment because the nitrogen chemical shift alone, on account of its low chemical shift dispersion, is not able to fulfil this role.

Materials and methods

The SH3 domain from chicken α-spectrin (Castellani et al. 2002), αB-crystallin (Jehle et al. 2009) and OmpG (Hiller et al. 2005) were expressed and purified as described previously. [U-^{13}C], [1,3-^{13}C]-glycerol and [2-^{13}C]-glycerol labelled samples were prepared for each protein. In addition, samples of OmpG with selectively labelled amino acids were produced. A sample in which glycine and alanine were uniformly [^{13}C,^{15}N] labelled and phenylalanine and tyrosine were [2,3-^{13}C,^{15}N] labelled (OmpG-GAFY) was prepared as described previously (Hiller et al. 2008). A

sample in which glycine, alanine, valine, leucine and serine were uniformly [^{13}C,^{15}N] labelled and all other amino acids were uniformly [^{15}N] labelled was produced following the same protocol, but expressing the protein using a fermenter (Fiedler et al. 2007).

All SH3 2D ^{13}C-^{13}C correlation spectra were collected on a Bruker DMX-750 narrow bore spectrometer equipped with a 4-mm double-resonance CP/MAS probe. 50 ms of PDSD mixing were used for the [2-^{13}C] and [1,3-^{13}C]-glycerol labelled samples and 10 ms PDSD mixing for the uniformly labelled sample. 2D and 3D NCACX and NCOCX spectra were collected for the glycerol samples on a Bruker DMX-400 spectrometer, equipped with a 4-mm triple-resonance CP/MAS probe. PDSD mixing times of 50, 200 and 500 ms were used. All experiments were performed at 280 K.

The ^{13}C-^{13}C correlation spectra, NCACX, NCOCX and NCACB spectra recorded for αB-crystallin have been described previously (Jehle et al. 2009).

^{13}C-^{13}C correlation spectra of OmpG were recorded on a Bruker narrow bore 900 MHz spectrometer equipped with a 3.2-mm triple resonance CP/MAS probe and using 13 kHz MAS. DARR mixing for a duration of 50 ms, 200 ms and 400 ms was used for the glycerol samples. 25 ms of PDSD mixing was used for the uniformly labelled sample. 20 ms and 700 ms DARR mixing was used for the OmpG-GAFY and OmpG-GAVLS samples. The B_1 fields during the ^{1}H-^{13}C CP step (1.5 ms) were typically around 55 kHz and 65 kHz for ^{1}H and ^{13}C, respectively. 3D NCACX and NCOCX spectra were recorded on a Bruker DMX-400 wide bore spectrometer equipped with a 4-mm triple-resonance CP/MAS probe using a MAS speed of 8 kHz. The following spectra were recorded: [1,3-^{13}C]-glycerol: NCACX-PDSD (100 ms), NCOCX-DARR (100, 200 ms); [2-^{13}C]-glycerol: NCACX-PDSD (100 ms), NCOCX-DARR (50, 100 ms); OmpG-GAFY: NCACX-PDSD (20 ms), NCACX-DARR (300, 500 ms), 2D NCOCX-PDSD (50 ms); OmpG-GAVLS: NCACX-DARR (20 ms), NCOCX-DARR (30 ms). SPINAL64 decoupling with a power level of around 90 kHz was used for all spectra and a temperature of 280 K was used throughout.

All data were processed with Topspin versions 1.3 and 2.1 (Bruker, Karlsruhe, Germany). Spectra were analysed and annotated using SPARKY version 3.110 (T.D. Goddard and D.G. Kneller, University of California, San Fransisco) and CCPNmr Analysis (Vranken et al. 2005).

Acknowledgments This work was supported by the Deutsche Forschungsgemeinschaft (SFB449). J.F. gratefully acknowledges funding from the National Science Foundation International Research Fellowship Program (Award #0402114).

References

Agarwal V, Reif B (2008) Residual methyl protonation in perdeuterated proteins for multi-dimensional correlation experiments in MAS solid-sate NMR spectroscopy. J Mag Res 194:16–24

Becker J, Ferguson N, Flinders J, van Rossum BJ, Fersht AR, Oschkinat H (2008) A sequential assignment procedure for proteins that have intermediate line widths in MAS NMR spectra: amyloid fibrils of human CA150.WW2. ChemBiochem 9:1946–1952

Bennett AE, Ok JH, Griffin RG, Vega S (1992) Chemical-shift correlation spectroscopy in rotating solids: radio frequency-driven dipolar decoupling and longitudinal exchange. J Chem Phys 96:8624–8627

Bennett AE, Rienstra CM, Griffiths JM, Zhen WG, Lansbury PT, Griffin RG (1998) Homonuclear radio frequency-driven recoupling in rotating solids. J Chem Phys 108:9463–9479

Bloembergen N (1949) On the interaction of nuclear spins in a crystalline lattice. Physica 15:386–426

Böckmann A, Lange A, Galinier A, Luca S, Giraud N, Juy M, Heise H, Montserret R, Penin F, Baldus M (2003) Solid state NMR sequential resonance assignments and conformational analysis of the 2 × 10.4 kDa dimeric form of the *Bacillus subtilis* protein Crh. J Biomol NMR 27:323–339

Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. Nature 420:98–102

Castellani F, van Rossum BJ, Diehl A, Rehbein K, Oschkinat H (2003) Determination of solid-state NMR structures of proteins by means of three-dimensional N-15-C-13-C-13 dipolar correlation spectroscopy and chemical shift analysis. Biochemistry 42:11476–11483

De Paëpe G, Lewandowski JR, Loquet A, Böckmann A, Griffin RG (2008) Proton assisted recoupling and protein structure determination. J Chem Phys 129:245101-11–245101-21

Etzkorn M, Martell S, Andronesi OC, Seidel K, Engelhard M, Baldus M (2007) Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. Angew Chem Int Ed 46:459–462

Ferguson N, Becker J, Tidow H, Tremmel S, Sharpe TD, Krause G, Flinders J, Petrovich M, Berriman J, Oschkinat H, Fersht AR (2006) General structural motifs of amyloid protofilaments. Proc Natl Acad Sci USA 103:16248–16253

Fiedler S, Knocke C, Vogt J, Oschkinat H, Diehl A (2007) HCDF as a protein-labeling methodology—production of H-2-, C-13-, and N-15-labeled OmpG via high cell density fermentation. Gen Eng Biotech News 27:54

Franks WT, Zhou DH, Wylie BJ, Money BG, Graesser DT, Frericks HL, Sahota G, Rienstra CM (2005) Magic-angle spinning solid-state NMR spectroscopy of the beta 1 immunoglobulin binding domain of protein G (GB1): N-15 and C-13 chemical shift assignments and conformational analysis. J Am Chem Soc 127:12291–12305

Franks WT, Wylie BJ, Schmidt HLF, Nieuwkoop AJ, Mayrhofer RM, Shah GJ, Graesser DT, Rienstra CM (2008) Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR. Proc Natl Acad Sci USA 105:4621–4626

Goldbourt A, Gross BJ, Day LA, McDermott AE (2007) Filamentous phage studied by magic-angle spinning NMR: resonance assignment and secondary structure of the coat protein in Pf1. J Am Chem Soc 129:2338–2344

Gullion T, Schaefer J (1989) Rotational-echo double-resonance NMR. J Mag Res 81:196–200

Hiller M, Krabben L, Vinothkumar KR, Castellani F, van Rossum BJ, Kühlbrandt W, Oschkinat H (2005) Solid-state magic-angle spinning NMR of outer-membrane protein G from *Escherichia coli*. ChemBiochem 6:1679–1684

Hiller M, Higman VA, Jehle S, van Rossum BJ, Kühlbrandt W, Oschkinat H (2008) 2,3-C-13-labeling of aromatic residues – getting a head start in the magic-angle-spinning NMR assignment of membrane proteins. J Am Chem Soc 130:408–409

Hohwy M, Jakobsen HJ, Eden M, Levitt MH, Nielsen NC (1998) Broadband dipolar recoupling in the nuclear magnetic resonance of rotating solids: a compensated C7 pulse sequence. J Chem Phys 108:2686–2694

Hong M (1999) Determination of multiple phi-torsion angles in proteins by selective and extensive C-13 labeling and two-dimensional solid-state NMR. J Mag Res 139:389–401

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

Igumenova TI, McDermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ (2004a) Assignments of carbon NMR resonances for microcrystalline ubiquitin. J Am Chem Soc 126:6720–6727

Igumenova TI, Wand AJ, McDermott AE (2004b) Assignment of the backbone resonances for microcrystalline ubiquitin. J Am Chem Soc 126:5323–5331

Jehle S, van Rossum B, Stout JR, Noguchi SM, Falber K, Rehbein K, Oschkinat H, Klevit RE, Rajagopal P (2009) αB-crystallin: a hybrid solid-state/solution-state NMR Investigation reveals structural aspects of the heterogeneous oligomer. J Mol Biol 385:1481–1497

Krabben L, van Rossum B-J, Jehle S, Bocharov E, Lyukmanova EN, Schulga AA, Arseniev A, Hucho F, Oschkinat H (2009) Loop 3 of short neurotoxin II is an additional interaction site with membrane-bound nicotinic acetylcholine receptor as detected by solid-state NMR spectroscopy. J Mol Biol 390:662–671

Lange A, Becker S, Seidel K, Giller K, Pongs O, Baldus M (2005) A concept for rapid protein-structure determination by solid-state NMR spectroscopy. Angew Chem Int Ed 44:2089–2092

Lee YK, Kurur ND, Helmle M, Johannessen OG, Nielsen NC, Levitt MH (1995) Efficient dipolar recoupling in the NMR of rotating solids—a sevenfold symmetrical radiofrequency pulse sequence. Chem Phys Lett 242:304–309

LeMaster DM, Kushlan DM (1996) Dynamical mapping of *E-coli* thioredoxin via C-13 NMR relaxation analysis. J Am Chem Soc 118:9255–9264

Lewandowski JR, De Paëpe G, Griffin RG (2007) Proton assisted insensitive nuclei cross polarization. J Am Chem Soc 129:728–729

Li Y, Berthold DA, Frericks HL, Gennis RB, Rienstra CM (2007) Partial C-13 and N-15 chemical-shift assignments of the disulfide-bond-forming enzyme DsbB by 3D magic-angle spinning NMR spectroscopy. ChemBiochem 8:434–442

Li Y, Berthold DA, Gennis RB, Rienstra CM (2008) Chemical shift assignment of the transmembrane helices of DsbB, a 20-kDa integral membrane enzyme, by 3D magic-angle spinning NMR spectroscopy. Prot Sci 17:199–204

Linser R, Fink U, Reif B (2008) Proton-detected scalar coupling based assignment strategies in MAS solid-state NMR spectroscopy applied to perdeuterated proteins. J Mag Res 193:89–93

Loquet A, Bardiaux B, Gardiennet C, Blanchet C, Baldus M, Nilges M, Malliavin T, Böckmann A (2008) 3D Structure determination of the Crh protein from highly ambiguous solid-state NMR restraints. J Am Chem Soc 130:3579–3589

Lundström P, Teilum K, Carstensen T, Bezsonova I, Wiesner S, Hansen DF, Religa TL, Akke M, Kay LE (2007) Fractional C-13 enrichment of isolated carbons using 1-C-13- or 2-C-13-glucose facilitates the accurate measurement of dynamics at backbone C-alpha and side-chain methyl positions in proteins. J Biomol NMR 38:199–212

Marulanda D, Tasayco ML, McDermott A, Cataldi M, Arriaran V, Polenova T (2004) Magic angle spinning solid-state NMR spectroscopy for structural studies of protein interfaces. Resonance assignments of differentially enriched *Escherichia coli* thioredoxin reassembled by fragment complementation. J Am Chem Soc 126:16608–16620

Neri D, Szyperski T, Otting G, Senn H, Wuthrich K (1989) Stereospecific nuclear magnetic-resonance assignments of the methyl-groups of valine and leucine in the DNA-binding domain of the 434-repressor by biosynthetically directed fractional C-13 labeling. Biochemistry 28:7510–7516

Pauli J, Baldus M, van Rossum B, de Groot H, Oschkinat H (2001) Backbone and side-chain C-13 and N-15 signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. ChemBiochem 2:272–281

Schneider R, Ader C, Lange A, Giller K, Hornig S, Pongs O, Becker S, Baldus M (2008) Solid-state NMR spectroscopy applied to a chimeric potassium channel in lipid bilayers. J Am Chem Soc 130:7427–7435

Schubert M, Manolikas T, Rogowski M, Meier BH (2006) Solid-state NMR spectroscopy of 10% ^{13}C labeled ubiquitin: spectral simplification and sterospecific assignment of isopropyl groups. J Biomol NMR 35:167–173

Takegoshi K, Nakamura S, Terao T (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 344:631–637

Takegoshi K, Nakamura S, Terao T (2003) C-13-H-1 dipolar-driven C-13-C-13 recoupling without C-13 rf irradiation in nuclear magnetic resonance of rotating solids. J Chem Phys 118:2325–2341

van Gammeren AJ, Hulsbergen FB, Hollander JG, de Groot HJM (2004) Biosynthetic site-specific C-13 labeling of the light-harvesting 2 protein complex: a model for solid state NMR structure determination of transmembrane proteins. J Biomol NMR 30:267–274

van Gammeren AJ, Hulsbergen FB, Hollander JG, de Groot HJM (2005) Residual backbone and side-chain C-13 and N-15 resonance assignments of the intrinsic transmembrane light-harvesting 2 protein complex by solid-state magic angle spinning NMR spectroscopy. J Biomol NMR 31:279–293

Verel R, Baldus M, Ernst M, Meier BH (1998) A homonuclear spin-pair filter for solid-state NMR based on adiabatic-passage techniques. Chem Phys Lett 287:421–428

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

Zech SG, Wand AJ, McDermott AE (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. J Am Chem Soc 127:8618–8626

Zhou DH, Shah G, Cormos M, Mullen C, Sandoz D, Rienstra CM (2007a) Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning. J Am Chem Soc 129:11791–11801

Zhou DH, Shea JJ, Nieuwkoop AJ, Franks WT, Wylie BJ, Mullen C, Sandoz D, Rienstra CM (2007b) Solid-rate protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. Angew Chem Int Ed 46:8380–8383