

Nearest-neighbor NMR spectroscopy: categorizing spectral peaks by their adjacent nuclei

Soumya P. Behera^{1,5}, Abhinav Dubey^{1,2,5}, Wan-Na Chen¹, Viviane S. De Paula³, Meng Zhang¹, Nikolaos G. Sgourakis³, Wolfgang Bermel⁴, Gerhard Wagner¹, Paul W. Coote^{1,2}✉ & Haribabu Arthanari^{1,2}✉

Methyl-NMR enables atomic-resolution studies of structure and dynamics of large proteins in solution. However, resonance assignment remains challenging. The problem is to combine existing structural information with sparse distance restraints and search for the most compatible assignment among the permutations. Prior classification of peaks as either from isoleucine, leucine, or valine reduces the search space by many orders of magnitude. However, this is hindered by overlapped leucine and valine frequencies. In contrast, the nearest-neighbor nuclei, coupled to the methyl carbons, resonate in distinct frequency bands. Here, we develop a framework to imprint additional information about passively coupled resonances onto the observed peaks. This depends on simultaneously orchestrating closely spaced bands of resonances along different magnetization trajectories, using principles from control theory. For methyl-NMR, the method is implemented as a modification to the standard fingerprint spectrum (the 2D-HMQC). The amino acid type is immediately apparent in the fingerprint spectrum. There is no additional relaxation loss or an increase in experimental time. The method is validated on biologically relevant proteins. The idea of generating new spectral information using passive, adjacent resonances is applicable to other contexts in NMR spectroscopy.

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. ²Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. ³Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA. ⁴Magnetic Resonance Spectroscopy NMR Application, Bruker BioSpin GmbH, 76287 Rheinstetten, Germany. ⁵These authors contributed equally: Soumya P. Behera, Abhinav Dubey. ✉email: paul_coote@hms.harvard.edu; hari@hms.harvard.edu

NMR spectroscopy enables study of the structure and dynamics of biomolecules in near-native conditions^{1–4}. In particular, NMR can probe dynamic events like protein folding, protein–protein, and protein–ligand interactions, and conformational exchange, and characterize lowly populated states^{5–8}. However, deriving backbone resonance assignment using traditional triple resonance NMR experiments are restricted to lower molecular weight proteins (e.g., <~150 kDa). This size limit is due to the rapid relaxation of the magnetization in multidimensional experiments, which is in turn related to slow molecular tumbling of larger molecules.

The advent of ¹H–¹³C methyl-TROSY experiments and isotope labeling has extended the range of solution NMR to proteins up to 1 MDa in size^{9–11}. The CH₃ protons undergo rapid motion about the methyl axis compared with the rotational tumbling time of proteins. The three chemically equivalent methyl protons give a single peak, which has relatively better sensitivity than backbone resonances¹². The heteronuclear multiple quantum coherence (HMQC) experiment takes advantage of the cross-correlated relaxation among the methyl protons, known as the methyl-TROSY effect, to transfer the magnetization through slow-relaxing pathways, resulting in narrow linewidths and high intensity peaks⁴. In addition, selective methyl protonation in an otherwise deuterated protein minimizes ¹H–¹H dipole-induced relaxation^{13–15}. These two advances together make large protein complexes amenable for NMR studies¹⁶. The amino acids with methyl resonances (Ile, Val, Leu, Ala) typically make up a substantial and well-distributed portion of the protein structure, so these resonances are excellent probes to study dynamics¹⁷.

Meaningful interpretation of the methyl-NMR data requires that NMR signals for the HMQC experiment have been assigned to sequence specific methyl bearing residue in the protein. For large proteins (>150 kDa), this has been accomplished by mutating individual residues and tracking changes in the spectrum¹⁸; however, this process is expensive and time consuming. Moreover, specific mutations may interfere with the ability of the protein to fold properly and/or affect expression yields. Recently, developed computational techniques use inter-methyl distance restraints derived from nuclear overhauser effect (NOE) experiments, compared with distances extracted from the protein structure as a means to obtain resonance assignments for methyl residues^{19–26}. The central idea is to systematically compare and match candidate assignments for consistency between the known set of inter-methyl distances and NOE distance estimates and choose the best overall match. The search space is typically astronomical; *n* resonances can be matched to *n* methyl groups in *n!* different ways.

Knowledge of the amino-acid type significantly reduces the size of the search space for the computational strategy described above. For instance, compare the matching problem for 30 arbitrary/unclassified amino acids versus jointly matching 15 valines and 15 leucines. The number of combinations, 15! × 15!, is over 100 million times smaller than 30!—a vast difference that can be traded for reduced computational time and complexity, and/or improved reliability of the eventual solution though better coverage of the search space. Fortunately, alanine and isoleucine resonances are easily distinguished by their characteristic spectral frequencies. However, the resonances corresponding to valine and leucine occupy the same spectral space. One way to distinguish leucine from valine peaks in spectrum is by making two different samples using specialized precursors for selective labeling of leucine^{27–29}, which is costly and labor intensive, especially for membrane proteins. In addition, unlabeled leucine will add to the proton density, which will affect the relaxation rates of the valines and one has to resort to an expensive deuterated leucine. Similarly, one can also distinguish leucine from valine peaks by

preparing an additional sample where valine methyl is labeled with NMR inactive ¹²C rendering those peaks absent from the spectrum. This approach is slightly cheaper than one described before but still it is labor intensive. Another way is to use three-dimensional (or higher) NMR experiments. However, these add extra delays for coherence transfer and encoding of additional dimensions^{30–32}. Extra delays result in severe relaxation losses, which can make these approaches unsuitable for large or challenging proteins.

Here, we show that the distinction between methyl resonances corresponding to leucine and valines can be obtained without the drawbacks discussed above using specially designed pulse for selective homonuclear decoupling. We demonstrate the applicability of this selective pulse to distinguish leucine and valines on a variety of proteins.

Results

Design of homonuclear decoupling pulse, which selectively affects valine resonances. In this article, we demonstrate that the leucine and valine signals can be clearly distinguished using a specially designed selective homonuclear decoupling pulse during the indirect ¹³C chemical shift encoding delay. The pulse refocuses the ¹³C_{Val}–¹³C_{Val}^β *J*-coupling but does not affect the encoding of the methyl chemical shift frequencies (Fig. 1a). The leucine peaks are not decoupled. As the selective pulse is applied during an existing delay in the experiment, there is no additional

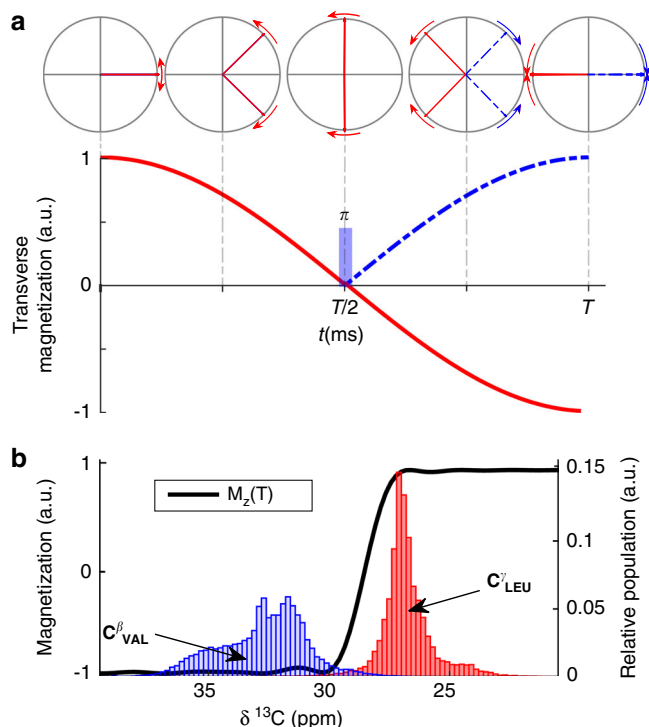


Fig. 1 Theory of decoupling using selective inversion to distinguish leucine from valine. **a** Trajectory of the ¹³CH₃ leucine (red) and valine (blue, dashed) magnetization during a constant time encoding period of duration $T = 1/J_{CC} = 28$ ms. Selective decoupling of valine at time $T/2$ refocuses the coupling and produces positive peaks for valine at time T . Leucine is not decoupled, and the coupling interaction leads to negative peaks at time T . **b** The distributions of C^{β} and C^{γ} chemical shifts of valine (blue) and leucine (red) are distinct. Assuming an initial state of I_z , our decoupling pulse selectively inverts the C^{β}_{Val} , which selectively decouples the valine methyl resonance. The encoding of the methyl-carbon chemical shift continues unaffected during the selective homonuclear decoupling pulse.

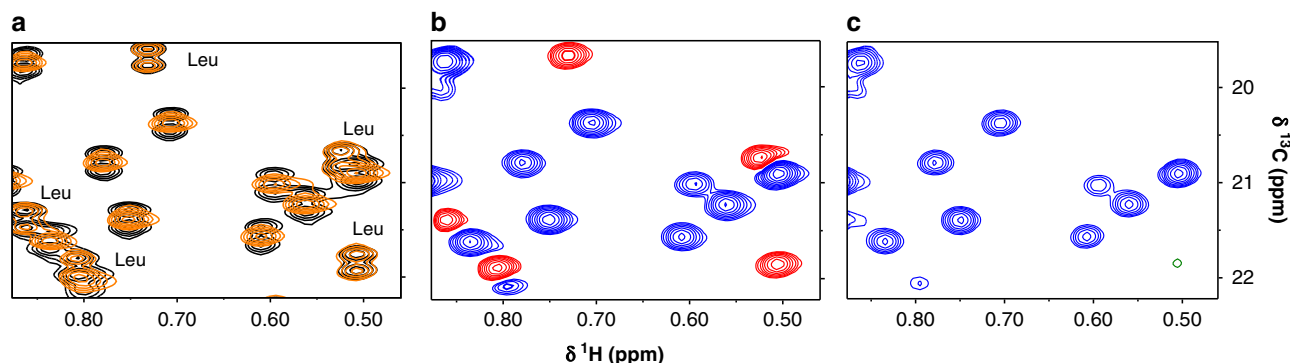


Fig. 2 Testing the selective decoupling pulse using real time and constant-time ^1H - ^{13}C HMQC on MBP. **a** Overlay of a section of a real-time 2D ^1H - ^{13}C methyl HMQC experiment with (orange) and without (black) the selective decoupling pulse on a 500 μM MBP (42.5 kDa) protein. The leucine peaks remain as doublets, whereas the valine peaks collapse into singlets. **b** Constant-time ^1H - ^{13}C HMQC experiment with indirect constant-time period set to $T = 28$ ms ($1/J_{\text{CC}}$) and shaped pulse applied at time $T/2$. Valine resonances are positive (blue) whereas leucine peaks are negative (red). **c** A version of the experiment with constant-time evolution period is set to $T = 42$ ms ($3/(2 J_{\text{CC}})$) completely eliminates the leucine peaks from the spectrum with positive valine peaks remaining (blue). Full spectra are in Supplementary Fig. 5.

loss of transverse relaxation introduced by the pulse compared with a traditional HMQC experiment on the similar isotopically labeled sample. This implies that the molecular weight limit for this approach is the same as for methyl-NMR in general. The idea presented here can be used in both constant-time and real-time indirect encoding schemes, and in many kinds of methyl-NMR experiments including HSQC and HMQC.

Chemical shift data from the BMRB database³³ show that while distribution of the leucine and valine methyl chemical shifts have considerable overlap, their adjacent nearest neighbor carbon atom (C_{Leu}^{β} and C_{Val}^{β} , respectively) have well-separated chemical shifts (Fig. 1b and Supplementary Fig. 1). One way to distinguish methyls of leucine and valine is by transferring magnetization to and from the nearest neighbor. However, that requires additional coherence transfer delays and associated relaxation losses. Instead, selectively inverting C_{Val}^{β} during the methyl chemical shift encoding leads to a clear distinction between valine and leucine methyl resonances, via selectively refocusing or evolving their one-bond homonuclear J -coupling. This difference can be observed as a doublet collapsing into a singlet, in real-time encoding, or a complete sign reversal of only the leucine resonances, in constant time chemical shift encoding (Fig. 1).

Although previous studies have effectively distinguished methyl peaks of threonine and alanine by selective decoupling^{34,35}, the small frequency separation between C_{Leu}^{β} , C_{Val}^{β} , and the encoded methyl region made the approach ineffective in this case; available decoupling pulses were not selective enough to reliably distinguish leucine and valine³⁰ and also produced severe spectral artefacts³⁶. Modern pulse-design methods, such as optimal control theory, allow precise engineering of pulse selectivity, and dependable Val–Leu distinction is now achievable.

The pulse shape was designed using recently developed homonuclear decoupling techniques from optimal control theory³⁷, which permit selective decoupling of closely spaced bands of frequencies without causing any undesirable off-resonance effects, such as Bloch–Siegert shifts³⁸. The pulse is applied in the middle of the ^{13}C indirect encoding period. During the pulse, the transverse component of the methyl magnetization (for all Ile, Leu, Val, Ala residues) evolve under their intrinsic chemical shift frequencies (Supplementary Fig. 2) for an apparent delay of $T' = aT$, where T is the duration of the pulse and a ($0 < a \leq 1$) is a constant chosen during pulse design. The encoding delays on either side of the relatively short pulse can easily absorb an extra $(1-a)T$ of chemical shift encoding, so that the overall

increment is exactly compatible with the Nyquist grid, Fourier transform, and all further downstream data processing. In this case, we determined by numerical exploration that $a = 0.84$ and $T = 2.3$ ms were suitable to design a valine-selective decoupling pulse.

To numerically optimize the pulse, an ensemble of 50 spins with uniformly spaced chemical shifts ω in range of [29.2, 44] ppm was sampled to represent the C_{Val}^{β} . An initial state of $\rho(0) = I_z$ and a desired final state of $\lambda(T) = -I_z$ was set for these spins. Inversion of the passive spins is sufficient for decoupling³⁷. To ensure the pulse achieves proper chemical shift encoding in the nearby methyl region, we sampled 130 spins from $\omega \in [8.5, 27.5]$ ppm and alternated the initial state $\rho(0)$ between I_x and I_y . A desired final state given by $\lambda(T) = U\rho(0)U^\dagger$ where $U = \exp(-i\omega T I_z)$ is set individually for the spins based on their intrinsic chemical shift frequency ω . The simultaneous optimization of inversion and encoding behavior ensures that decoupling does not cause any off-resonance effects on the methyl spins. The optimized pulse shape was generated using the toggling frame implementation^{37–39} of the GRAPE (gradient ascent pulse engineering) algorithm^{40–44}. The optimization was seeded with a random pulse shape. The algorithm maximizes the agreement between the target state $\lambda(T)$ and the simulated final state $\rho(T)$, by systematically changing the pulse shape using gradient ascent optimization. The resulting radio frequency amplitude and phase, and a simulation of the performance in the methyl chemical shift region are given in Supplementary Fig. 2. The various spins follow highly intricate trajectories during the pulse, but all end up at their respective desired final states at time T (Supplementary Fig. 4).

Testing the designed pulse on maltose-binding protein. We tested this method on selective ILV-methyl labeled and otherwise deuterated 42 kDa maltose-binding protein (MBP). A high-resolution HMQC (<35 Hz resolution in ^{13}C dimension) has doublets owing to the evolution of J_{CC} homonuclear coupling. This coupling was selectively refocused for valine methyl by using the selective pulse described above (Fig. 2a). The increased crowding owing to peak splitting may be undesirable for large proteins. In such cases, distinction can be achieved by imparting an opposite sign to valine versus leucine methyl resonances by using constant time encoding of methyl chemical shifts (Fig. 1a). The spectra in Fig. 2b, c were recorded using a constant-time two-dimensional (2D) HMQC pulse sequence (Supplementary Fig. 3)

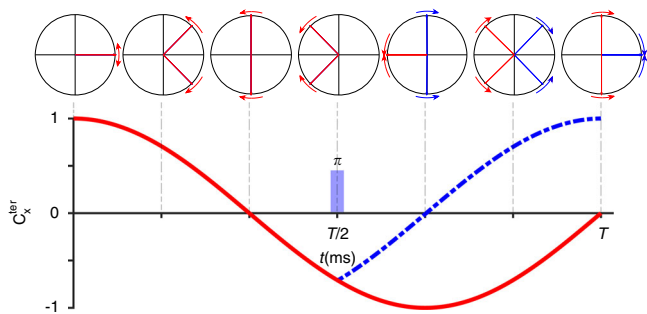


Fig. 3 Schematic explaining disappearance of isoleucine and leucine peaks when the constant-time duration is set to $T = 3/(2 \, ^1J_{CC})$. The simulation shows the transverse projection of the $^{13}\text{CH}_3$ Leu (red), Val (blue) magnetization vector, with arrows showing the sense of rotation with respect to the intrinsic chemical shift frequency. The resulting $^{13}\text{CH}_3$ -x magnetization for Leu (red), Val (blue, dashed) are shown below. The constant-time evolution period T is set to 42 ms ($3/(2 \, ^1J_{CC})$). The shaped pulse at the $T/2$ selectively decouples C_{Val}^{γ} from C_{Val}^{β} , resulting in the appearance of the valine peaks. The disappearance of isoleucine and leucine peaks in the spectrum is due to the cancelling out of the anti-phase term arising because of the ^{13}C - ^{13}C spin coupling.

with constant-time T set as $1/{}^1J_{CC}$ and $3/(2 \, ^1J_{CC})$, respectively. The latter spectrum has only valine methyl peaks, enabling identification of any leucine-valine overlapping peaks. The theory explaining the disappearance of leucine methyls when $T = 3/(2 \, ^1J_{CC})$ is shown in Fig. 3.

Validating the designed pulse on other challenging proteins.

We expanded our test set by including range of biologically relevant protein samples and different magnetic field strengths. It should be noted that the same decoupling pulse can be used at different field strengths by linearly scaling the pulse length and power level accordingly. The results observed are as expected on a suite of four proteins, namely the HNH domain of Cas9, the REC1-2 domain of Cas9, human-IL2, and eIF4A (eukaryotic initiation factor A) (Fig. 4, full spectra are shown in Supplementary Figs. 6 and 7). We used the valine methyl-selective decoupling pulse in the SOFAST-HMQC for acquiring spectra on these proteins⁴⁵ to achieve increased sensitivity per unit time. Where assignments are known, our method correctly identifies the valine and leucine peaks. In the absence of assignment, we matched the expected number of peaks for valine and leucine methyl with the positive and negative peaks in the spectrum. The broadened peaks in eIF4A reflect the dynamic nature of protein, which is confirmed in the control spectrum (without selective decoupling pulse). As described before, we can see in an example of Rec1-2 domain of Cas9 structure the valine and leucine methyl are distributed throughout the protein, enabling them to act as NMR probes for different protein domains. Another point to note from Fig. 4e is that valine and leucine are found in close proximity on the surface of the protein. As valine and leucine also overlap in chemical shift space, it is often challenging to derive resonance assignments solely based on distance constraints obtained through NOEs. Thus, the ability to distinguish them by any other NMR experiment, as shown in Fig. 4b is invaluable for obtaining unambiguous methyl assignments.

Discussion

The method we have presented here can provide amino-acid classification for resonance assignment using a variety of different NMR experiments (notably HMQC in real-time, constant-time and SOFAST). However, the basic principle is easily adapted to

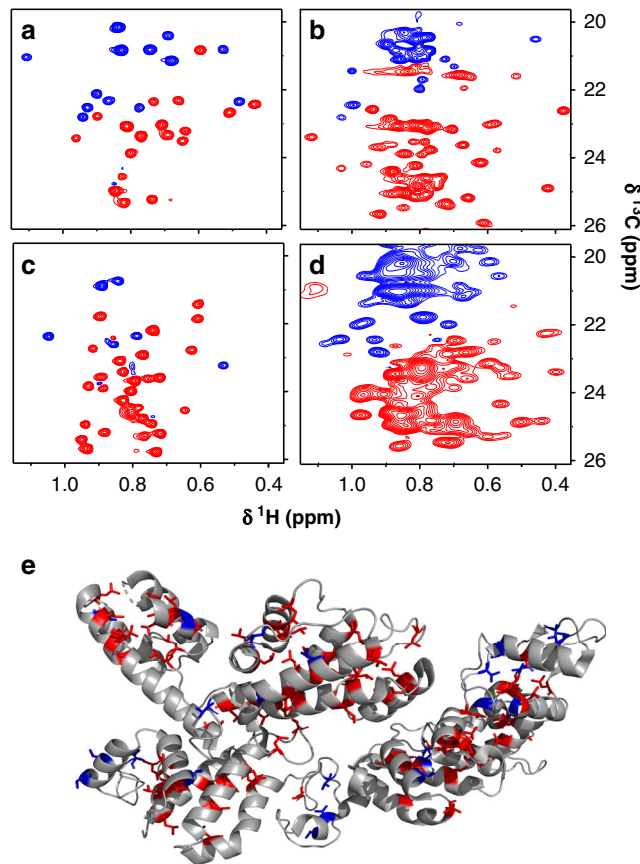


Fig. 4 Application of the valine-selective decoupling pulse on biologically relevant proteins. Constant-time 2D ^1H - ^{13}C methyl HMQC spectra of four biologically relevant proteins are shown. **a** Cas9 HNH domain (130 μM , 15 kDa); **b** Cas9 REC1-2 domain (250 μM , 52 kDa); **c** Human IL2 (920 μM , 15.4 kDa); **d** eIF4A (50 μM , 46 kDa). Unambiguous distinction between the valine (blue) and leucine (red) resonances is achieved. Only Leu-Val overlapping region of the HMQC spectrum is shown here. The full spectra of these proteins are presented in Supplementary Figs. 6 and 7. **e** Here we show the 3D structure for the REC1-2 domain of Cas9 (PDB ID: 4UN3)⁵² where the valine and leucine residues are highlighted in blue and red, respectively.

related contexts, provided that there is an indirect encoding delay for the carbon methyl resonances for example, depending on choice of constant-time duration and/or phase cycle, either valine or leucine peaks can be selectively removed or moved in the spectrum. We can also selectively remove valine methyl peaks from an HMQC spectrum by taking difference of two spectra recorded with and without the pulse in an interleaved manner. We can selectively shift the resonances of valine methyl to another less crowded part of the spectrum; for example, by using the decoupling pulse only in one of the quadrature components of the indirect acquisition scheme, or only on every second indirect increment⁴⁶. Such tricks can be used to obtain clear distinction between overlapping leucine and valine methyl peaks. This may be useful to alleviate crowding or ambiguity in NOE experiments or in chemical shift perturbation measurements.

The same selective pulse can be used at different magnetic field strengths by proper scaling of pulse duration and pulse power. We acquired the data for MBP on a 750 MHz NMR spectrometer and for other proteins on an 800 MHz NMR spectrometer. In addition, we acquired constant-time HMQC with constant time period set to 28 ms for MBP on a 600 MHz and an 800 MHz NMR spectrometer, with the selective pulse, to demonstrate the

pulse applicability across different magnetic field strengths (Supplementary Fig. 8). The pulse duration and power for different magnetic field strengths are listed in Supplementary Table 1.

The method can be used in both real time and constant time versions of the HMQC as shown above. In the real-time experiments, the differentiation can be seen in the peak shape, by the presence or absence of the splitting. In particular, the valine residues will appear as a singlet and the leucine as a doublet. In the constant time implementation, the differentiation is much more dramatic and there is a sign inversion between the leucine and valines peaks. However, the constant-time implementation suffers from relaxation losses when we deal with very large proteins, which can be compensated by additional scans and the use of non-uniform sampling (NUS).

We theoretically estimated transverse relaxation for perdeuterated protein where only CD1 of Ile, CG1 of Val, and CD1 of Leu protonated. We used the following relaxation estimates from a recent review by Schutz et al.⁴⁷ For a protein tumbling with rotational tumbling time 20 ns, the transverse relaxation rate (R_2) of CH multiple quantum coherence is estimated to be ~5 Hz. And for this R_2 of 5 Hz, the loss of signal height due to constant time encoding relative to real time encoding is 8%. Similarly, for a protein tumbling with rotational tumbling time 180 ns, the transverse relaxation rate (R_2) of CH multiple quantum coherence is estimated to be ~42 Hz. And for this $R_2 = 42$ Hz, the loss of signal height due to constant time encoding relative to real time encoding is 46%. To put into perspective of molecular size, 180 ns is the tumbling time of proteasome (650 kDa) at 65 °C. Here the user can employ a real time implementation of the experiment, which will allow to distinguish leucine from valine or use constant time implementation with more scans and additionally use NUS for time savings.

The implementation described here, in either HMQC or HSQC form, can be appended to NOESY experiments; for instance, a 3D NOESY-HSQC or a 4D HMQC-NOESY-HMQC. These types of experiments can be used to distinguish leucine-derived spatial constraints from those originating from valine. Simple, reliable amino-acid determination promises to improve the utility and reliability of existing and emerging methyl-NMR assignment software. Amino-acid selectivity can also be used to simplify and clarify methyl-NMR spectra in a range of experimental contexts.

Methods

Streptococcus pyogenes Cas9. The coding sequences for Cas9 HNH domain (residues 776–908) and RECI-2 domain (residues 56–505) were synthesized and codon optimized for expression in *Escherichia coli* (Supplementary Table 3). HNH domain was expressed with a C-terminal 6× His tag, and RECI-2 domain was fused with a MBP tag, an N-terminal 6× His and a tobacco etch virus (TEV) protease cleavage site. Each protein was expressed in *E. coli* BL21 (DE3) containing chaperone plasmid pG-KJE8 (TAKARA, 3340) to enhance protein folding. In brief, when cells reached an OD₆₀₀ of ~0.6, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce protein expression. Cells were then grown for an additional 18 h at 23 °C. Collected cells were resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% (v/v) glycerol and 1 mM Tris(2-carboxy-ethyl) phosphine (TCEP) containing an ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor tablet (Roche). The cell suspension was sonicated on ice and clarified by centrifugation at 15,000 × *g* for 30 min. The soluble lysate fraction was bound in batch to nickel-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen). The resin was washed extensively with 20 mM Tris, pH 7.5, 500 mM NaCl, 1 mM TCEP, 10 mM imidazole, and 5% (vol/vol) glycerol, and the bound protein was eluted in 20 mM Tris, pH 7.5, 500 mM NaCl, 1 mM TCEP, 300 mM imidazole, and 5% (vol/vol) glycerol. For RECI-2 domain, the His₁₀-MBP affinity tag was removed with His₁₀-tagged TEV protease (1 mg of TEV protease was added per 50 mg of fusion protein) during overnight dialysis against 20 mM Tris, pH 7.5, 500 mM NaCl, 1 mM TCEP, and 5% (vol/vol) glycerol. The protein was then flowed over Ni-NTA agarose to remove TEV protease and the cleaved affinity tag and further purified by size-exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare) in 20 mM Tris, pH 7.5, 200 mM KCl, 1 mM TCEP, and 5% (vol/vol) glycerol. For HNH domain, following the Ni-NTA

purification, the protein was further purified on a Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES (pH 7.5) and 80 mM KCl. Protein concentrations were determined using A₂₈₀ measurements on a NanoDrop with extinction coefficients estimated with the ExpASY ProtParam tool.

Interleukin-2. Codon optimized DNA encoding the human IL2 (Supplementary Table 3) was expressed in *E. coli* BL21 (DE3) competent cells as inclusion bodies. Protein expression was achieved by induction with 1 mM IPTG at an OD₆₀₀ of 0.6 followed by cell growth at 37 °C for 5 h at 200 r.p.m. For in vitro refolding, ~30 mg of inclusion bodies was dropped diluted into 200 mL of refolding buffer (1.1 M guanidine, 6.5 mM cysteamine, 0.65 mM cystamine, 110 mM Tris pH 8.0) at 4 °C while stirring. Refolding proceeded overnight at 4 °C without stirring. The solution was dialyzed into a buffer of 20 mM MES, pH 6.0, 25 mM sodium chloride. Purification of refolded IL2 was performed by cation exchange chromatography with a CAPTO-SP column using a 25 mM to 1 M NaCl gradient in a buffer with 25 mM MES pH 6.0 followed by size-exclusion chromatography on a Superdex 75 column (GE) in 50 mM NaCl, 20 mM sodium phosphate, pH 6.0.

HNH, RECI-2 and hIL-2 were overexpressed in M9 minimal media culture in ²H₂O containing 2 g l⁻¹ ²H¹³C glucose (Sigma #552151) and 1 g l⁻¹ ¹⁵NH₄Cl (Sigma #299251). Selective methyl labeling referred to as ILV*, was achieved by the addition of appropriate precursors (ISOTEC Stable Isotope Products (Sigma-Aldrich) as detailed previously^{48,49}. The selective labeling of ILV* methyls (Ile ¹³Cδ₁ only; Leu ¹³Cδ₁/¹³Cδ₂; Val ¹³Cγ₁/¹³Cγ₂ in an otherwise U-[¹⁵N, ¹³C, ²H] background) was achieved by adding 60 mg l⁻¹ 2-ketobutyric acid-¹³C_{4,3,3-2H₂ (Sigma #607541) for Ile and 120 mg l⁻¹ 2-keto-3-(methyl-d3)-butyric acid-1,2,3,4-¹³C₄, 3-2H (Sigma #637858) for Leu/Val 1 h prior to induction. For each uniformly ILV*-labeled protein sample, the concentration and buffer composition are as follows:}

- The 0.13 mM HNH domain (15 kDa) in 20 mM HEPES (pH 7.5), 80 mM KCl, 0.01% Na₃N, 10% D₂O.
- The 0.25 mM RECI-2 domain (52 kDa) in 20 mM HEPES (pH 7.5), 200 mM KCl, 5% deuterated glycerol-d₈, 1 mM TCEP, 0.01% Na₃N, 10% D₂O.
- The 0.9 mM hIL-2 (15 kDa) in 20 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.01% Na₃N, 10% D₂O.

MBP. Several single colonies of BL21 (DE3) cells carrying MBP encoding plasmid was grown overnight at 37 °C in 10 ml LB media with 100% D₂O. The cells were pelleted and resuspended in 10 mL M9 medium in D₂O containing 2 g of U[¹³C, ²H] labeled glucose and U[¹⁵N] labeled NH₄Cl and grown for 12 hours at 37 °C. The culture was expanded to 100 mL in the same medium and continue to grow at 37 °C. When OD₆₀₀ reach 0.4, 70 mg/l of α-ketobutyric acid (CDLM-4611 from Cambridge Isotope Laboratories) and 120 mg/l of α-ketoisovaleric acid (CDLM-4418 from Cambridge Isotope Laboratories) was added. Both these precursors were uniformly ¹³C labeled and deuterated, with the exception of the methyl group, which was protonated. The protein expression was conducted for overnight at 18 °C after induced with 1 mM IPTG at OD₆₀₀ 0.6–0.8. The cells were harvested and lysed using sonication. The lysate was passed through immobilized amylose beads and was eluted using 10 mM maltose in 50 mM Tris-HCl, pH 8.0. The eluted protein was buffer exchanged into 10 mM HEPES, 1 mM EDTA, pH 6.5. The amides were back exchanged with protons by incubating it with 1 M urea for 24 h at 37 °C. This was followed by buffer exchange to NMR buffer and further purified on size-exclusion column (Superdex 75 10/300 GL). The final sample was prepared by concentrating MBP to 0.5 mM into NMR buffer (1 mM EDTA, 2 mM β-cyclodextrin, 10 mM HEPES, pH 6.5).

eIF4A. The human translation initiation factor eIF4A (46 kDa) was expressed using a N-terminal His₆-GB1-tag followed by TEV cleavage site in *E. coli*. Protein expression and isotope labeling were performed as described above for the case of MBP, except that 500 mL cell culture was used instead. eIF4A was purified by loading the lyse in buffer A (25 mM sodium phosphate, pH 7.5, 300 mM NaCl, 2 mM beta mercaptoethanol, 5% glycerol) plus 10 mM imidazole to Ni-NTA resin, washed with buffer A containing 30 mM imidazole and eluted with buffer A plus 300 mM imidazole. The elution was dialyzed in presence of His₆-TEV protease at 1:40 (w/w) ratio overnight against 4 L buffer A. When the cleavage of GB1-tag is completed, the reaction mixture was purified by running through fresh Ni-NTA resin. Flow-through containing eIF4A was collected, concentrated and further polished using superdex 75 10/300 GL column equilibrated in buffer B (20 mM Tris-HCl at pH 7.5, 150 mM NaCl, 5 mM DTT). The final NMR sample is prepared at 50 μM concentration in buffer B with 10% D₂O.

NMR data acquisition and analysis. All NMR experiments were acquired on Bruker NMR machines operating at 600 MHz, 750 MHz, or 800 MHz ¹H Larmor frequency and equipped with cryogenically cooled triple resonance 5 mm probes. The samples were doped with 5% D₂O for the purpose of locking. 2D ¹H-¹³C HMQC or ¹H-¹³C SOFAST-HMQC NMR experiments were recorded on different protein samples with the acquisition parameters provided in Supplementary Table 2. The spectra were acquired using Topspin 3 acquisition software from

Bruker. They were processed using NMRPipe⁵⁰ and analyzed using in house MATLAB scripts and CCPNMR⁵¹. The selective homonuclear decoupling pulse, real-time and constant-time version of HMQC pulse sequence used in this manuscript is provided as Supplementary Data 1–3, respectively.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The shaped file used in this manuscript is provided as Supplementary Data 1. NMR data sets are available from the corresponding authors upon reasonable request. Pulse sequences, shaped pulse files, and related parameter sets, along with detailed instructions can be downloaded from the lab website <http://artlab.dana-farber.org/downloads.html>.

Code availability

Pulse sequences are provided as Supplementary Data 2 and 3.

Received: 25 May 2020; Accepted: 1 October 2020;

Published online: 03 November 2020

References

- Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814 (2009).
- Barrett, P. J. et al. The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry* **52**, 1303–1320 (2013).
- Bax, A. & Clore, G. M. Protein NMR: boundless opportunities. *J. Magn. Reson.* **306**, 187–191 (2019).
- Kay, L. E. Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view. *J. Magn. Reson.* **210**, 159–170 (2011).
- Alderson, T. R. & Kay, L. E. Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.* **60**, 39–49 (2020).
- Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–968 (2011).
- Mallis, R. J. et al. NMR: an essential structural tool for integrative studies of T cell development, pMHC ligand recognition and TCR mechanobiology. *J. Biomol. NMR* **73**, 319–332 (2019).
- Tugarinov, V., Sprangers, R. & Kay, L. E. Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 1743–1750 (2007).
- Mas, G. et al. Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci. Adv.* **4**, eaau4196 (2018).
- Rosenzweig, R. & Kay, L. E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* **83**, 291–315 (2014).
- Ruschak, A. M. & Kay, L. E. Proteasome allostery as a population shift between interchanging conformers. *Proc. Natl Acad. Sci.* **109**, E3454–E3462 (2012).
- Gorman, S. D., Sahu, D., O'Rourke, K. F. & Boehr, D. D. Assigning methyl resonances for protein solution-state NMR studies. *Methods* **148**, 88–99 (2018).
- Gans, P. et al. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem.* **122**, 2002–2006 (2010).
- Tugarinov, V. & Kay, L. E. An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* **28**, 165–172 (2004).
- Lichtenecker, R., Ludwiczek, M. L., Schmid, W. & Konrat, R. Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing. *J. Am. Chem. Soc.* **126**, 5348–5349 (2004).
- Kerfah, R., Plevin, M. J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113–122 (2015).
- Wen, J., Zhou, P. & Wu, J. Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl–methyl NOESY for large proteins. *J. Magn. Reson.* **218**, 128–132 (2012).
- Amero, C. et al. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR* **50**, 229–236 (2011).
- Chao, F.-A. et al. FLAMEnGO 2.0: an enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. *J. Magn. Reson.* **245**, 17–23 (2014).
- Poulsen, F. M., Hoch, J. C. & Dobson, C. M. Structural study of the hydrophobic box region of lysozyme in solution using nuclear Overhauser effects. *Biochemistry* **19**, 2597–2607 (1980).
- Xu, Y. & Matthews, S. MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins. *J. Biomol. NMR* **55**, 179–187 (2013).
- Monneau, Y. R. et al. Automatic methyl assignment in large proteins by the MAGIC algorithm. *J. Biomol. NMR* **69**, 215–227 (2017).
- Pritišanac, I., Alderson, T. R. & Güntert, P. Automated assignment of methyl NMR spectra from large proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **118–119**, 54–73 (2020).
- Pritišanac, I. et al. Automatic assignment of methyl-NMR spectra of supramolecular machines using graph theory. *J. Am. Chem. Soc.* **139**, 9523–9533 (2017).
- Pritišanac, I., Würz, J. M., Alderson, T. R. & Güntert, P. Automatic structure-based NMR methyl resonance assignment in large proteins. *Nat. Commun.* **10**, 1–12 (2019).
- Sherman, B. C. *Automated Resonance Assignment Via Boolean Satisfiability and Quadratic Programming*. (University of California, Santa Cruz, 2019).
- Lichtenecker, R. J., Coudeville, N., Konrat, R. & Schmid, W. Selective isotope labelling of leucine residues by using α -ketoacid precursor compounds. *ChemBioChem* **14**, 818–821 (2013).
- Lichtenecker, R. J. et al. Independent valine and leucine isotope labeling in *Escherichia coli* protein overexpression systems. *J. Biomol. NMR* **57**, 205–209 (2013).
- Senn, H. et al. Stereospecific assignment of the methyl ¹H NMR lines of valine and leucine in polypeptides by nonrandom ¹³C labelling. *FEBS Lett.* **249**, 113–118 (1989).
- Hu, W., Namanja, A. T., Wong, S. & Chen, Y. Selective editing of Val and Leu methyl groups in high molecular weight protein NMR. *J. Biomol. NMR* **53**, 113–124 (2012).
- Kerfah, R., Hamelin, O., Boisbouvier, J. & Marion, D. CH 3-specific NMR assignment of alanine, isoleucine, leucine and valine methyl groups in high molecular weight proteins using a single sample. *J. Biomol. NMR* **63**, 389–402 (2015).
- Siemons, L., Mackenzie, H. W., Shukla, V. K. & Hansen, D. F. Intra-residue methyl–leucine correlations for valine and leucine residues in large proteins from a 3D-HMBC-HMQC experiment. *J. Biomol. NMR* **73**, 749–757 (2019).
- Ulrich, E. L. et al. BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408 (2007).
- Guo, C. & Tugarinov, V. Selective ¹H-¹³C NMR spectroscopy of methyl groups in residually protonated samples of large proteins. *J. Biomol. NMR* **46**, 127–133 (2010).
- Van Melckebeke, H., Simorre, J.-P. & Brutscher, B. Amino acid-type edited NMR experiments for methyl–methyl distance measurement in ¹³C-labeled proteins. *J. Am. Chem. Soc.* **126**, 9584–9591 (2004).
- Van Melckebeke, H., Simorre, J.-P. & Brutscher, B. Suppression of artifacts induced by homonuclear decoupling in amino-acid-type edited methyl ¹H-¹³C correlation experiments. *J. Magn. Reson.* **170**, 199–205 (2004).
- Coote, P. W. et al. Optimal control theory enables homonuclear decoupling without Bloch–Siegert shifts in NMR spectroscopy. *Nat. Commun.* **9**, 1–9 (2018).
- Bloch, F. & Siegert, A. Magnetic resonance for nonrotating fields. *Phys. Rev.* **57**, 522 (1940).
- Coote, P., Anklin, C., Masefski, W., Wagner, G. & Arthanari, H. Rapid convergence of optimal control in NMR using numerically-constructed toggling frames. *J. Magn. Reson.* **281**, 94–103 (2017).
- Conolly, S., Nishimura, D. & Macovski, A. Optimal control solutions to the magnetic resonance selective excitation problem. *IEEE Trans. Med. Imag.* **5**, 106–115 (1986).
- Skinner, T. E., Reiss, T. O., Luy, B., Khaneja, N. & Glaser, S. J. Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR. *J. Magn. Reson.* **163**, 8–15 (2003).
- Skinner, T. E., Reiss, T. O., Luy, B., Khaneja, N. & Glaser, S. J. Reducing the duration of broadband excitation pulses using optimal control with limited RF amplitude. *J. Magn. Reson.* **167**, 68–74 (2004).
- Mao, J., Mareci, T. H., Scott, K. N. & Andrew, E. R. Selective inversion radiofrequency pulses by optimal control. *J. Magn. Reson.* (1969) **70**, 310–318 (1986).
- Rosenfeld, D. & Zur, Y. Design of adiabatic selective pulses using optimal control theory. *Magn. Reson. Med.* **36**, 401–409 (1996).
- Schanda, P., Kupče, Ě. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
- Sakhaii, P. & Bermel, W. A different approach to multiplicity-edited heteronuclear single quantum correlation spectroscopy. *J. Magn. Reson.* **259**, 82–86 (2015).
- Schütz, S. & Sprangers, R. Methyl TROSY spectroscopy: a versatile NMR approach to study challenging biological systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **116**, 56–84 (2020).

48. De Paula, V. S. et al. Interleukin-2 druggability is modulated by global conformational transitions controlled by a helical capping switch. *Proc. Natl Acad. Sci.* **117**, 7183–7192 (2020).
49. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **1**, 749–754 (2006).
50. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
51. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
52. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).

Acknowledgements

H.A. acknowledges funding from the Claudia Adams Barr Program for Innovative Cancer Research and from NIH (GM136859). G.W. acknowledges support from NIH grant GM129026 and GM132079. This paper is dedicated to the memory of our dear friend Dr. Hanudatta Atreya, who recently passed away.

Author contributions

S.P.B., A.D., P.W.C., G.W., H.A., designed the experiments and wrote the manuscript. A.D., W.B., and H.A., designed the NMR pulse sequence. S.P.B. and P.W.C., designed the pulse. W.C., M.Z., V.S.D., prepared the samples. A.D., V.S.D., N.G.S., W.B., and H.A., performed the NMR experiments. All authors analyzed NMR data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19325-4>.

Correspondence and requests for materials should be addressed to P.W.C. or H.A.

Peer review information *Nature Communications* thanks D Flemming Hansen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020