Structure Elucidation

DOI: 10.1002/ange.200501884

SOLARIA: A Protocol for Automated Cross-Peak Assignment and Structure Calculation for Solid-State Magic-Angle Spinning NMR Spectroscopy

Michele Fossi, Federica Castellani, Michael Nilges, Hartmut Oschkinat, and Barth-Jan van Rossum**

Recently, it has been shown that structures of proteins can be determined by magic-angle-spinning (MAS) solid-state NMR spectroscopy.[1,2] Central to most structure-determination procedures is the collection of a set of distance restraints that is sufficiently large to achieve convergence in the calculations. In our previous solid-state NMR spectroscopic structure investigation on a microcrystalline preparation of the α-spectrin SH3 domain these distance restraints were obtained by manual assignment of cross-peaks from 2D and 3D proton-driven spin diffusion (PDSD) correlation experiments.[1,3] Extensively ^{13}C-labeled preparations obtained by protein expression on a medium containing either [1,3-^{13}C]glycerol or [2-^{13}C]glycerol as the carbon source were used for all the experiments. In such preparations, only a few amino acid types have ^{13}C labels in adjacent positions, which led to a substantial suppression of dipolar attenuation effects[4,5] and enabled a straightforward detection of long-range ^{13}C-^{13}C correlations (Figure 1). The spectra used in these studies displayed a high level of resonance overlap, which resulted in many assignment options per signal and led to the exclusion of a large fraction of ambiguous cross-peaks that could not be assigned manually. Automation of the cross-peak assignment would be beneficial to avoid this problem and to speed-up the structure-determination process.

In liquid-state NMR spectroscopy, software packages such as ARIA,[6–10] CANDID,[11] DYANA,[12] KNOWNOE,[13] NOAH,[14,15] and AUTOSTRUCTURE[16,17] have provided a way to handle ambiguous cross-peaks to derive distance restraints in an automated fashion. In ARIA, ambiguous NOE interactions are handled by treating them as "ambiguous distance restraints" (ADR).[18,19] The calculated structures are then refined in an iterative manner, each time using the structures of the latest round of calculations to find new assignments as the input for the next round.

We developed SOLARIA,[+] a MAS NMR-dedicated version of ARIA for solid-state NMR spectroscopic studies.

[*] M. Fossi, Dr. F. Castellani, Prof. Dr. H. Oschkinat, Dr. B.-J. van Rossum
Forschungsinstitut für Molekulare Pharmakologie (FMP)
Robert-Rössle-Strasse 10, Berlin (Germany)
Fax: (+49) 3094793169
E-mail: Oschkinat@fmp-berlin.de
brossum@fmp-berlin.de
Prof. Dr. M. Nilges
Institut Pasteur, 25–28 Rue du Docteur Roux
75015 Paris (France)

[+] Available from www.fmp-berlin.de.

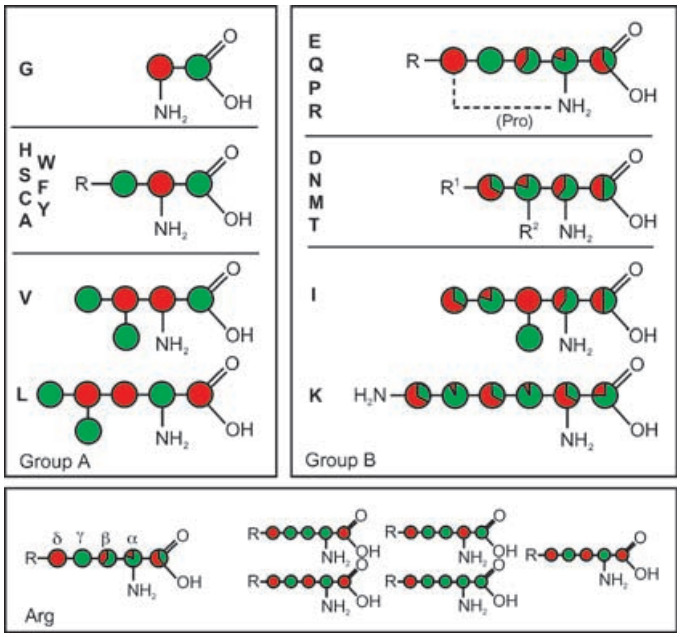

Figure 1. Labeling patterns for the different amino acids in the α-spectrin SH3 domain preparations used in these studies. Schematic representation of the effective ^{13}C enrichment for the different residues, as obtained by protein expression in *E. coli* BL21 (DE3). The green color corresponds to the degree of ^{13}C labeling obtained by growth of the bacteria on [1,3-^{13}C]-glycerol; the opposite labeling pattern, obtained by growth on [2-^{13}C]glycerol, is represented in red. There are two groups of amino acids (A and B). In group A, the various carbon sites are either approximately 100% or 0% labeled, in group B the residues show fractional labeling. This fractional labeling is the result of the production of isotopomers of residues with different labeling patterns, as illustrated for arginine at the bottom. These percentages were estimated from extensive solution NMR studies using 2D and 3D heteronuclear NMR techniques.

The two programs share the same overall architecture consisting of nine iterations of cross-peak assignment and structure calculations. However, there are also significant differences. First, SOLARIA can handle peak lists containing not only ^{1}H-^{1}H, but also typical solid-state ^{13}C-^{13}C, ^{15}N-^{15}N, and ^{13}C-^{15}N correlations. Second, in contrast to solution NMR spectroscopy, it is very difficult to find a satisfying relationship between experimental volumes and distances in solid-state NMR spectroscopy. Peak volumes do not depend exclusively on the distance; they are strongly affected by mobility, which interferes with dipolar couplings. This situation can significantly alter the dipolar polarization transfer efficiency and often makes it difficult to achieve a uniform excitation of all the relevant spins. The proton environment may also influence the efficiency of mixing. Further complications may arise through offset-dependent transfer processes, interference with heteronuclear decoupling schemes, and sample heterogeneity. Finally, even with the reduced labeling, dipolar attenuation effects are not fully suppressed and the transfer efficiency between two spins can still be largely affected by coupling to other nearby spin systems. For these reasons, the program does not use volumes to set the boundaries for distance restraints, as done in ARIA. Therefore, all distance constraints are represented by the same lower boundary and the same generous upper boundary, and input peak lists do not require the presence of cross-peak volumes or intensities. Third, the labeling pattern in samples made by using 2- or 1,3-labeled glycerol is exploited by SOLARIA for better convergence of the automated assignment process. This is achieved by removal of all assignment options which are not allowed according to the labeling pattern. In principle, other labeling patterns can be easily implemented, for example, for proteins obtained from growth media containing selectively ^{13}C-labeled succinic acid as a precursor.[20]

We tested SOLARIA on lists with the coordinates of manually picked peaks from PDSD-type spectra of the α-spectrin SH3 domain. Intermolecular cross-peaks were identified previously on an experimental basis[1] and removed from the lists. To face the computational difficulties inherent to the use of highly ambiguous solid-state peak lists (the average number of assignment candidates per signal was approximately 6 for the two 3D spectra, and 16 and 19 for the 2-CC and 1,3-CC 2D spectra, respectively) and very loose boundaries (all constraints are restrained between 2.5 and 6.5 Å, independent of the peak volume), we decided to enhance the convergence capability of the software by substantially increasing the number of cooling steps.[21] We raised the number from 9000, which is the commonly used value, to 100000. Under these conditions and using standard ARIA1.2 input values for all other parameters SOLARIA produced convergent results. The 11 lowest-energy structures (blue) overlaid with the X-ray structure (red) as reference are shown in Figure 2. Despite the use of extremely generous boundaries for the distance restraints, well-defined (0.73-Å ensemble backbone rmsd) and accurate (1.3-Å backbone rmsd to the X-ray reference) structures could be obtained.

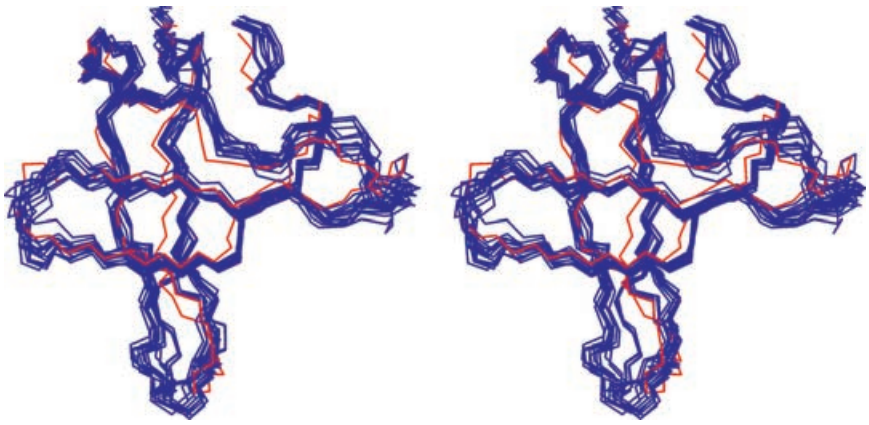

Figure 2. Stereoview of the 11 lowest-energy solid-state NMR structures of the α-spectrin SH3 domain calculated by SOLARIA (blue). For comparison, the X-ray reference structure is included (PDB entry: 1SHG; red) and overlaid with the family of solid-state structures by fitting the backbone atoms to the average solid-state structure. The calculations were performed on peak lists where intermolecular cross-peaks were manually removed.

These structures were calculated in approximately 12 h, which is in striking contrast to the several months required in our previous work for the manual assignment of the same spectra.[1] More importantly, SOLARIA allowed the assignment of approximately 20% more cross-peaks than in the previous manual assignment procedure (Figure 3). All

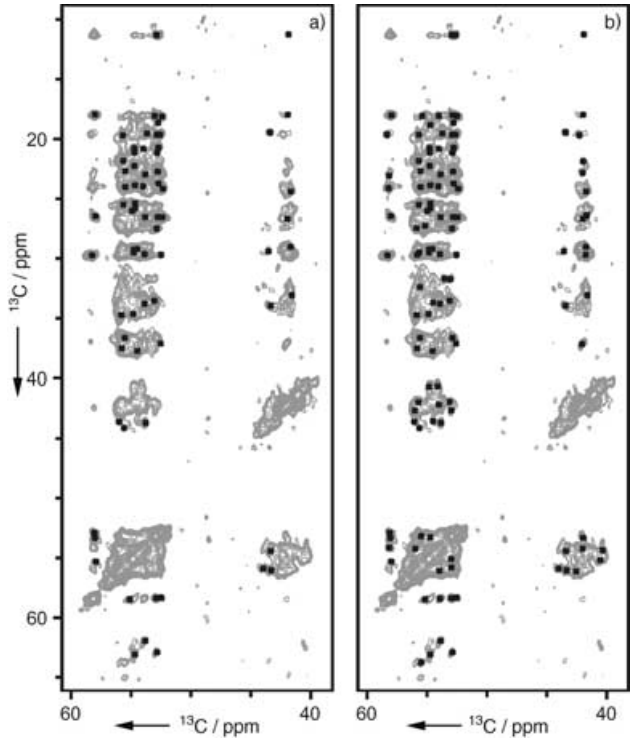

Figure 3. Visualization of the assignments. A strip of a two-dimensional PDSD spectrum recorded using [1,3-^{13}C]glycerol-grown α-spectrin SH3 domain sample is shown.[1] In (a), the cross-peaks that could be assigned manually using 2D and 3D data are indicated. In (b), cross-peaks assigned and used for the structure calculations by SOLARIA using the same dataset are indicated.

manual assignments obtained from the evaluation of 2D and 3D spectra are indicated in Figure 3a, displayed in a region of a 2D experiment, and all automatically assigned peaks are indicated in Figure 3b.

In summary, the program SOLARIA produced, in a few hours, accurate structures (1.3 Å rmsd to the X-ray reference) using unassigned peak lists from 2D and 3D MAS-NMR spectra of the α-spectrin SH3 domain, where intermolecular cross-peaks were discarded prior to the calculation. These structures exemplify a first successful attempt to introduce automation to structure determination of solid proteins by MAS NMR spectroscopy. The automation of the cross-peak assignment resulted in a dramatic speed-up of the whole procedure and, most importantly, provided a way to handle ambiguous cross-peaks in MAS NMR spectra. Hence, we expect that this work will open new possibilities for the use of MAS NMR to determine the structures of larger systems, such as membrane proteins, where the number of cross-peaks may become unmanageably large for manual assignment methods. For these reasons, we believe that SOLARIA represents an important step forward for the rapidly developing MAS NMR technique.

Experimental Section

Sample preparation and solid-state MAS NMR spectroscopy: The main characteristics of the SH3 domain and the spectra used in this work are summarized in Table 1. The sample preparation is described in detail elsewhere.[22] Peak lists for the structure calculations were generated from solid-state 2D ^{13}C-^{13}C and 3D ^{15}N-^{13}C-^{13}C correlation spectra.[1,3]

Table 1: Spectra used in the calculation.[a]

Spectrum	Dimensionality	Labeling	Number of peaks
1,3-CC	2D	1,3-glycerol	566
2-CC	2D	2-glycerol	461
1,3-NCOCX	3D	1,3-glycerol	477
2-NCACX	3D	2-glycerol	377

[a] Peaks were picked manually. The numbers in parenthesis refer to the total number of peak list entries after the manual removal of intermolecular cross-peaks. No manual assignment was included in the lists.

Structure calculations: The software SOLARIA was used to perform the automatic assignment of manually picked cross-peaks and to generate structures. The calculations were performed on a SGI Origin 2100 cluster at the FMP, Berlin. The chemical shift tolerances were set to 0.4 ppm for carbon and 0.5 ppm for nitrogen dimentions. The number of calculated structures was 20 for the first eight iterations and 100 for the final one. The results of the calculations were evaluated by computing a pairwise backbone rmsd (precision) and a backbone rmsd (accuracy) to the X-ray reference-structure[23] (PDB entry: 1SHG) of the 15 lowest-energy structures.

Received: May 31, 2005

Keywords: computer chemistry · NMR spectroscopy · proteins · solid-state structures · structure elucidation

[1] F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, *Nature* **2002**, *420*, 98 – 102.
[2] A. E. McDermott, *Curr. Opin. Struct. Biol.* **2004**, *14*, 554 – 561.
[3] F. Castellani, B. J. van Rossum, A. Diehl, K. Rehbein, H. Oschkinat, *Biochemistry* **2003**, *42*, 11476 – 11483.
[4] P. Hodgkinson, L. Emsley, *J. Magn. Reson.* **1999**, *139*, 46 – 59.
[5] S. Kiihne, M. A. Mehta, J. A. Stringer, D. M. Gregory, J. C. Shiels, G. P. Drobny, *J. Phys. Chem. A* **1998**, *102*, 2274 – 2282.
[6] J. P. Linge, S. I. O'Donoghue, M. Nilges, *Nucl. Magn. Reson. Biol. Macromol. Part B* **2001**, *339*, 71 – 90.
[7] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, *Bioinformatics* **2003**, *19*, 315 – 316.
[8] M. Nilges, M. J. Macias, S. I. O'Donoghue, H. Oschkinat, *J. Mol. Biol.* **1997**, *269*, 408 – 422.
[9] M. Nilges, S. I. O'Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 107 – 139.
[10] A. Kharrat, M. J. Macias, T. J. Gibson, M. Nilges, A. Pastore, *EMBO J.* **1995**, *14*, 3572 – 3584.
[11] T. Herrmann, P. Guntert, K. Wuthrich, *J. Mol. Biol.* **2002**, *319*, 209 – 227.
[12] P. Guntert, C. Mumenthaler, K. Wuthrich, *J. Mol. Biol.* **1997**, *273*, 283 – 298.

[13] W. Gronwald, S. Moussa, R. Elsner, A. Jung, B. Ganslmeier, J. Trenner, W. Kremer, K. P. Neidig, H. R. Kalbitzer, *J. Biomol. NMR* **2002**, *23*, 271 – 287.
[14] C. Mumenthaler, W. Braun, *J. Mol. Biol.* **1995**, *254*, 465 – 480.
[15] C. Mumenthaler, P. Guntert, W. Braun, K. Wuthrich, *J. Biomol. NMR* **1997**, *10*, 351 – 362.
[16] H. N. Moseley, D. Monleon, G. T. Montelione, *Methods Enzymol.* **2001**, *339*, 91 – 108.
[17] H. N. Moseley, G. T. Montelione, *Curr. Opin. Struct. Biol.* **1999**, *9*, 635 – 642.
[18] M. Nilges, *J. Mol. Biol.* **1995**, *245*, 645 – 660.
[19] M. Nilges, *Proteins Struct. Funct. Genet.* **1993**, *17*, 297 – 309.
[20] A. J. van Gammeren, F. B. Hulsbergen, J. G. Hollander, H. J. de Groot, *J. Biomol. NMR* **2004**, *30*, 267 – 274.
[21] M. Fossi, H. Oschkinat, M. Nilges, L. J. Ball, *J. Magn. Reson.* **2005**, *175*, 92 – 102.
[22] J. Pauli, B. van Rossum, H. Forster, H. J. de Groot, H. Oschkinat, *J. Magn. Reson.* **2000**, *143*, 411 – 416.
[23] A. Musacchio, M. Noble, R. Pauptit, R. Wierenga, M. Saraste, *Nature* **1992**, *359*, 851 – 855.

