Protein–Protein Interfaces Probed by Methyl Labeling and Proton-Detected Solid-State NMR Spectroscopy

Maximilian Zinke,[a] Pascal Fricke,[a] Sascha Lange,[a] Sophie Zinn-Justin,[c] and Adam Lange*[a, b]

Proton detection and fast magic-angle spinning have advanced biological solid-state NMR, allowing for the backbone assignment of complex protein assemblies with high sensitivity and resolution. However, so far no method has been proposed to detect intermolecular interfaces in these assemblies by proton detection. Herein, we introduce a concept based on methyl labeling that allows for the assignment of these moieties and for the study of protein–protein interfaces at atomic resolution.

Solid-state NMR (ssNMR) spectroscopy is a powerful technique to study the structure and dynamics of macromolecular protein assemblies, such as bacterial cytoskeletal filaments,[1] phage capsids[2] and amyloid fibrils.[3–8] Although ssNMR linewidths, in contrast to solution NMR, are not limited by the molecular weight of the protein complex, many protein samples remain beyond the scope of this method. This is due to heterogeneous line-broadening, a sample-dependent property, and the spectral crowding caused by a large number of residues. These effects lead to peak overlap, rendering spectral data ambiguous, and ssNMR as a tool ineffective for many biological systems. Recently, protein deuteration and magic angle spinning (MAS) at frequencies at or above 40 kHz have enabled proton-detected experiments that offer a higher sensitivity and dimensionality than their traditional carbon-detected counterparts.[9–12] While these methodologies allow for unambiguous spectral assignments (via a "backbone walk") of large proteins,[13–15] structural data such as long-distance restraints, which are required for 3D structure determination, are still collected in a highly ambiguous and/or insensitive manner. To this date, most restraints are extracted from carbon-detected experiments (e.g. 2D carbon-carbon correlations with homonuclear mixing

through space)[16,17] or proton-detected 4D HNhhNH experiments with radio frequency driven recoupling mixing (RFDR mixing)[18] between protons (correlating amide groups through space).[19,20] The latter experiment offers excellent resolution due to its four-dimensionality but lacks sensitivity compared to 3D proton-detected experiments (since increasing dimensionality decreases signal-to-noise).[21] However, its 3D version, the 3D HNhH, yields highly ambiguous data for uniformly labeled proteins due to the ubiquity of protons in proteins. If unambiguous data are to be obtained in 2D or 3D proton-detected experiments, this methodology clearly needs to be combined with sophisticated labeling strategies and experimental designs.[22]

An exceptional advantage of ssNMR as a structural biology method is that it allows for the study of protein-protein interfaces by heterogeneous isotope labeling.[23,24] Traditionally, these structural features have been analyzed by carbon-detected experiments such as spin dilution experiments[4] and mixed labeling schemes with 50% ^{13}C labeled protein species and 50% ^{15}N labeled protein species (e.g. through NhhC experiments[25]). Due to their carbon-detected nature, these experiments suffer from low sensitivity and dimensionality, requiring many weeks of data acquisition, and often resulting in highly ambiguous data.

In this work, we use methyl labeling[26]—which has proven itself as an indispensable tool in solution NMR of large proteins[27,28]—to simplify spectral data and facilitate proton-detection.[29,30] The methodology presented here allows for the collection of long-distance restraints with 3D proton-detected ssNMR experiments between the following: 1.) methyl and amide groups globally; 2.) methyl and amide groups at protein-protein interfaces.

Additionally, we present an assignment method for isoleucine Cδ1 methyl groups by their correlation with backbone resonances (which ought to be known from a preliminary backbone walk) using a specific isoleucine precursor ("assignment precursor") and homonuclear transfer steps. In this manner, time-consuming assignment by mutagenesis can be avoided, opening the way to study assemblies with very large subunits and a high number of isoleucine residues. The presented strategy is used to assign the isoleucine Cδ1 methyl groups in gp17.1, the major tail tube protein of the bacteriophage SPP1,[31] and to identify a protein–protein interface within the gp17.1 polymer that is crucial for protein oligomerization.

By supplementation of the precursor 2-Ketobutyric acid-4-^{13}C-3,3-d$_2$ during protein expression in deuterated media,[32]

[a] M. Zinke, Dr. P. Fricke, Dr. S. Lange, Prof. A. Lange
Department of Molecular Biophysics
Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP), Berlin, Germany
E-mail: alange@fmp-berlin.de
[b] Prof. A. Lange
Institut für Biologie, Humboldt-Universität zu Berlin, Berlin, Germany
[c] Dr. S. Zinn-Justin
Institute for Integrative Biology of the Cell (I2BC)
CEA, CNRS, Université Paris-Sud
Université Paris-Saclay, Gif-sur-Yvette Cedex, France

isoleucine Cδ1 methyl groups are ^{13}C and ^{1}H labeled, and thus rendered "NMR-visible" (Figure 1A and Figure S3 in the Supporting Information). Isoleucine residues are well distributed over the primary structure of gp17.1 (Figure 1B) and are, due to their limited number, fully resolved by their Hδ1 chemical shift as shown in the 2D hCH correlation spectrum in Figure 1C (each isoleucine Cδ1 methyl group in gp17.1 results in one peak in the spectrum). If all subunits in the prepared gp17.1 protein assembly are ^{15}N and isoleucine Cδ1 methyl labeled, proton–proton mixing from amide to methyl protons can occur within and in between gp17.1 subunits (left panel of Figure 1D). If half of the subunits are ^{15}N labeled and the other half are isoleucine Cδ1 methyl labeled, this mixing can only occur between two different subunits (right panel of Figure 1D). Figure 1E shows 2D planes of two 3D HNhH experiments with RFDR mixing superimposed with the 2D hNH fingerprint spectrum of gp17.1 (grey spectrum in the background, Figure S1). The 2D plane corresponding to the Hδ1 methyl frequency of Ile18 (as represented by the dashed blue line) shows magnetization transfer from nearby amide protons to the Hδ1 protons of Ile18. In the left spectrum (homogeneous labeling) all nearby amide groups are visible. However, in the right spectrum (mixed labeling) only intermolecular contacts are visible, limiting the number of peaks, and revealing a protein-protein interface between Ile18 and the C-terminus of gp17.1.

For the isoleucine Cδ1 methyl assignment, the precursor *2-Ketobutyric acid-^{13}C$_4$-3,3-d$_2$* ("assignment precursor", Figure S3)

is supplemented during protein expression, which in addition to the NMR-visible isoleucine Cδ1 methyl group leads to the incorporation of further ^{13}C species at the Cγ1, Cα and CO sites of all isoleucine residues, without concomitant introduction of additional protons. Using a long cross polarization (CP) step of 6 ms, magnetization can be transferred from Hδ1 to the labeled carbons in the isoleucine side chain and back to the methyl protons resulting in a 2D hCH correlation spectrum as shown in Figure 2A. This procedure enables the correlation of all labeled carbon atoms within the isoleucine side chain and Cα as part of the backbone. Additionally, in a second experiment, magnetization is transferred from Hδ1 to Cγ1, and subsequently to Cα by homonuclear dipolar recoupling enhanced by amplitude modulation (DREAM) transfer[33] traversing the unlabeled Cβ through space - a transfer which would be impossible in solution NMR by through-bond mixing. By cαN and nH CP steps, magnetization is finally transferred to the backbone amide protons for detection, resulting in a 2D HccanH correlation spectrum (Figures 2B and S2). These two experiments allow for a straightforward assignment of the isoleucine Cδ1 methyl groups by their correlation with backbone resonances. If these spectra are not sufficient for assignment, the 2D HccanH spectrum can be expanded by further indirect dimensions on either Cα or amide N.

The described methodology allowed us to assign all visible isoleucine Cδ1 methyl groups, including multiple conformations, of gp17.1 (cf. the Supporting Information Table S4) and, on that basis, to identify 105 unambiguous long-range
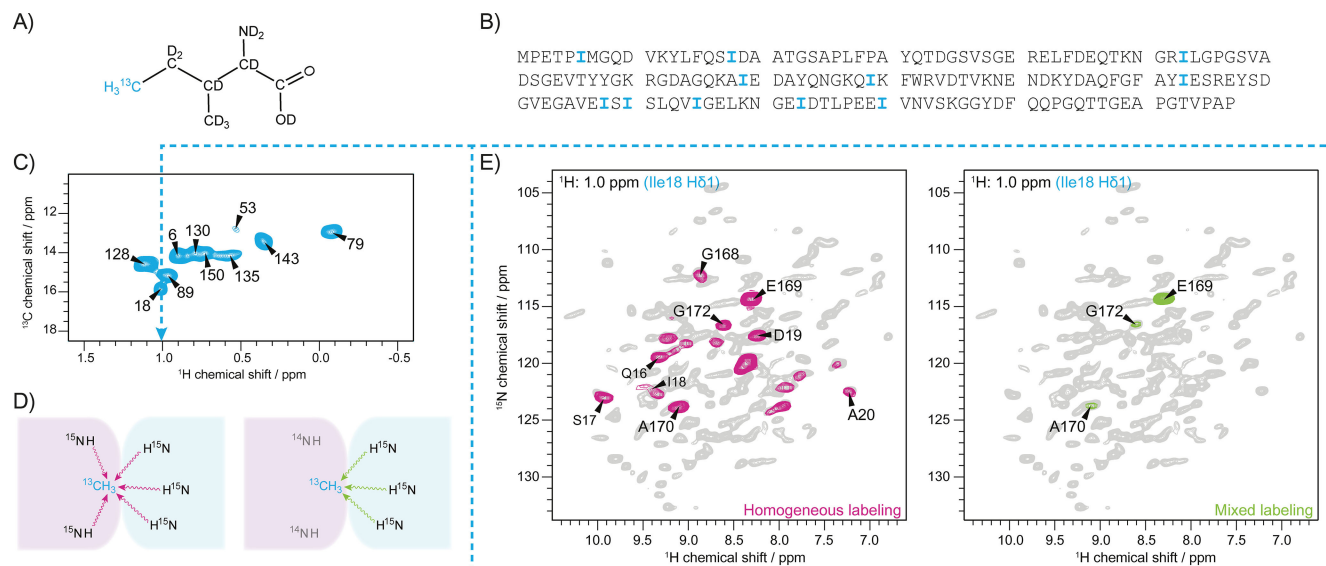

Figure 1. Isoleucine-methyl labeling as a tool to map protein interfaces. A) 2-Ketobutyric acid-4-^{13}C-3,3-d$_2$ as a precursor leads to labeling of isoleucines in the protein as indicated. Only the Cδ1 methyl group carries a ^{13}C isotope as well as protons and is thus, by the choice of suitable experiments, "NMR visible". B) Isoleucine residues are well distributed over the primary structure of gp17.1 C) 2D hCH correlation spectrum of isoleucine-Cδ1 labeled gp17.1 at 40 kHz MAS and 900 MHz external magnetic field. The limited number of isoleucines in gp17.1 results in fully resolved unique resonances in the proton dimension. D) Schematic representation of the labeling strategy to identify intermolecular contacts. In the first sample (left), all subunits are ^{15}N and Ile-Cδ1 labeled. Magnetization transfer through space occurs between and within subunits. In the second sample (right), half of the subunits are ^{15}N and the other half Ile-Cδ1 labeled. Magnetization transfer through space can only occur between different subunits. E) 2D planes of 3D HNhH spectra with RFDR mixing superimposed with a 2D hNH correlation spectrum. The 2D planes correspond to the ^{1}H methyl frequency of Ile18 (as represented by the blue dashed line) showing magnetization transfer from amide protons to the protons attached to Ile18 Cδ1. In the left case, all NH groups close to the methyl group of Ile18 are revealed. However, in the right case, only intermolecular contacts can be detected limiting the amount of peaks. This approach reveals a protein-protein interface between Ile18 of one subunit and the C-terminus (E169, A170 and G172) of another subunit.
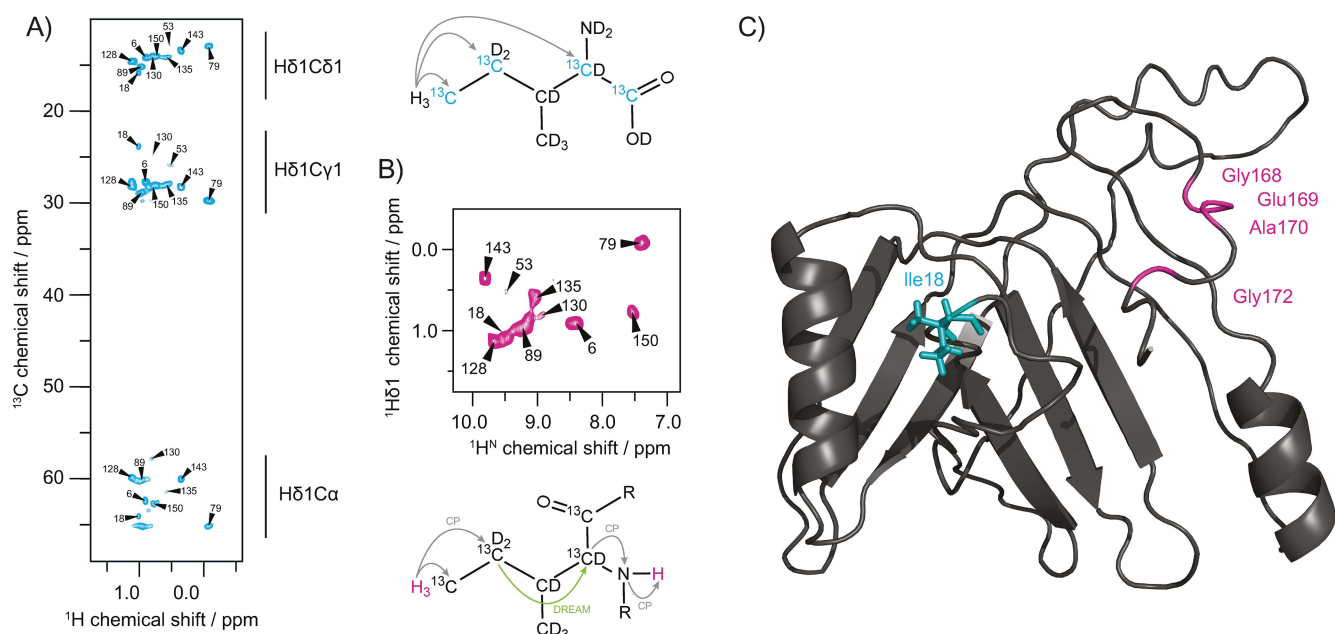
Figure 2. Assignment of isoleucine Cδ1 methyl groups and location of the protein–protein interface. A) 2D hCH correlation spectrum of isoleucine Cδ1 methyl labeled gp17.1 at 40 kHz MAS and 900 MHz external magnetic field. 2-Ketobutyric acid-$^{13}C_4$-3,3-d_2 was used as a precursor resulting in the labeling of isoleucine residues as shown. A 6 ms hC cross-polarization time enabled magnetization transfer from the Hδ1 methyl protons throughout the side-chain and back. B) 2D HccanH spectrum to correlate methyl Hδ1 protons with their backbone amide protons for assignment. Magnetization is initially transferred from Hδ1 to Cδ1 and Cγ1 via CP, and subsequently to Cα via DREAM. Finally, magnetization transfer to H^N is achieved by cαN and nH CP. C) Mapping of the determined tail tube interface onto the homology model of a polymerized gp17.1 subunit (taken from Langlois et al.[31]). The location of the Ile18 residue in the middle of the subunit suggests that the C-terminus forms a molecular bridge that embraces the next monomer within the tail tube quaternary structure.

restraints between amides and isoleucine Cδ1 methyl groups, 7 of which are unambiguously intermolecular (Figure S4, including chemical shift-ambiguous restraints: 179/12). The determined intermolecular contacts between Ile18 and Glu169, Ala170 and Gly172 (and Gly168 which is below the contour levels in Figure 1E) are intriguing since the C-terminus is predicted to be disordered and lacks homology to any other structurally-described bacteriophage tail protein. Also, Ile18 is not located in the outer β-strands of the homology model of polymerized gp17.1, which would resemble a "classic" protein interface in such an assembly, but in the middle of the subunit (Figure 2C). To check the significance of this newly discovered interface, we removed the last 7 and 14 C-terminal residues of gp17.1 (1–170 and 1–163) and observed that both mutants are incapable of polymerization, but remain monomeric (see size-exclusion chromatography data in Figure S5 and solution NMR data in Figure S6 in the Supporting Information).[34] This emphasizes the proposition that the C-terminus forms a molecular bridge that embraces the next monomer within the tube,[35] similar to what was described for the N-terminus of the T4 phage tail tube protein gp19.[36]

We expect that the methodology introduced here will turn out useful in a variety of solid-state NMR investigations of supramolecular assemblies or membrane-integrated proteins or protein complexes. We note that our straightforward approach to assign isoleucines can be modified to assign the methyl groups of leucine and valine (LV), threonine, alanine or methionine. This makes tedious mutation studies for methyl group assignment redundant, provided that the backbone assignment is available. Furthermore, we think that methyl groups as tools for long-distance restraints will become indispensable for the analysis of similar or even larger protein systems. The mixed-labeling strategy for the identification of protein–protein interfaces may also be extended to include differently methyl-labeled subunits, e.g 50% isoleucine Cδ1 methyl labeled; 50% LV methyl labeled.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

[1] C. Shi, P. Fricke, L. Lin, V. Chevelkov, M. Wegstroth, K. Giller, S. Becker, M. Thanbichler, A. Lange, *Sci. Adv.* **2015**, *1*, e1501087.

[2] T. Aharoni, A. Goldbourt, *Chem. – A Eur. J.* **2018**, DOI 10.1002/chem.201800532.

[3] C. M. Quinn, T. Polenova, *Q. Rev. Biophys.* **2017**, *50*, 1–44.

[4] A. K. Schütz, T. Vagt, M. Huber, O. Y. Ovchinnikova, R. Cadalbert, J. Wall, P. Güntert, A. Böckmann, R. Glockshuber, B. H. Meier, *Angew. Chem. Int. Ed.* **2015**, *54*, 331–335; *Angew. Chem.* **2015**, *127*, 337–342.

[5] M. T. Colvin, R. Silvers, B. Frohm, Y. Su, S. Linse, R. G. Griffin, *J. Am. Chem. Soc.* **2015**, *137*, 7509–7518.

[6] J. R. Lewandowski, M. E. Halse, M. Blackledge, L. Emsley, *Science.* **2015**, *348*, 578–581.

[7] T. Theint, P. S. Nadaud, D. Aucoin, J. J. Helmus, S. P. Pondaven, K. Surewicz, W. K. Surewicz, C. P. Jaroniec, *Nat. Commun.* **2017**, *8*, DOI 10.1038/s41467-017-00794-z.

[8] D. T. Murray, M. Kato, Y. Lin, K. R. Thurber, I. Hung, S. L. Mcknight, R. Tycko, D. T. Murray, M. Kato, Y. Lin, *Cell* **2017**, 171, 615–627.

[9] E. Barbet-Massin, A. J. Pell, J. S. Retel, L. B. Andreas, K. Jaudzems, W. T. Franks, A. J. Nieuwkoop, M. Hiller, V. Higman, P. Guerry, *J. Am. Chem. Soc.* **2014**, *136*, 12489–12497.

[10] D. H. Zhou, A. J. Nieuwkoop, D. A. Berthold, G. Comellas, L. J. Sperling, M. Tang, G. J. Shah, E. J. Brea, L. R. Lemkau, C. M. Rienstra, *J. Biomol. NMR* **2012**, *54*, 291–305.

[11] V. Chevelkov, B. Habenstein, A. Loquet, K. Giller, S. Becker, A. Lange, *J. Magn. Reson.* **2014**, *242*, 180–188.

[12] S. Penzel, A. A. Smith, V. Agarwal, A. Hunkeler, M. L. Org, A. Samoson, A. Böckmann, M. Ernst, B. H. Meier, *J. Biomol. NMR* **2015**, *63*, 165–186.

[13] H. Fraga, C. A. Arnaud, D. F. Gauto, M. Audin, V. Kurauskas, P. Macek, C. Krichel, J. Y. Guan, J. Boisbouvier, R. Sprangers, *ChemPhysChem* **2017**, *18*, 2697–2703.

[14] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307–1311.

[15] P. Fricke, V. Chevelkov, M. Zinke, K. Giller, S. Becker, A. Lange, *Nat. Protoc.* **2017**, *12*, 764–782.

[16] A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, *Nature* **2012**, *486*, 276–9.

[17] F. Castellani, B. Van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, *Nature* **2002**, *420*, 23–26.

[18] A. E. Bennett, C. M. Rienstra, J. M. Griffiths, W. Zhen, P. T. Lansbury, R. G. Griffin, *J. Chem. Phys.* **1998**, *108*, 9463–9479.

[19] M. Huber, S. Hiller, P. Schanda, M. Ernst, A. Böckmann, R. Verel, B. H. Meier, *ChemPhysChem* **2011**, *12*, 915–918.

[20] M. Huber, A. Böckmann, S. Hiller, B. H. Meier, *Phys. Chem. Chem. Phys.* **2012**, *14*, 5239.

[21] R. Linser, B. Bardiaux, V. Higman, U. Fink, B. Reif, *J. Am. Chem. Soc.* **2011**, *133*, 5905–5912.

[22] M. Jain, D. Lalli, J. Stanek, C. Gowda, S. Prakash, S. Tom, *J. Phys. Chem. Lett.* **2017**, *8*, 2399–2405.

[23] M. Etzkorn, A. Böckmann, A. Lange, M. Baldus, *J. Am. Chem. Soc.* **2004**, *136*, 14746–14751.

[24] A. J. Nieuwkoop, C. M. Rienstra, *J. Am. Chem. Soc.* **2010**, *132*, 7570–7571.

[25] A. Lange, S. Luca, M. Baldus, *J. Am. Chem. Soc.* **2002**, *124*, 9704–9705.

[26] H. K. Fasshuber, J.-P. Demers, V. Chevelkov, K. Giller, S. Becker, A. Lange, *J. Magn. Reson.* **2015**, *252*, 10–19.

[27] R. Sprangers, A. Velyvis, L. E. Kay, *Nat. Methods* **2007**, *4*, 697–703.

[28] R. Kerfah, M. J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122.

[29] V. Kurauskas, E. Crublet, P. Macek, R. Kerfah, D. F. Gauto, J. Boisbouvier, P. Schanda, *Chem. Commun.* **2016**, *52*, 9558–9561.

[30] S. Wang, S. Parthasarathy, Y. Nishiyama, Y. Endo, T. Nemoto, K. Yamauchi, T. Asakura, M. Takeda, T. Terauchi, M. Kainosho, *PLoS One* **2015**, *10*, 1–12.

[31] C. Langlois, S. Ramborarina, A. Cukkemane, I. Auzat, B. Chagot, B. Gilquin, A. Ignatiou, I. Petitpas, E. Kasotakis, M. Paternostre, *J. Biol. Chem.* **2014**, *290*, jbc.M114.613166.

[32] K. H. Gardner, L. E. Kay, *J. Am. Chem. Soc.* **1997**, *119*, 7599–7600.

[33] R. Verel, M. Ernst, B. H. Meier, *J. Magn. Reson.* **2001**, *150*, 81–99.

[34] J. Anglister, S. Grzesiek, H. Ren, C. B. Klee, A. Bax, *J. Biomol. NMR* **1993**, *3*, 121–126.

[35] M. Zinke, P. Fricke, C. Samson, S. Hwang, J. S. Wall, S. Lange, S. Zinn-Justin, A. Lange, *Angew. Chem. Int. Ed.* **2017**, 1–6.

[36] N. M. I. Taylor, N. S. Prokhorov, R. C. Guerrero-Ferreira, M. M. Shneider, C. Browning, K. N. Goldie, H. Stahlberg, P. G. Leiman, *Nature* **2016**, *533*, 346–352.