VIP Solid-State NMR Spectroscopy Very Important Paper

DOI: 10.1002/anie.201405730

De Novo 3D Structure Determination from Sub-milligram Protein Samples by Solid-State 100 kHz MAS NMR Spectroscopy**

Vipin Agarwal, Susanne Penzel, Kathrin Szekely, Riccardo Cadalbert, Emilie Testori, Andres Oss, Jaan Past, Ago Samoson, Matthias Ernst,* Anja Böckmann,* and Beat H. Meier**

Abstract: *Solid-state NMR spectroscopy is an emerging tool for structural studies of crystalline, membrane-associated, sedimented, and fibrillar proteins. A major limitation for many studies is still the large amount of sample needed for the experiments, typically several isotopically labeled samples of 10–20 mg each. Here we show that a new NMR probe, pushing magic-angle sample rotation to frequencies around 100 kHz, makes it possible to narrow the proton resonance lines sufficiently to provide the necessary sensitivity and spectral resolution for efficient and sensitive proton detection. Using restraints from such spectra, a well-defined de novo structure of the model protein ubiquitin was obtained from two samples of roughly 500 μg protein each. This proof of principle opens new avenues for structural studies of proteins available in microgram, or tens of nanomoles, quantities that are, for example, typically achieved for eukaryotic membrane proteins by in-cell or cell-free expression.*

Reducing the sample amount for structural studies of proteins by solid-state NMR spectroscopy[1] by a factor of 40, which is what we describe here, in a standard setup with a 3.2 mm MAS rotor that contains 20 mg of protein would reduce the signal-to-noise ratio by the same factor. Alternatively, by linearly scaling down the size of the coil and sample container, the reduction approach reduces the sensitivity by a factor of $40^{2/3} \approx 10$,[2] still making experiments unacceptably long. Using ^{1}H instead of ^{13}C detection is an appealing way to improve the sensitivity of the experiment, taking advantage of the four times higher gyromagnetic ratio of the proton. The corresponding gain in sensitivity of about an order of magnitude ($4^{3/2} = 8$) is sufficient to offset the reduction in sample amount, but it can only be realized if the proton line is as narrow as the ^{13}C line. This presents a challenge since the strong proton–proton dipolar interactions must be adequately suppressed by the experimental scheme. A conceptually simple approach is to increase magic-angle sample spinning (MAS)[3] frequencies, because, in the relevant spinning regime, the observed proton linewidth decreases linearly with the spinning frequency. Since in the past, technically feasible frequencies did not allow sufficient averaging, previous approaches often combined MAS with isotopic dilution,[4] which, however, results in undesired loss of sensitivity. At MAS frequencies of 50 kHz and above, proton detection of deuterated and fully backprotonated proteins (meaning amide and exchangeable side-chain protons are protonated) becomes feasible.[3b,5] Spinning frequencies of up to 60 kHz can be reached with 1.3 mm rotors, which contain roughly 2.0 mg protein, that is, four times more than the amount in the the 0.8 mm rotors used for 100 kHz spinning. The use of the 0.8 mm rotors is thus associated with a sensitivity loss of $4^{2/3} \approx 2.5$. This is almost offset by the proton linewidth which narrows roughly by a factor of 2 (vide infra) when going from 50 to 100 kHz. MAS data at 55 kHz and 60 kHz have been used, together with data from larger rotors or paramagnetic restraints, to derive NMR structures.[4c,6]

We show here that with MAS frequencies of 100 kHz[7] proton lines are narrowed sufficiently to provide the necessary resolution and sensitivity for de novo 3D structure determination of sub-milligram amounts of ubiquitin as a model protein. With protein samples less than 500 μg five experiments for the assignment of protein backbone signals were recorded in about two days, while ^{1}H–^{1}H distance restraints extracted from 3D/4D spectra were recorded in about eight days of measurement time. For structure determination, we rely on the combination of three elements:[8] 1) TALOS-N[9] dihedral-angle backbone restraints based on chemical shifts to define secondary-structure elements; 2) H^N–H^N distance restraints particularly useful to define the β-sheets; 3) and proton distance restraints between methyl groups central for defining the hydrophobic core and providing the global fold.

Figure 1a shows the 0.8 mm diameter rotor used in this study, compared to a standard 3.2 mm rotor. The active volume of the 0.8 mm rotor is roughly 20 times smaller and

[*] Dr. V. Agarwal,[+] S. Penzel,[+] K. Szekely, R. Cadalbert, E. Testori, Prof. M. Ernst, Prof. B. H. Meier
Laboratory of Physical Chemistry, ETH Zurich
8093 Zurich (Switzerland)
E-mail: maer@ethz.ch
beme@ethz.ch

A. Oss, J. Past, Prof. A. Samoson
NMR Instituut, Tartu Teaduspark, Tehnomeedikum
Tallinn University of Technology
Akadeemia tee 15a, 19086, Tallinn (Estonia)
E-mail: ago.samoson@ttu.ee

Dr. A. Böckmann
IBCP, UMR 5086 CNRS/Université de Lyon 1
7 Passage du Vercors, 69367 Lyon (France)
E-mail: a.bockmann@ibcp.fr

[+] These authors contributed equally to this work.

[**] We acknowledge financial support from the Swiss National Science Foundation (Grant 200020_146757), the CNRS, ETH Zurich, the Estonian Science Foundation, and ETAG. We also acknowledge support from the European Commission under the Seventh Framework Programme (FP7; contract Bio-NMR 261863). MAS = magic-angle spinning.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201405730.

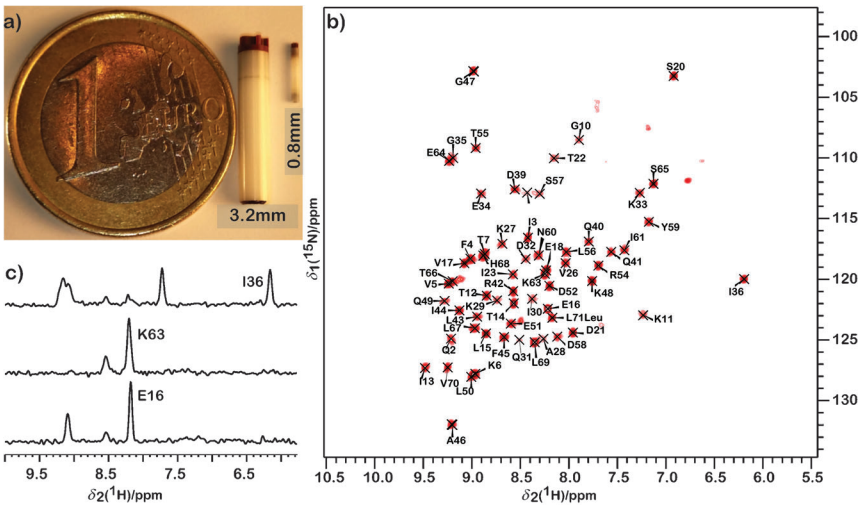

Figure 1. a) Comparison of a 0.8 mm rotor to a 3.2 mm rotor, which can be considered the standard rotor for structural studies. b) ^{1}H-detected 2D ^{1}H–^{15}N spectra of a 100% back-exchanged U-(^{2}H,^{13}C,^{15}N)-ubiquitin sample at a MAS frequency of 99.2 kHz recorded in about 10 min using less than 0.5 mg of protein sample. c) Three typical row traces from the 2D experiment to illustrate the signal-to-noise ratio in the spectrum.

accommodates 0.7 μL of sample. Taking into account the amount of water in such samples, this corresponds to less than 500 μg of protein. Figure 1b shows the proton-detected 2D ^{1}H,^{15}N HSQC spectrum of ubiquitin recorded at 99 kHz MAS within 9 min. The repetition rate was limited by the T_1 of the protons and could be further reduced by paramagnetic doping of the sample.[10] The proton linewidths in the spectrum are in the range from 19 to 73 Hz for the different resonances, with a mean of 41 Hz, while ^{15}N linewidths are 30–68 Hz with a mean of 39 Hz (all values measured without apodization, see Table S1). The average signal-to-noise (S/N) ratio of the peaks in Figure 1b is 31 ± 11. Even around MAS frequencies of 100 kHz, the ^{1}H linewidth is still found to be inversely proportional to the MAS frequency and would, therefore, benefit from spinning at even higher frequencies.[11] Between 60 kHz and 93 kHz MAS frequency, the average experimental proton T_2' was found to increase from 8.4 ms to 13.5 ms, confirming the linear decrease in homogeneous linewidth. For a ubiquitin sample in solution, typical proton linewidths are 6–9 Hz, while ^{15}N linewidths can vary between 3–15 Hz depending on the protein concentration.[12] Thus, the observed linewidths in solids are, under these conditions, within a factor of 3–10 of those in solution. For larger proteins, solution spectra become broader whereas in solids no change is expected.

Five 3D experiments were recorded for backbone assignment, namely (H)CONH (H)CANH (H)CA(CO)NH,[3b] all using dipolar polarization-transfer steps, a (H)(CA)CB(CA)-NH[5a] with a combination of dipolar and scalar transfer steps, as well as a J-transfer-based out-and-back (H)(N)(CA)CB(CA)NH,[12] where all polarization-transfer steps, except the HN transfer, were mediated by the J-coupling. A measurement time of 58 hours was used to record the five spectra. The assignment, together with the resulting TALOS-N[9] prediction for the backbone torsion angles are given in Table S2. A total of 13 peaks in the methyl [^{1}H–^{13}C] HSQC spectrum could be assigned based on unique ^{13}C solid-state chemical shifts, while proton resonances obtained under similar conditions in solution were used to help in assigning the other resonances (details in Table S3).[13] No significant changes in resonance frequencies were detected compared to spectra at lower spinning frequency, indicating that 100 kHz spinning has no effect on the protein.

We used the (^{2}H,^{15}N,^{13}C) sample with 100% protonation at exchangeable sites to observe amide–amide contacts. Correlations between amide protons can be resolved in 3D ^{15}N-^{1}H-^{1}H spectra and they provide distance restraints for structure calculation. The H^N–H^N polarization transfer was achieved by means of spin diffusion in the rotating frame (see the Experimental Section in the Supporting Information). A total of 386 correlation signals were automatically picked in the spectrum,[14] from which 199 cross-peaks could be uniquely assigned. Representative strips from the 3D spectrum, along with the assignment, are shown in Figure 2. In Figure 2 the corresponding inter-proton vectors, mainly between amide protons in amino acids involved in β-sheets, are visualized on the ubiquitin X-ray structure (PDB code: 3ONS).[15] A complete list of assigned peaks is provided in the Supporting Information (Table S4).

All 33 ILV methyl correlations expected for ubiquitin are observed in the 2D ^{1}H–^{13}C HSQC spectrum (Figure 3a).

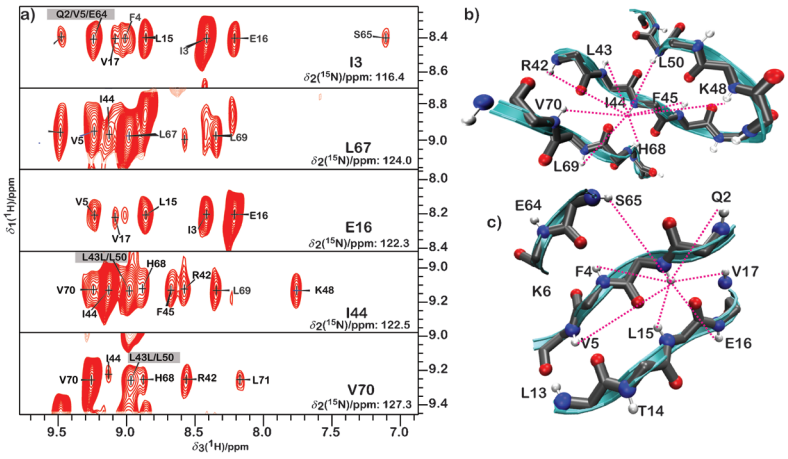

Figure 2. Amide restraints from 3D spectra of backprotonated (^{2}H,^{13}C,^{15}N)-ubiquitin. a) Strip plot from a ^{1}H-detected 3D ^{15}N-resolved rotating-frame [^{1}H,^{1}H] spin-diffusion experiment. The ambiguous crosspeaks are highlighted by gray boxes around the label; all other labeled crosspeaks are spectrally unambiguous. b,c) Map of the correlations observed in the strip plots in (a) to the structural element from the X-ray structure (PDB code: 3ONS).

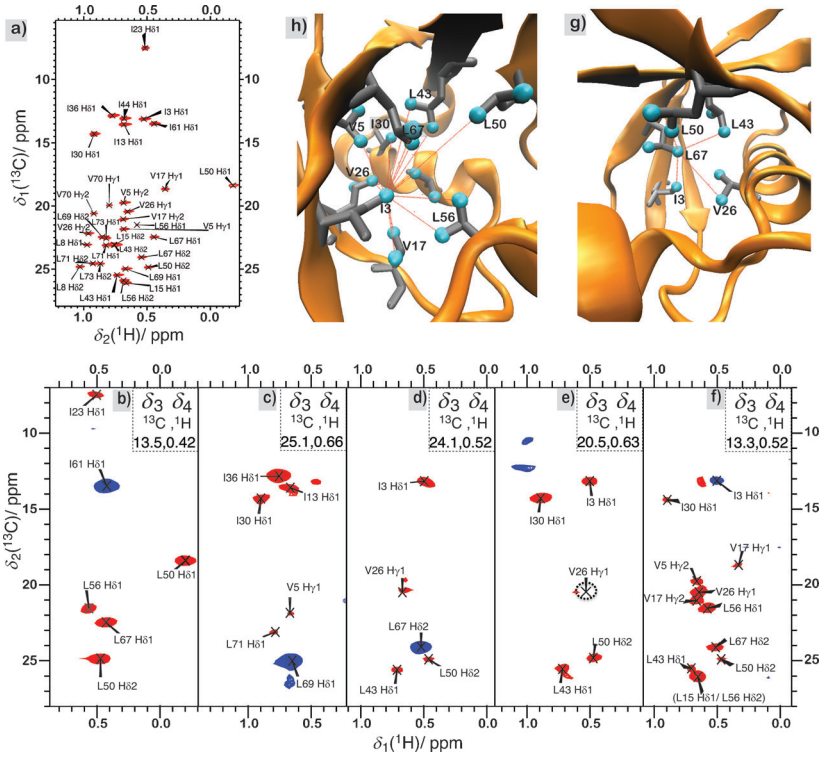

Figure 3. Spectra of ILV-labeled ubiquitin. a) 2D [^{1}H,^{13}C] HSQC spectrum at a MAS frequency of 94.5 kHz and a sample temperature of 20 °C. b–f) 2D planes (f_1–f_2 dimensions) of the nonuniformly sampled 4D HSQC-DREAM-HSQC experiment. Positive peaks (diagonal peaks) are depicted in blue while negative peaks (cross-peaks) are depicted in red. 2D planes of b) I61 Hδ1, c) L69 Hδ1, d) L67 Hδ2, e) V26 Hγ1, and f) I3 Hδ1. g,h) Representation of the methyl–methyl correlations observed in the 4D experiments for the L67δ2 and I3δ1 methyl groups by red lines on the X-ray structure. CHD_2 groups are depicted as cyan spheres.

Figure 3 b–f shows five ^{1}H–^{13}C 2D planes from a 4D HSQC-DREAM-HSQC experiment, which correlates the HSQC spectra of the source and the destination spin, providing the detailed polarization-transfer pathways.[8] The cross-peaks in the 2D planes shown identify the source of the polarization (by its ^{1}H and ^{13}C frequency), which is subsequently transferred to the methyl group corresponding to the signal with blue contours in each panel. In most globular proteins, methyl groups are predominantly located in the hydrophobic core.[16] Cross-peaks in the 4D spectrum correspond almost exclusively to long-range H^{Met}–H^{Met} contacts and provide key distance restraints. Eighty-two cross peaks were identified in the spectrum. Of these, 73 could be spectrally unambiguously assigned, while nine cross peaks have two assignment possibilities. A table with the observed contacts is provided in the Supporting Information (Table S5). As an example, the correlations observed for the L67 Hδ2 and the I3 Hδ1 protons in the 2D planes are shown on the X-ray structure in Figure 3 g,h).

The restraints summarized in Table 1 are sufficient to obtain the well-defined structure shown in Figure 4 a) with a backbone rmsd of (0.6 ± 0.1) Å which is comparable to the reference X-ray structure (3ONS) in Figure 4 b. Details of the structure calculation are given in Table S7. The backbone rmsd between the NMR and X-ray structures is 2.0 Å (using residues 2–71) and predominantly caused by shifts in the relative position of the well-defined helices and sheets and by differences in the semiflexible loop. The stability of the structure calculation against input variation was confirmed by independently varying the upper-distance limit of the methyl–methyl contacts between 4.5 and 7 Å and of the amide–amide contacts between 5 and 7 Å on a 2D grid. Each calculation was also repeated for different initial seeds for the calculation. The rmsd for the bundle of these structures was (0.7 ± 0.2) Å. Despite the smaller amount of sample, the number of long-range H^N–H^N (H^{Met}–H^{Met}) distance restraints obtained at ca. 100 kHz is about a factor of 3.0 (1.5) times larger than those previously obtained at 60 kHz MAS frequency with significantly larger amounts of protein.[4c]

We have demonstrated the feasibility of de novo protein structure determination using less than 500 μg (59 nanomol) of protein sample by solid-state NMR spectroscopy. The comparison in proton transverse relaxation times confirmed the expected improvement in homogeneous linewidth with faster spinning, improving the spectral

Table 1: Number of structural restraints used for the structure calculation.

Type of restraint	H^N–H^N	HMet–HMet
total Distance restraints	130	82
automated[a]/manual	automated	manual
unambiguous	130	73
short range ($\|i-j\| \leq 1$)	44	0
medium range ($2 \leq \|i-j\| < 5$)	50	5
long range ($\|i-j\| > 5$)	36	68
dihedral angle restraints (TALOS-N)	120	

[a] For a detailed list of the automatically assigned peaks see Table S4.

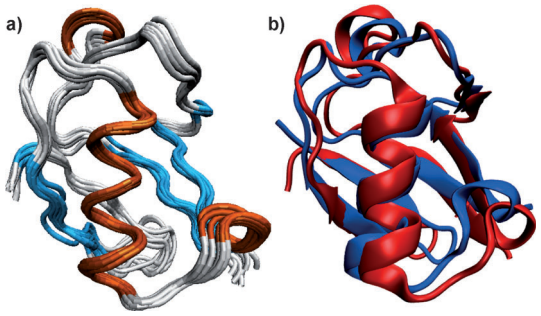

Figure 4. a) Bundle of ten lowest-energy structures determined by NMR experiments, derived from an ensemble of 200 calculated structures. The cyan, orange, and the gray colors depict β-sheets, α-helices, and loops in the protein. b) Overlay of the solid-state NMR structure (red) and X-ray structure (blue, PDB code: 3ONS) of ubiquitin.

resolution and sensitivity: The total measurement time for all spectra was 11 days (2.4 days for assignment, 8.5 days for structure restraints). This leaves room for progressing to larger proteins, protein complexes, or proteins diluted in lipids, as the quality of the spectrum, when defined in terms of linewidth, is expected to be independent of the protein size although it may be dependent on dynamic aspects of the protein investigated. We believe that future technical advances will further improve the linewidth and signal-to-noise ratio. The methodology described here will expand the applicability of solid-state NMR spectroscopy to proteins that are difficult to express in multi-milligram quantities, like many eukaryotic membrane proteins, where amounts often do not surpass milligram amounts per liter (in-cell expression) or milliliter (cell-free expression).

Received: May 30, 2014
Published online: September 15, 2014

Keywords: magic-angle spinning · protein structures · proton detection · solid-state NMR spectroscopy

[1] a) F. Castellani, B. Van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschklnat, *Nature* **2002**, *420*, 98; b) A. Loquet, B. Bardiaux, C. Gardiennet, C. Blanchet, M. Baldus, M. Nilges, T. Malliavin, A. Böckmann, *J. Am. Chem. Soc.* **2008**, *130*, 3579; c) T. Manolikas, T. Herrmann, B. H. Meier, *J. Am. Chem. Soc.* **2008**, *130*, 3959; d) C. Gardiennet, A. K. Schutz, A. Hunkeler, B. Kunert, L. Terradot, A. Bockmann, B. H. Meier, *Angew. Chem.* **2012**, *124*, 7977; *Angew. Chem. Int. Ed.* **2012**, *51*, 7855; e) I. Bertini, C. Luchinat, G. Parigi, E. Ravera, B. Reif, P. Turano, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10396; f) C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, *Science* **2008**, *319*, 1523; g) H. Van Melckebeke, C. Wasmer, A. Lange, E. Ab, A. Loquet, A. Böckmann, B. H. Meier, *J. Am. Chem. Soc.* **2010**, *132*, 13765; h) S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kuhne, J. R. Stout, V. A. Higman, R. E. Klevit, B. J. van Rossum, H. Oschkinat, *Nat. Struct. Mol. Biol.* **2010**, *17*, 1037; i) L. Shi, I. Kawamura, K.-H. Jung, L. S. Brown, V. Ladizhansky, *Angew. Chem.* **2011**, *123*, 1338; *Angew. Chem. Int. Ed.* **2011**, *50*, 1302.

[2] A. Samoson, T. Tuherm, J. Past, A. Reinhold, I. Heinmaa, T. Anupõld, M. E. Smith, K. J. Pike, *eMagRes*, Wiley, Hoboken, **2010**.

[3] a) J. R. Lewandowski, J.-N. Dumez, U. Akbey, S. Lange, L. Emsley, H. Oschkinat, *J. Phys. Chem. Lett.* **2011**, *2*, 2205; b) M. J. Knight, A. L. Webber, A. J. Pell, P. Guerry, E. Barbet-Massin, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, L. Emsley, A. Lesage, T. Herrmann, G. Pintacuda, *Angew. Chem.* **2011**, *123*, 11901; *Angew. Chem. Int. Ed.* **2011**, *50*, 11697; c) D. Zhou, A. Nieuwkoop, D. Berthold, G. Comellas, L. Sperling, M. Tang, G. Shah, E. Brea, L. Lemkau, C. Rienstra, *J. Biomol. NMR* **2012**, *54*, 291.

[4] a) V. Agarwal, A. Diehl, N. Skrynnikov, B. Reif, *J. Am. Chem. Soc.* **2006**, *128*, 12620; b) V. Chevelkov, K. Rehbein, A. Diehl, B. Reif, *Angew. Chem.* **2006**, *118*, 3963; *Angew. Chem. Int. Ed.* **2006**, *45*, 3878; c) M. Huber, S. Hiller, P. Schanda, M. Ernst, A. Böckmann, R. Verel, B. H. Meier, *ChemPhysChem* **2011**, *12*, 915; d) S. Asami, K. Szekely, P. Schanda, B. H. Meier, *J. Biomol. NMR* **2012**, *54*, 155.

[5] a) E. Barbet-Massin, A. Pell, K. Jaudzems, W. T. Franks, J. Retel, S. Kotelovica, I. Akopjana, K. Tars, L. Emsley, H. Oschkinat, A. Lesage, G. Pintacuda, *J. Biomol. NMR* **2013**, *56*, 379; b) A. Marchetti, S. Jehle, M. Felletti, M. J. Knight, Y. Wang, Z.-Q. Xu, A. Y. Park, G. Otting, A. Lesage, L. Emsley, N. E. Dixon, G. Pintacuda, *Angew. Chem.* **2012**, *124*, 10914; *Angew. Chem. Int. Ed.* **2012**, *51*, 10756.

[6] M. J. Knight, A. J. Pell, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, T. Herrmann, L. Emsley, G. Pintacuda, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11095 – 11100.

[7] a) V. Agarwal, T. Tuherm, A. Reinhold, J. Past, A. Samoson, M. Ernst, B. H. Meier, *Chem. Phys. Lett.* **2013**, *583*, 1; b) Y. Q. Ye, M. Malon, C. Martineau, F. Taulelle, Y. Nishiyama, *J. Magn. Reson.* **2014**, *239C*, 75.

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

[9] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56*, 227.

[10] N. P. Wickramasinghe, M. Kotecha, A. Samoson, J. Past, Y. Ishii, *J. Magn. Reson.* **2007**, *184*, 350.

[11] B. H. Meier, A. Samoson, A. Böckmann, M. Ernst, V. Agarwal, M. Huber, S. Penzel, F. Ravotti in *54th Experimental Nuclear Magnetic Resonance conference (ENC)* (Ed.: M. Garwood), Asilamor, CA, **2013**.

[12] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy*, Elsevier Academic Press, Amsterdam, **2007**.

[13] a) T. I. Igumenova, A. E. McDermott, K. W. Zilm, R. W. Martin, E. K. Paulson, A. J. Wand, *J. Am. Chem. Soc.* **2004**, *126*, 6720; b) A. J. Wand, J. L. Urbauer, R. P. McEvoy, R. J. Bieber, *Biochemistry* **1996**, *35*, 6116.

[14] T. J. Stevens, R. H. Fogh, W. Boucher, V. A. Higman, F. Eisenmenger, B. Bardiaux, B. J. Van Rossum, H. Oschkinat, E. D. Laue, *J. Biomol. NMR* **2011**, *51*, 437.

[15] K. Y. Huang, G. A. Amodeo, L. Tong, A. McDermott, *Protein Sci.* **2011**, *20*, 630.

[16] a) V. Tugarinov, L. E. Kay, *Chembiochem* **2005**, *6*, 1567; b) A. M. Ruschak, L. E. Kay, *J. Biomol. NMR* **2010**, *46*, 75.

