J Biomol NMR (2015) 62:281–290
DOI 10.1007/s10858-015-9943-6

ARTICLE

Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps

Subrata H. Mishra[1] · **Dominique P. Frueh**[1]

Received: 17 March 2015 / Accepted: 28 April 2015 / Published online: 8 May 2015

Abstract Methyl groups have become key probes for structural and functional studies by nuclear magnetic resonance. However, their NMR signals cluster in a small spectral region and assigning their resonances can be a tedious process. Here, we present a method that facilitates assignment of methyl resonances from assigned amide groups. Calculating the covariance between sensitive methyl and amide 3D spectra, each providing correlations to C^{α} and C^{β} separately, produces 4D correlation maps directly correlating methyl groups to amide groups. Optimal correlation maps are obtained by extracting residue-specific regions, applying derivative to the dimensions subject to covariance, and multiplying 4D maps stemming from different 3D spectra. The latter procedure rescues weak signals that may be missed in traditional assignment procedures. Using these covariance correlation maps, nearly all assigned isoleucine, leucine, and valine amide resonances of a 52 kDa nonribosomal peptide synthetase cyclization domain were paired with their corresponding methyl groups.

Keywords NMR resonance assignment · Large proteins · Methyl groups · Nonribosomal peptide synthetases

Electronic supplementary material The online version of this article (doi:10.1007/s10858-015-9943-6) contains supplementary material, which is available to authorized users.

✉ Dominique P. Frueh
dfrueh@jhmi.edu

[1] Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 701 Hunterian, 725 North Wolfe St., Baltimore, MD 21205, USA

Introduction

Nuclear magnetic resonance (NMR) has become a method of choice for molecular studies of proteins of ever increasing sizes, and efforts at overcoming the limitations of the technique for large proteins have been spurred by the wealth of information that NMR provides at an atomic level. In particular, strategies that rely on detection of methyl groups have been keys to recent successes since they help overcome the low sensitivity of NMR experiments for large proteins for many reasons. First, three methyl protons contribute to the amplitude of each methyl group signal, which increases the sensitivity of related experiments. Second, so-called TROSY methods can be designed to combat relaxation (Tugarinov et al. 2003; Tugarinov and Kay 2005), the process by which NMR signals are broadened and reduced in intensity for large proteins. Finally, to detect methyl resonances and further minimize relaxation, proteins can be uniformly deuterated with the exception of a subset of methyls that are selectively protonated, e.g. isoleucine (δ1-only), leucine and valine, as used in this work (Gardner and Kay 1997; Goto et al. 1999). Such samples have permitted kinetic, thermodynamic, binding and structural studies of various large biological systems such as the 204 kDa translocase SecA (Gelis et al. 2007; Gouridis et al. 2009), 50 kDa trigger factor chaperone molecules (Saio et al. 2014), the 670 kDa proteasome (Ruschak et al. 2010; Sprangers and Kay 2007), and a 37 kDa di-domain excised from a non-ribosomal peptide synthetase (NRPS) (Frueh et al. 2008). The protein discussed in this article, a 52 kDa cyclization domain (Cy1), also belongs to an NRPS system. NRPSs employ a modular, multi-domain organization to select and condense various substrates into important microbial secondary metabolites that oftentimes have pharmaceutical applications (Fischbach and Walsh 2006; Mootz et al. 2002; Walsh 2004). Recent studies indicate that NRPSs are

subject to inter- (Frueh et al. 2008; Tanovic et al. 2008) and intra-domain (Bloudoff et al. 2013; Bruner et al. 2002; Gulick 2009; Koglin et al. 2006) dynamics, and hence NMR is the preferred method for structural and functional studies of Cy1. The size of this domain warrants a selective methyl labelling strategy. However, methyl NMR resonances must be assigned prior to undertaking such studies, and this task is demanding for large proteins. For proteins up to 25–30 kDa, experiments exist to *directly* correlate unassigned methyl signals with assigned amide resonances, so methyl resonances are readily assigned (Grzesiek and Bax 1993). However, for larger proteins, transverse relaxation deteriorates the sensitivity of these experiments, and a suite of alternative experiments is required. In this case, resonances of amide and methyl groups are each separately correlated to C^{α} and C^{β} carbons. Thus, methyl resonances are paired to assigned amide resonances in an indirect manner, by identifying correlations to C^{α} and C^{β} carbons that are common to amide and methyl 3D spectra (Sprangers and Kay 2007; Tugarinov and Kay 2003). Successful assignment of methyl resonances relies either on accurate peak picking of several spectra or on a careful visual inspection of the data sets (vide infra). Unfortunately, both approaches can be time-consuming and prone to human error, and assignment of methyl resonances remains a major bottleneck in NMR studies of large proteins.

Here, we present a processing method that employs the most sensitive spectra and provides a direct correlation between assigned amide resonances and unassigned methyl resonances. Using covariance NMR (Blinov et al. 2005; Brüschweiler 2004; Brüschweiler and Zhang 2004; Chen et al. 2010; Kupce and Freeman 2006; Lescop and Brutscher 2007; Snyder and Brüschweiler 2009; Trbovic et al. 2004; Zhang and Bruschweiler 2004), indirect correlations to C^{α} and C^{β} nuclei are translated into direct correlations between methyl and amide groups (Fig. 1). Many strategies are employed to produce reliable correlation maps that are easy to analyze. False correlations resulting from spectral crowding in the original spectra can be minimized by extracting C^{α} and C^{β} spectral regions specific to each residue. Similarly, tailoring amide and methyl spectral regions to residue types provides access to otherwise prohibitively large 4D correlation maps. False correlations due to residues with similar but non-identical C^{α} and C^{β} signals are minimized by taking spectral derivatives of original spectra prior to calculating covariance (Harden et al. 2014). Finally, the information provided separately by C^{α} and C^{β} nuclei is combined by multiplying two 4D arrays. In the end, we provide a specific 4D correlation map for each residue type (e.g. Ile) from four distinct conventional 3D spectra. With these maps, simple inspection of a two-dimensional correlation map reveals the methyl resonances corresponding to an assigned amide signal. We demonstrate this method on the 52 kDa NRPS cyclization domain, Cy1.

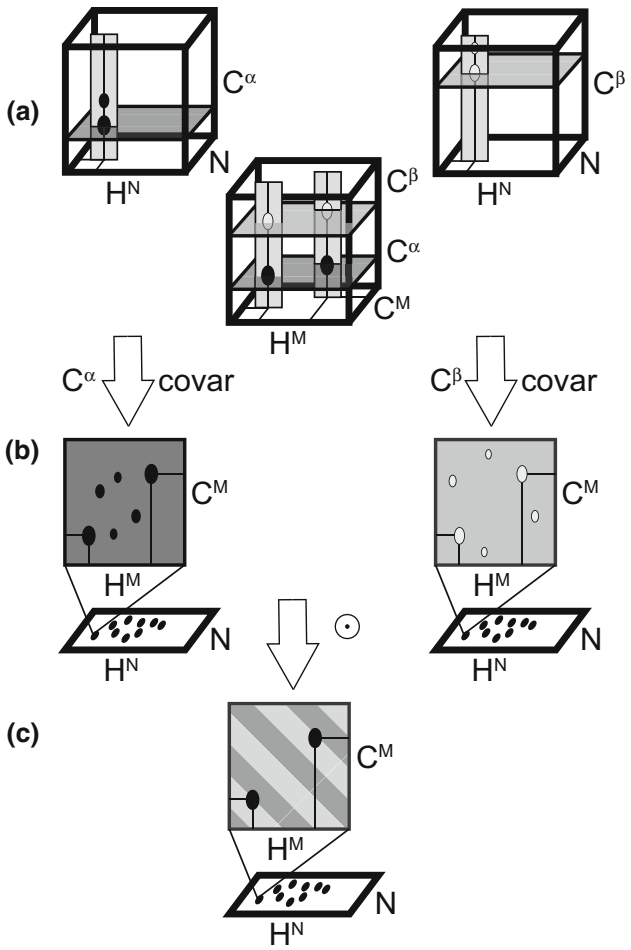

Fig. 1 Generation of covariance correlation maps for assigning methyl resonances. **a** For a residue with given C^{α} and C^{β} frequencies, H^N/N planes in HNCA (*top left*, *dark grey*) and HN(CA)CB (*top right*, *light grey*) feature the (H^N, N) correlation of that residue; the corresponding H^M/C^M planes in HMCM experiments feature (H^M, C^M) correlations for that residue. Simultaneous inspection of H^N/N and H^M/C^M may lead to assignment of methyl signals but the process is tiresome. **b** Calculating the covariance between H^N/C^{α} and H^M/C^{α} planes and between H^N/C^{β} and H^M/C^{β} planes produces two 4D-arrays correlating methyl and amide groups. Tailoring the process to residue-types produces reliable and practicable 4Ds. **c** The direct product between both arrays produces a 4D-array that conveys the information obtained through both C^{α} and C^{β} signals and for all spectra in **a**. An H^M/C^M plane at the (H^N, N) coordinates of the residue will reveal its methyl resonances

Materials and methods

Cloning, expression and purification of Cy1

All NMR experiments were recorded on a 52 kDa monomeric protein, a cyclization domain (Cy1) excised from the Yersiniabactin non-ribosomal peptide synthetase (Keating et al. 2000). Cy1 was cloned into the pET30a expression vector to create a construct that expressed Cy1 with LEHHHHHH appended to the C-terminus (Cy1H6)

(Mishra et al. 2014). Two Cy1H6 NMR samples were used for data acquisition in our studies; one with uniform 2H, ^{15}N and ^{13}C labeling (CDN) and the other with the methyl side-chains of Leu, Val and Ile (δ1 position only) labeled with ^{1}H and ^{13}C (ILV) in an otherwise uniform ^{2}H, ^{15}N and ^{13}C labeling. For both labeling schemes Cy1H6 was expressed and purified as discussed previously (Mishra et al. 2014) with one minor modification for the ILV sample. For the ILV sample, sodium salts of $^{13}C_4$ α-ketobutyric acid (75 mg, CIL, CDLM-4611) and $^{13}C_5$ α-ketoisovaleric acid (125 mg, CIL, CDLM4418) were added to the expression media prior to induction at O.D.600 ~ 0.5. The NMR buffer for both samples was 20 mM sodium phosphate pH 7.0, 10 mM NaCl, 1 mM EDTA, 5 mM DTT and 5 % D_2O. The protein concentrations for the CDN and ILV NMR samples were 650 and 640 μM respectively.

NMR data acquisition

All NMR experiments were recorded at 25 °C on an 800 MHz Varian spectrometer equipped with a chiliprobe®. The ILV sample was used to record the HNCA, HMCMCGCB, HMCM(CG)CBCA, and HC(C)H-TOCSY datasets while the HN(CA)CB was acquired on the CDN sample. The HNCA was recorded with spectral widths of 20 ppm for ^{1}H (1200 complex points, carrier at 4.758 ppm), 27.5 ppm for ^{13}C (60 complex points, carrier at 58 ppm), and 35 ppm for ^{15}N (48 complex points, carrier at 118 ppm). Eight scans were accumulated with a recycling delay of 1 s. The HN(CA)CB was recorded with spectral widths of 18.76 ppm for ^{1}H (825 complex points, carrier at 4.758 ppm), 62 ppm for ^{13}C (56 complex points, carrier at 44 ppm), and 34 ppm for ^{15}N (29 complex points, carrier at 118 ppm). 16 scans were accumulated with a recycling delay of 1 s.

Both HMCMCGCB and HMCM(CG)CBCA pulse sequences were kindly provided by Lewis E. Kay, University of Toronto, Canada. The HMCMCBCA and the HMCM (CG)CBCA pulse phases were updated for samples with uniform ^{13}C labeling as recommended by Tugarinov and Kay (2003). The HMCMCGCB and HMCM(CG)CBCA were both recorded with spectral widths of 16 ppm for ^{1}H (832 complex points, carrier at 4.761 ppm), 20 ppm for ^{13}C methyl (48 complex points, carrier at 41.5 ppm), and 65 ppm for ^{13}C aliphatic (50 complex points, carrier at 41.5 ppm), and a recycling delay of 1 s. 16 scans were accumulated for HMCMCGCB and 32 scans were accumulated for HMCM(CG)CBCA. The HC(C)H-TOCSY was recorded with spectral widths of 17.5 ppm for ^{1}H direct dimension (910 complex points, carrier at 4.773 ppm), 3 ppm for ^{1}H methyl (50 complex points, carrier at 0.5 ppm), and 21 ppm for ^{13}C methyl (65 complex points, carrier at 19 ppm). Eight scans were accumulated with a recycling delay of 1 s. A 3D time-shared NOESY experiment (Frueh et al. 2008) was used to record the NOESY-HC-HSQC and NOESY-HN-TROSY spectra on a 750 μM sample with the methyls of Leu, Val and Ile (δ1 position only) labeled with ^{1}H and ^{13}C (ILV), the sidechains of Phe and Tyr protonated, and an otherwise uniform ^{2}H, ^{15}N and ^{12}C labeling (Mishra et al. 2014). The NOESY-HC-HSQC was acquired with spectral widths of 16.88 ppm for ^{1}H direct dimension (877 complex points, carrier at 4.761 ppm), 13.75 ppm for ^{1}H indirect dimension (165 complex points, carrier at 4.761 ppm), and 21 ppm for ^{13}C (30 complex points, carrier at 19 ppm). The mixing time was set to 150 ms and eight scans were accumulated with a recycling delay of 1 s.

Processing of NMR spectra and covariance correlation maps

All dimensions were apodized with a cosine squared bell function. Here, HMCMCGCB and HMCM(CG)CBCA are referred to as HMCM experiments whereas HNCA and HNCACB are denoted by HN experiments. 4D arrays relying on C^{β} or C^{α} signals in original spectra are denoted by $H^{N}NH^{M}C^{M}$cb and $H^{N}NH^{M}C^{M}$ca, respectively. For successful calculation of covariance matrices, it is critical that signals observed in $C^{\alpha/\beta}$ dimensions of HMCM and HN experiments coincide. To account for differences in spectral widths, each spectrum has to be appropriately zero-filled, and regions must be extracted so that signal maxima occur at the same coordinate (in points). Of note, isotopic shifts, if present, may be accommodated by matching spectra differently for each residue-type and nucleus (see below). For successful combination of $H^{N}NH^{M}C^{M}$ca and $H^{N}NH^{M}C^{M}$cb, the amide nitrogen and proton dimensions of HNCA and HN(CA)CB must be matched and similarly those of methyl carbons and protons in HMCMCGCB and HMCM(CG)CBCA. These constraints are independent of residue-specific regions, which are defined below.

The HNCA and HN(CA)CB nitrogen dimensions were zero-filled to 102 and 100 points, respectively, and their proton dimensions were zero-filled to 2184 and 2048 points, respectively. 473 points were extracted between 10.42 and 6.10 ppm in the amide proton dimensions. 100 points were extracted in the ^{15}N dimensions. The methyl proton dimensions in HMCMCGCB and HMCM(CG)CBCA were first zero-filled to 2048 points, of which 269 points were extracted between 1.47 and −0.63 ppm. The methyl carbon dimensions were zero-filled to 100 points. 33 points were extracted between 15.52 and 9.12 ppm for Ile, and 61 points were extracted between 28.86 and 16.85 ppm for Val and Leu. The ^{13}C aliphatic dimensions were zero-filled to 256 (HNCA), 578 (HN(CA)CB), and 606 (HMCMCGCB and HMCM(CG)CBCA) points. A circular shift was applied to HNCA to account for differences in carrier offsets and spectral widths between HNCA and HMCM experiments.

HN(CA)CB was recorded with the CDN sample instead of the ILV sample, and different circular shifts were applied for different residue-types to account not only for differences in carrier offsets and spectral widths but also for isotopic shifts. To determine the values of the circular shifts, C^{β} signals were inspected separately for each residue type, and we verified that C^{β} signal maxima in HN(CA)CB coincided with those of HMCM experiments (again, looking at the coordinates in points). If possible, we advise to record all experiments with the same ILV-sample, especially if high resolution is achieved in ^{13}C. Here, circular shifts were sufficient to account for isotopic shifts and produce reliable correlation maps, as evidenced throughout the article. Following these adjustments and following calibration, the residue-specific regions for C^{β} were [31.23—40.77 ppm] (90 points) for Ile, [36.01—48.35 ppm] (116 points) for Leu, and [27.29—38.02 ppm] (101 points) for Val. Those for C^{α} were [56.67—67.39 ppm] (101 points) for Ile, [47.33—60.73 ppm] (126 points) for Leu, and [54.62–69.64 ppm] (141 points) for Val. Thus, 12 spectra were produced (we use our shorthand notation of HMCM and HN experiments, see above): IB-HMCM, LB-HMCM, VB-HMCM, IB-HN, LB-HN, and VB-HN contain residue-specific regions for C^{β} and IA-HMCM, LA-HMCM, VA-HMCM, IA-HN, LA-HN, and VA-HN contain residue-specific regions for C^{α}. All 3D spectra were stored as arrays of 2D $H/C^{\alpha/\beta}$ and $H^{M}/C^{\alpha/\beta}$ planes ($C^{\alpha/\beta} = C^{\alpha}$ or C^{β}), with 100 planes spanning the ^{15}N dimensions and either 33 (Ile) or 66 (Leu, Val) planes spanning the ^{13}C methyl dimensions.

The 4D covariance correlation maps were generated with an in-house MATLAB (Matlab Reference Guide 1992) script (available upon contacting the corresponding author). Briefly, (1) users define the location of IA-HMCM, IA-HN, IB-HMCM, and IB-HN spectra, for example. (2) For each nitrogen and for each methyl carbon point, H^{N}/C^{α} and H^{M}/C^{α} as well as H^{N}/C^{β} and H^{M}/C^{β} planes are read. (3) The covariance matrix between H^{N}/C^{α} and H^{M}/C^{α} planes as well as that between H^{N}/C^{β} and H^{M}/C^{β} planes are calculated with the covarnmr toolbox (Short et al. 2011), thus creating H^{N}/H^{M}ca and H^{N}/H^{M}cb 2D covariance matrices. (4) The element-wise product between H^{N}/H^{M}ca and H^{N}/H^{M}cb is calculated to provide H^{N}/H^{M}cacb planes. (5) Each plane is added to a user-defined output folder, thus building the 4D array $H^{N}NH^{M}C^{M}$cacb. 3D projections of the 4D array are calculated on the fly (Harden et al. 2014) and facilitate frequency referencing. Defining input spectra dictates how many maps are combined by the element-wise product. Thus, $H^{N}NH^{M}C^{M}$cb is calculated by only defining the location of C^{β} spectra, whereas $H^{N}NH^{M}C^{M}$cacb requires the definition of both C^{α} and C^{β} spectra. Calculation times were estimated in MATLAB 2015a on a 2011 iMac with a 3.44 GHz Intel Core i7 8 GB quad core processor. $H^{N}NH^{M}C^{M}$ca took 46, 98, and 89 s for I, L, and V, respectively. $H^{N}NH^{M}C^{M}$cb took 46, 98, and 85 s for I, L, and V, respectively. $H^{N}NH^{M}C^{M}$cacb took 87, 178, and 156 s for I, L, and V, respectively.

Results and discussion

Despite elegant and sensitive experiments, existing protocols to assign methyl resonances can be tedious and prone to human error. To assign methyl resonances of isoleucine, leucine, and valine in large proteins, Tugarinov and Kay have designed the HMCM(CG)CBCA (Tugarinov and Kay 2003) experiment, and Sprangers and Kay the modified HMCMCGCB (Sprangers and Kay 2007). The first experiment correlates methyl groups of isoleucine, leucine, and valine with their C^{α} and C^{β} carbons. HMCMCGCB is a more sensitive experiment and provides correlations (H^{M}, C^{M}, C^{γ}) and (H^{M}, C^{M}, C^{β}) for isoleucines and leucines. For valines, HMCMCGCB correlates methyl groups with C^{α} and C^{β} carbons and can be renamed V-HMCMCBCA. Here, we focus on the correlations (H^{M}, C^{M}, C^{α}) and (H^{M}, C^{M}, C^{β}) and we refer to either or both experiments as HMCM experiments. (H^{M}, C^{M}) correlations are assigned indirectly by matching C^{α} and C^{β} signals from HMCM spectra to those observed in HNCA and HN(CA)CB, referred together as HN experiments.

Two strategies are commonly employed to match C^{α} or C^{β} signals observed in HN spectra with those observed in HMCM spectra. In the first strategy, peak lists of HN and HMCM spectra are used to compare C^{α} and C^{β} chemical shifts (Tugarinov and Kay 2003). This method has the advantage of requiring minimal human intervention once the peak lists have been created but will fail if (1) any signal in any of the four experiments has not been picked, (2) sequential signals in HN spectra have been erroneously assigned to intra-residue correlations, (3) C^{α} or C^{β} resonances of (nearly) overlapping (H^{N}, N) and (H^{M}, C^{M}) correlations have been mis-paired (e.g. the C^{β} of residue i was paired with the (H^{N}, N) of residue j), or (4) the chemical shifts determined by peak-picking are inaccurate, for instance because a signal overlap shifts signal maxima.

The second strategy relies principally on concerted visual inspection of HN and HMCM spectra (Frueh 2014). The method is illustrated in Fig. 1a, and examples are shown in Fig. 3a–f, i–n. In this strategy, H^{M}/C^{M} planes of HMCM(CG)CBCA and HMCMCGCB are displayed at the frequencies $\omega_{C\alpha}$ and $\omega_{C\beta}$ determined from HN spectra assignment. To assign the methyl signals, *all* (H^{M}, C^{M}) correlations in both planes have to be inspected simultaneously (e.g. by synchronizing a cross-hair cursor) to identify methyl correlations found both at $\omega_{C\alpha}$ and $\omega_{C\beta}$. With this method methyl resonances may be assigned even if accurate peak picking of C^{α} and C^{β} signals is not

possible in these spectra. Unfortunately, the method may be tedious and prone to human error. Indeed, HM/CM planes taken at a specific $\omega_{C\alpha}$, for example, will display (H^M, C^M) correlations with maxima near $\omega_{C\alpha}$ in addition to correlations with maxima at exactly $\omega_{C\alpha}$. These correlations are referred to as "bleed-through" correlations because their maxima occur in other planes than the plane under inspection. They occur when different residues have different yet similar C^α and/or C^β signals. Bleed-through correlations in H^M/C^M planes must be eliminated by inspecting C^α and C^β dimensions for each (H^M, C^M) correlation and substantial effort may be spent identifying them, in contrast to the rapid peak picking method. Overall, it would be preferable to use a method that combines the advantages of each procedure; the method should not rely on preliminary peak picking but nevertheless be simple to utilize.

4D correlation maps for assigning methyl resonances

A complex assignment procedure can be translated into a simple correlation map featuring *direct correlations* between amide and methyl groups. The procedure described in the previous paragraph can be expressed as two distinct steps that can each be recast mathematically. The first step is formulated as "for each (H^N, N, C^α) correlation found in HNCA, find the corresponding (H^M, C^M, C^α) correlations in HMCM" and "for each (H^N, N, C^β) in HN(CA)CB find (H^M, C^M, C^β) in HMCM". The second step is defined as "identify (H^N, N) and (H^M, C^M) correlations that have *both* C^α and C^β signals in common". The mathematical equivalence of the first step is "for each nitrogen frequency in HNCA and for each methyl carbon frequency in HMCM, calculate the covariance matrix between H^N/C^α and H^M/C^α planes". Using the formalism of Brüschweiler and collaborators (Brüschweiler 2004; Short et al. 2011; Zhang and Bruschweiler 2004), we calculate the 4D array

$$4D\,H^N NH^M C^M ca(a,b,c,e) = \frac{1}{D-1}\sum_{d=1}^{D} HN\widetilde{CA}(a,b,d) \cdot HMCM\widetilde{CA}(c,e,d), \quad (1)$$

in which the indices a and b span the amide proton and nitrogen dimensions, c and e span the methyl proton and carbon dimensions, and d spans the common C^α dimension. "$\sim$" Indicates that the means along the C^α dimensions have been subtracted (Trbovic et al. 2004). D is the number of points in the common C^α dimension. In Eq. 1, HMCMCA represents the HMCM experiment that provides optimal C^α signals; for valine, HMCMCA corresponds to HMCMCGCB (V-HMCMCBCA) which provides (H^M, C^M, C^α) and (H^M, C^M, C^β) for this residue; for leucine and isoleucine HMCMCA corresponds to HMCM(CG)CBCA. For each (H^N, N) coordinate of the 4D $H^N NH^M C^M$ca, a 2D-H^M/C^M correlation map reveals the methyl resonances that possess the same C^α chemical shift as that of the amide group (Fig. 1b, left). Similarly, a 4D $H^N NH^M C^M$cb can be calculated using C^β signals from HN(CA)CB and HMCMCGCB (Fig. 1b, right). The second step in the assignment procedure, "find (H^N, N) and (H^M, C^M) correlations that share *both* C^α and C^β signals", can be translated mathematically as an element-wise product between 4D $H^N NH^M C^M$ca and 4D $H^N NH^M C^M$cb:

$$4D\,H^N NH^M C^M cacb = 4DH^N NH^M C^M ca \odot 4DH^N NH^M C^M cb. \quad (2)$$

In 4D $H^N NH^M C^M$cbca, (H^N, N, H^M, C^M) correlations only occur for amide and methyl groups that have *both* C^α and C^β signals in common. Thus, 4D $H^N NH^M C^M$cbca summarizes the entire assignment procedure with correlations that can be identified by visual inspection.

Residue-specific 4D maps and artifact removal

Two orthogonal preliminary treatments result in residue-specific 4D $H^N NH^M C^M$cbca correlation maps with minimal artifacts. First, because C^α and C^β signals of isoleucine, leucine, and valine residues are featured in well-defined spectral regions, these regions can be extracted from HN and HMCM spectra before performing covariance, according to each residue and according to each nucleus (Fig. 2). Thus, three 4D correlation maps, V-$H^N NH^M C^M$cbca, I-$H^N NH^M C^M$cbca, and L-$H^N NH^M C^M$cbca, can be calculated with tailored spectral regions. This step abrogates false correlations stemming from resonances outside the extracted regions, such as sequential C^α or C^β correlations in HNCA and HN(CA)CB or those stemming from C^γ and C^δ signals in HMCM experiments. The bounds of each region can be defined using statistical distributions of chemical shifts (e.g. from the Biological Magnetic Resonance Bank), or they can be tailored to the protein under study so as to only encompass assigned C^α and C^β signals. Likewise, H^M, and C^M dimensions can be reduced to exclusively encompass signals of interest for a given residue type. (H^M, C^M) correlations of isoleucine cluster in a distinct spectral region. Valine and leucine (H^M, C^M) correlations have different phases and spectral regions can be defined to only include the targeted signals. Finally, because (H^N, N) correlations are assigned, the amide regions can be defined to exclude irrelevant correlations. As an important computational benefit, memory load as well as disk space are reduced and a four-dimensional array can be both generated and used in a practical manner.

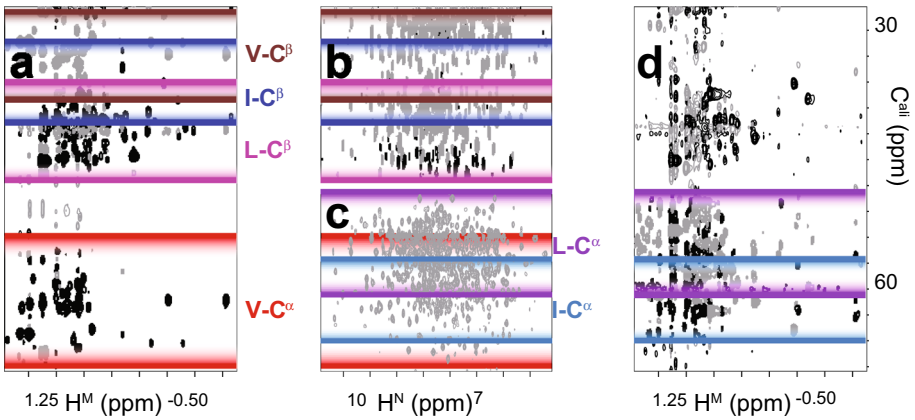

Fig. 2 Residue-specific spectral regions used to calculate residue-optimized correlation maps. H/C$^{\alpha/\beta}$ projections of *a* HMCMCGCB, *b* HN(CA)CB, *c* HNCA, and *d* HMCM(CG)CBCA. For each amino acid, a spectral region that is specific to its C$^{\beta}$ chemical shift and common to HMCMCGCB and HN(CA)CB is extracted. Signals in *grey* and *black* have opposite phases. *Brown*: V-C$^{\beta}$ (*grey* signals) *dark blue*: I-C$^{\beta}$ (*black*) and *magenta*: L-C$^{\beta}$ (*black*). Valine alpha carbons, V-C$^{\alpha}$, appear both in HMCMCGCB and HNCA and their spectral region is coloured in *red* (*black* signals in *a*). The remaining two regions, I-C$^{\alpha}$ and L-C$^{\alpha}$, are found both in HNCA and HMCM(CG)CBCA and are coloured in *light blue* and *purple*, respectively (*grey* signals in *d*)

The maps are further improved by performing derivation in all C$^{\alpha}$ and C$^{\beta}$ dimensions subject to covariance before calculating the four dimensional array. This procedure ensures that covariance reports on C$^{\alpha}$ or C$^{\beta}$ signals that have matching maxima in amide and methyl experiments rather than signals that may partially overlap (Harden et al. 2014). In H^{N}NH^{M}C^{M}cbca, correlations from methyls with C$^{\alpha}$ and C$^{\beta}$ signals similar but non-identical to those of amide groups are greatly reduced, vanish, or become negative (and hence are identified as artifacts, see supporting information Figure S1). In effect, this treatment mitigates the aforementioned bleed-through correlations that occur in original spectra. That is, bleed-through correlations seen in an H^{M}/C^{M} plane of HMCM(CG)CBCA, for example, may disappear in the corresponding H^{M}/C^{M} plane in H^{N}NH^{M}C^{M}ca. In the end, reliable, residue-specific 4D spectra can be used to pair methyl resonances with assigned amide groups.

Application to a 52 kDa protein

Covariance correlation maps provide rapid and reliable assignment of methyl signals. Figure 3 compares the traditional procedure that employs simultaneous visual inspection of HN and HMCM spectra with the novel procedure using 4D H^{N}NH^{M}C^{M}cbca. In the traditional procedure, H/C$^{\alpha}$ and H/C$^{\beta}$ strips from HN spectra (f, e) provide C$^{\alpha}$ and C$^{\beta}$ frequencies of an assigned (H^{N}, N) correlation, here V315 (g). H^{M}/C^{M} planes of HMCMCGCB (c, d) are then displayed for both frequencies. Each plane contains many correlations, 12 of which are seen in both planes. *For each of these 12 correlations*, we must display H^{M}/C$^{\alpha}$ (a) and H^{M}/C$^{\beta}$ (b) strips and verify that C$^{\alpha}$ and C$^{\beta}$ signals have the same maxima as found in H/C$^{\alpha}$ and H/C$^{\beta}$ strips from HN spectra (e, f). Eventually, after close inspections of all 12 correlations, we determine the coordinates of (H$^{\gamma 1}$, C$^{\gamma 1}$) and (H$^{\gamma 2}$, C$^{\gamma 2}$) (Fig. 3a, b, respectively). In contrast, assignment with V-H^{N}NH^{M}C^{M}cbca is strikingly trivial; an H^{M}/C^{M} plane from the 4D, at the (H^{N}, N) indirect coordinates of V315, immediately reveals the signals of (H$^{\gamma 1}$, C$^{\gamma 1}$) and (H$^{\gamma 2}$,C$^{\gamma 2}$) (Fig. 3h). Note that throughout the text the labels γ_1 and γ_2 are arbitrarily used to label each methyl group; stereochemical assignment cannot be performed with our method. The methyl groups of fifteen out of sixteen valines with assigned (H, N) correlations were rapidly assigned with V-H^{N}NH^{M}C^{M}cbca. The remaining valine, V338, has no signals in HN(CA)CB and could not provide correlations in V-H^{N}NH^{M}C^{M}cbca. However, V338 has a unique C$^{\alpha}$ frequency so its methyls could be assigned with V-H^{N}NH^{M}C^{M}ca.

Another advantage of 4D H^{N}NH^{M}C^{M}cbca arrays is illustrated when assigning the methyl of I428 (Fig. 3i–o). Both traditional and novel procedures are the same as described for V315, with C$^{\alpha}$ correlations provided by HMCM(CG)CBCA (Fig. 3k, i) instead of HMCMCGCB, as discussed above. However, the C$^{\alpha}$ signal of I428 is extremely weak in HNCA and was not picked (Fig. 3n). Hence, the methyl resonance of I428 could not be assigned before generating I-H^{N}NH^{M}C^{M}cbca. When calculating the correlation map, strong C$^{\beta}$ and C$^{\alpha}$ signals in HN(CA)CB and HMCM experiments rescue the sensitivity of HNCA for this residue, and the (H^{M}, C^{M}) correlation is easily assigned in I-H^{N}NH^{M}C^{M}cbca (Fig. 3o). The C$^{\alpha}$ frequency was found by inspecting HMCM(CG)CBCA (Fig. 3i) at the (H^{M}, C^{M}) coordinates identified with I-H^{N}NH^{M}C^{M}cbca. Using I-H^{N}-NH^{M}C^{M}cbca, we assigned the methyl resonances of eight of

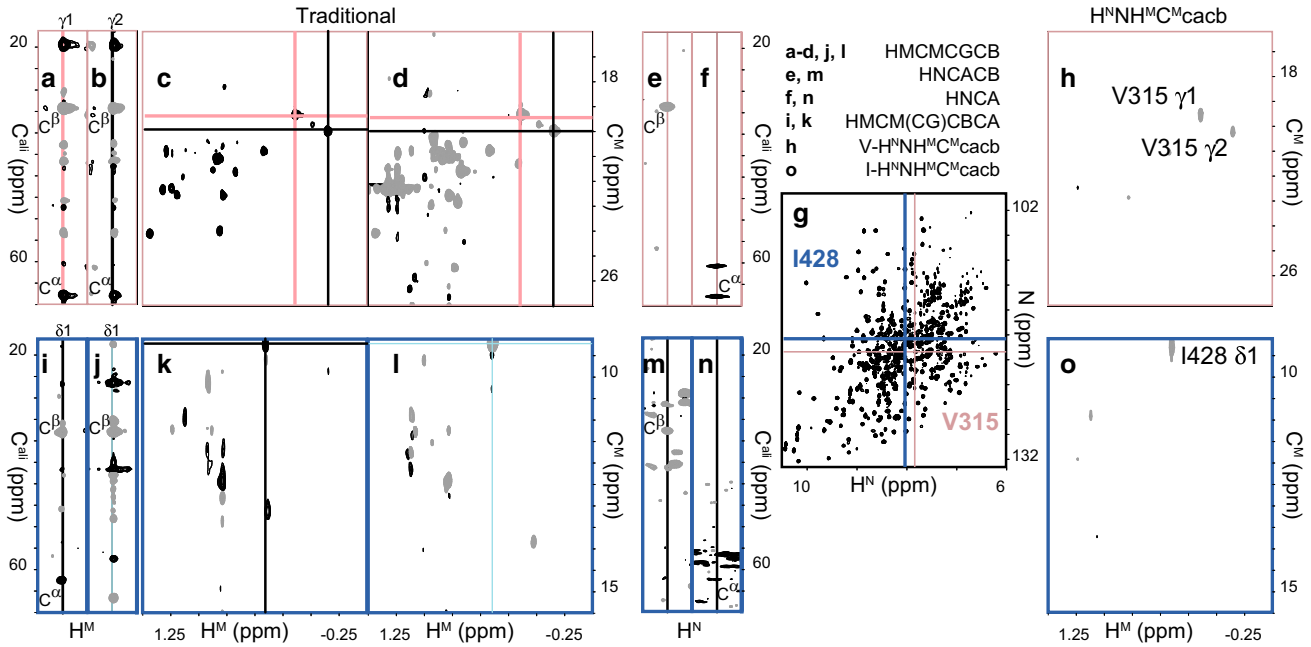

Fig. 3 Comparison of assignment procedures with original spectra (*a–f*, *i–n*) and with covariance correlation maps (*h*, *o*) for V315 (*a*, *h*) and I428 (*i*, *o*). C^{α} signals appear in *black* and C^{β} signals in *grey*. Cross-hairs in *c*, *d*, *k*, *l* indicate the position of the strips shown in *a*, *b*, *i*, and *j*, respectively. *a*, *b* $H^M/C^{\alpha/\beta}$ strips of HMCMCGCB perpendicular to H^M/C^M planes shown in *c*, *d*. *c*, *d*. H^M/C^M planes at the frequencies of V315 C^{α} and C^{β}, respectively. $H^N/C^{\alpha/\beta}$ strips of HNCA (*f*, *n*) and HN(CA)CB (*e*, *m*) are shown for the positions indicated by the cross-hairs on the HN-TROSY shown in *g*. $H^M/C^{\alpha/\beta}$ strips for all signals seen in *c* and *d* must be inspected for assigning V315 methyl resonances. *h* With V-$H^NNH^MC^M$cacb, the methyl resonances of V315 are revealed by visualisation of an H^M/C^M plane at the (H, N) coordinates defined in *g*. The labels γ_1 and γ_2 are arbitrarily used to label each methyl group and do not represent stereochemical assignments. *i* $H^M/C^{\alpha/\beta}$ strip of HMCM(CG)CBCA and *j* HMCMCGCB perpendicular to the planes shown in *k* and *l*, respectively. $H^M/C^{\alpha/\beta}$ strips for all signals seen in *k* and *l* must be inspected for assigning I428 methyl resonances. *o* The methyl resonance of I428 is seen in the H^M/C^M plane of I-$H^NNH^MC^M$cacb, at the (H, N) coordinates defined in *g*

ten Ile residues that had assigned amide groups. One residue had no C^{α} signal in HNCA while another showed no correlations in HMCM(CG)CBCA. Clearly, as for conventional methods, the methyl resonances cannot be paired with assigned amide groups when one or more spectra lack signals. However, as shown for I428, a very weak signal that may escape peak picking or visual inspection can lead to an unambiguous correlation in 4D $H^NNH^MC^M$cbca if other spectra feature strong signals for that residue. In the end, covariance correlation maps permitted the assignment of all valine and isoleucine methyls for residues that had no missing signal in any spectra.

Covariance correlation maps help alleviate limitations of the HMCM experiments for leucine residues. Although the short HMCMCGCB provides quality data for leucines, the longer experiment is problematic. In HMCM(CG)CBCA, leucines feature a large number of undesired multiple-quantum (MQ) correlations, in part due to branching at C^{γ} in uniformly ^{13}C-enriched samples. Thus, 16 undesired signals are seen with such samples while the desired correlation, (H^M, C^M, C^{α}), is often vanishingly small (Tugarinov and Kay 2003). In addition, we have found that many leucines had no signal in HMCM(CG)CBCA for either one or both methyls, e.g. because unusual $^1J_{CC}$ values deteriorate COSY transfers and further decrease the signal-to-noise. These shortfalls are expected to be reduced when using samples in which only one methyl is enriched in ^{13}C (Tugarinov and Kay 2003). However, such samples are expensive, and in our experiments both methyls have ^{13}C ^{1}H labeling. In summary, for leucines, information on C^{α} is often absent and HMCMCGCB must be analysed in concert with HN(CA)CB, HC(C)H-TOCSY (Kay et al. 1993), NOESY-HN-HSQC and NOESY-HC-HSQC spectra. Here, an L-$H^NNH^MC^M$cb map offers some relief by reducing the number of correlations that must be inspected. Indeed, the calculation of L-$H^NNH^MC^M$cb includes taking the derivatives in C^{β} dimensions before calculating a covariance matrix. As mentioned above, this procedure greatly reduces the number of bleed-through resonances when comparing an H^M/C^M plane in L-$H^NNH^MC^M$cb with that of HMCMCGCB (Fig. 4). Figure 4 highlights the advantages of L-$H^NNH^MC^M$cb with a challenging example in which two residues, L304 and L277, have degenerate C^{β} shifts and many others have similar C^{β} shifts. Clearly neither the traditional approach nor L-$H^NNH^MC^M$cb can unambiguously assign methyls specifically to L304 or L207, and spectra from HC(C)H-TOCSY, NOESY-HN-HSQC and NOESY-HC-HSQC must

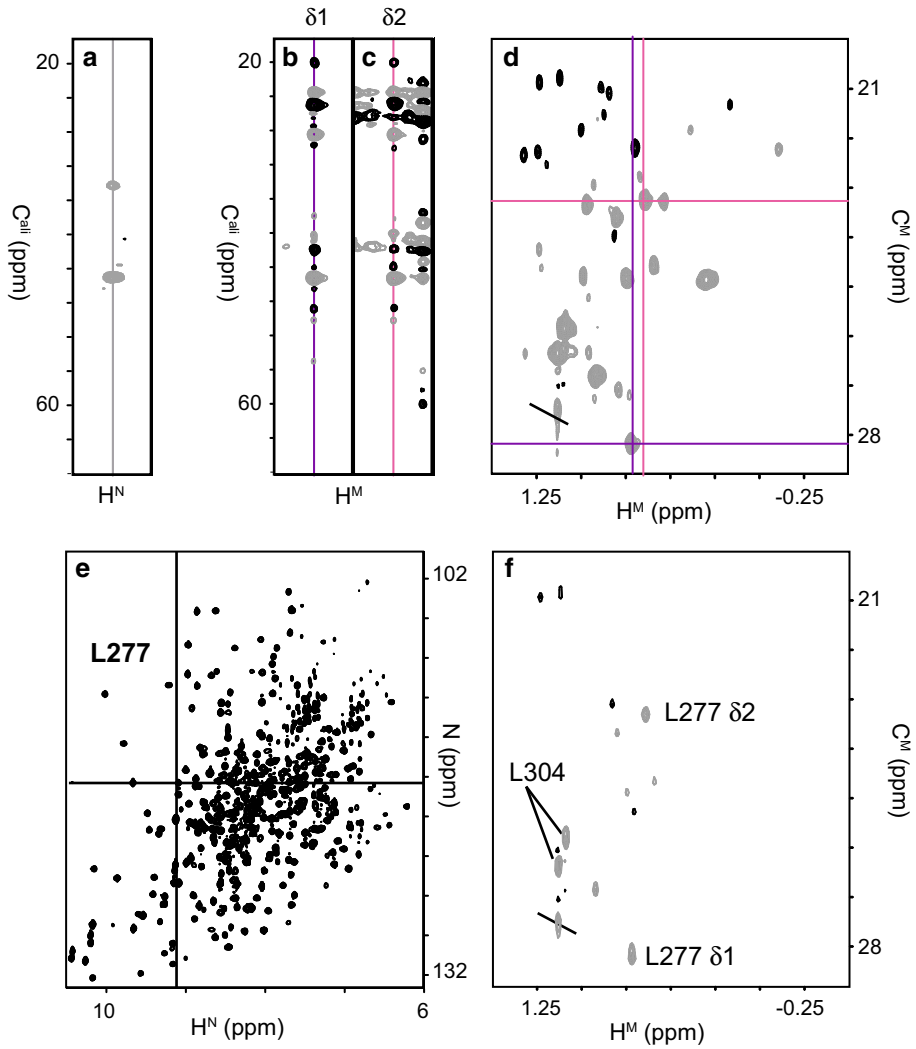

Fig. 4 Assignment of leucine methyls without using data reporting on C^{α}. *a–d* Conventional approach; *f* $H^{N}NH^{M}C^{M}$cb. *a* H^{N}/C^{β} strip of HN(CA)CB used in the conventional approach to identify C^{β} frequencies. *b*, *c* $H^{M}/C^{\alpha/\beta}$ strips of HMCMCGCB to be compared with *a*. Shown are strips later assigned to L277 and corresponding to the cross-hairs shown in *d*. *d* H^{M}/C^{M} plane of HMCMCGCB. Strips such as shown in *b* and *c* as well as NOESY strips must be displayed for all correlations to finalize the assignment. Note the profusion of correlations stronger than those later assigned to L277. *e* HN-TROSY of 52 kDa Cy1. The cross-hair denotes the signal of L277 and provides the coordinates for the strip displayed in *a* as well as the H^{M}/C^{M} plane shown in *f*. *f* H^{M}/C^{M} plane of $H^{N}NH^{M}C^{M}$cb. Only four correlations must be investigated with NOESY data to finalize the assignment. The labels δ_1 and δ_2 are arbitrarily used to label each methyl group and do not represent stereochemical assignments

also be inspected. However, 18 (H^{M}, C^{M}) correlations must be inspected in the HMCMCGCB H^{M}/C^{M} plane (Fig. 4d) (Fig. 4b, c shows the strips after assignment). In contrast, only five correlations must be inspected in L-$H^{N}NH^{M}C^{M}$cb and the methyl resonances of L304 and L277 are rapidly and unambiguously identified (Fig. 4f). With L-$H^{N}NH^{M}C^{M}$cb, HC(C)H-TOCSY, and NOESY spectra, we assigned 23 pairs of leucine methyls. In addition, 11 leucine methyl pairs could be assigned with $H^{N}NH^{M}C^{M}$cbca as described for valines and isoleucines. For three more residues, $H^{N}NH^{M}C^{M}$cbca featured a single methyl correlation and the second methyl was assigned with HC(C)H-TOCSY. Four more residues had characteristic C^{β} signals and they were assigned with L-$H^{N}NH^{M}C^{M}$cb alone, without resorting to TOCSY and NOESY spectra. In all, 91 % of the leucines that could potentially be assigned with the original data were rapidly assigned with help from covariance correlation maps.

Prior to the development of 4D-$H^{N}NH^{M}C^{M}$ covariance maps, substantial time and effort was invested to assign 8 isoleucine, 10 valine, and only 14 leucine methyl resonances (e.g. Figures 3a–f, i–n, 4a–d). Here, less than 6 h were used to assign 8 isoleucines, 16 valines, and 41 leucines with the assistance of covariance correlation maps (see supporting information, Table S1). Nevertheless, prior to a thorough analysis of the covariance maps, we do recommend rapidly inspecting the original spectra to identify missing signals and avoid unnecessary efforts. For example if C^{α} is not seen in HNCA, one should use $H^{N}NH^{M}C^{M}$cb rather than $H^{N}NH^{M}C^{M}$cacb, as described above for leucines. Overall, disregarding residues that have missing signals, the methyl resonances of 100 % of Ile, 100 % of Val, and 91 % of Leu could be paired with the corresponding assigned amide resonances with covariance maps.

Conclusion

We have designed residue-specific, reliable 4D correlation maps that correlate methyl resonances with assigned amide resonances and alleviate heretofore time-consuming

assignment procedures. Simple mathematical operations replace user intervention that is otherwise needed prior to assigning resonances. Thus, comparison of C^{α} (or C^{β}) frequencies in methyl and amide spectra, either visually or after peak picking, is replaced by covariance analysis. Likewise, combining the information provided separately by C^{α} and C^{β} nuclei is simply conveyed by the element-wise product of two 4D maps. The method exploits spectroscopic properties of residues both to provide reliable maps and to permit facile analysis with sparse, residue-specific 4D arrays. Thus, nearly all assigned amide resonances of a 52 kDa protein were paired with methyl resonances in about half a day. Limitations reflect the quality of the original spectra, which in turn is governed by labeling schemes. We expect that $H^{N}NH^{M}C^{M}$cbca correlation maps can be further improved when methyl spectra are obtained with specific ^{13}C labeling of a single methyl. The method can easily be adapted to assign methyl resonances of other residues. T-HMCMCBCA (Guo and Tugarinov 2010) and γ2-I-HMCMCBCA (Sheppard et al. 2009) can be used with HNCA and HN(CA)CB as described above for valines. For alanine methyls, HMCMCA (Sheppard et al. 2009) can be combined with HNCA to produce an A-$H^{M}C^{M}H^{N}$Nca map. For proteins up to 40 kDa, the methyl groups of all residues can be correlated to C^{α} and C^{β} nuclei with a TOCSY experiment and a single fractionally deuterated sample (Otten et al. 2010). Again, residue-specific $H^{M}C^{M}H^{N}$Ncacb can be generated by adapting our protocol. Similarly, although we have calculated correlations mediated by C^{α} and C^{β} nuclei, the method can readily include correlations conveyed by carbonyl carbons and provided by the experiments HMCM(CGCBCA)CO (Tugarinov and Kay 2003) and HN(CA)CO. Thus, our protocol can be used to assign the methyl groups of any residue (save methionines), and it ensures that all experimental data available is exploited for optimal results. The method will permit facilitated assignment of NMR methyl resonances, which are critical probes in structural and functional studies of large proteins by NMR.

Acknowledgments We thank Dr Lewis Kay for providing the pulse sequences of HMCMCGCB and HMCM(CG)CBCA, and Brad Harden, Scott Nichols and Andy Goodrich for carefully reading the manuscript. This research was supported by the National Institutes of Health (Grant GM 104257).

References

Blinov KA, Larin NI, Kvasha MP, Moser A, Williams AJ, Martin GE (2005) Analysis and elimination of artifacts in indirect covariance NMR spectra via unsymmetrical processing. Magn Reson Chem 43:999–1007. doi:10.1002/mrc.1674

Bloudoff K, Rodionov D, Schmeing TM (2013) Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. J Mol Biol 425:3137–3150. doi:10.1016/j.jmb.2013.06.003

Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. Structure (Camb) 10:301–310

Brüschweiler R (2004) Theory of covariance nuclear magnetic resonance spectroscopy. J Chem Phys 121:409–414

Brüschweiler R, Zhang F (2004) Covariance nuclear magnetic resonance spectroscopy. J Chem Phys 120:5253–5260

Chen K, Delaglio F, Tjandra N (2010) A practical implementation of cross-spectrum in protein backbone resonance assignment. J Magn Reson 203:208–212. doi:10.1016/j.jmr.2009.12.018

Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106:3468–3496

Frueh DP (2014) Practical aspects of NMR signal assignment in larger and challenging proteins. Prog Nucl Magn Reson Spectrosc 78:47–75. doi:10.1016/j.pnmrs.2013.12.001

Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. Nature 454:903–906. doi:10.1038/nature07162

Gardner KH, Kay LE (1997) Production and incorporation of 15N, 13C, 2H (1H-1 methyl) isoleucine into proteins for multidimensional NMR studies. J Am Chem Soc 119:7599–7600

Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131:756–769. doi:10.1016/j.cell.2007.09.039

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. J Biomol NMR 13:369–374

Gouridis G, Karamanou S, Gelis I, Kalodimos CG, Economou A (2009) Signal peptides are allosteric activators of the protein translocase. Nature 462:363–367

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

Gulick AM (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. ACS Chem Biol 4:811–827. doi:10.1021/cb900156h

Guo C, Tugarinov V (2010) Selective 1H-13C NMR spectroscopy of methyl groups in residually protonated samples of large proteins. J Biomol NMR 46:127–133. doi:10.1007/s10858-009-9393-0

Harden BJ, Nichols SR, Frueh DP (2014) Facilitated assignment of large protein NMR signals with covariance sequential spectra using spectral derivatives. J Am Chem Soc 136:13106–13109. doi:10.1021/ja5058407

Kay LE, Xu GY, Singer AU, Muhandiram DR, Formankay JD (1993) A Gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. J Magn Reson Ser B 101:333–337. doi:10.1006/jmrb.1993.1053

Keating TA, Miller DA, Walsh CT (2000) Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of yersiniabactin synthetase. Biochemistry 39:4729–4739

Koglin A et al (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science 312:273–276

Kupce E, Freeman R (2006) Hyperdimensional NMR spectroscopy. J Am Chem Soc 128:6020–6021. doi:10.1021/ja0609598

Lescop E, Brutscher B (2007) Hyperdimensional protein NMR spectroscopy in peptide-sequence space. J Am Chem Soc 129:11916–11917. doi:10.1021/ja0751577

Matlab reference guide (1992). Natick, MA

Mishra SH, Harden BJ, Frueh DP (2014) A 3D time-shared NOESY experiment designed to provide optimal resolution for accurate assignment of NMR distance restraints in large proteins. J Biomol NMR 60:265–274. doi:10.1007/s10858-014-9873-8

Mootz HD, Schwarzer D, Marahiel MA (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chembiochem 3:490–504. doi:10.1002/1439-7633(20020603)3:6<490:AID-CBIC490>3.0.CO;2-N

Otten R, Chu B, Krewulak KD, Vogel HJ, Mulder FA (2010) Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. J Am Chem Soc 132:2952–2960. doi:10.1021/ja907706a

Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE (2010) The proteasome antechamber maintains substrates in an unfolded state. Nature 467:868–871. doi:10.1038/nature09444

Saio T, Guan X, Rossi P, Economou A, Kalodimos CG (2014) Structural basis for protein antiaggregation activity of the trigger factor chaperone. Science 344:1250494. doi:10.1126/science.1250494

Sheppard D, Guo C, Tugarinov V (2009) Methyl-detected 'out-and-back' NMR experiments for simultaneous assignments of Alaβ and Ileγ2 methyl groups in large proteins. J Biomol NMR 43:229–238. doi:10.1007/s10858-009-9305-3

Short T, Alzapiedi L, Brüschweiler R, Snyder D (2011) A covariance NMR toolbox for MATLAB and OCTAVE. J Magn Reson 209:75–78. doi:10.1016/j.jmr.2010.11.018

Snyder DA, Brüschweiler R (2009) Generalized indirect covariance NMR formalism for establishment of multidimensional spin correlations. J Phys Chem A 113:12898–12903. doi:10.1021/jp9070168

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622. doi:10.1038/nature05512

Tanovic A, Samel SA, Essen LO, Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. Science 321:659–663

Trbovic N, Smirnov S, Zhang F, Bruschweiler R (2004) Covariance NMR spectroscopy by singular value decomposition. J Magn Reson 171:277–283

Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc 125:13868–13878

Tugarinov V, Kay LE (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. Chembiochem 6:1567–1577. doi:10.1002/cbic.200500110

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125:10420–10428

Walsh CT (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science 303:1805–1810

Zhang F, Bruschweiler R (2004) Indirect covariance NMR spectroscopy. J Am Chem Soc 126:13180–13181