

Rapid prediction of multi-dimensional NMR data sets

Sabine Gradmann · Christian Ader · Ines Heinrich · Deepak Nand ·
Marc Dittmann · Abhishek Cukkemane · Marc van Dijk · Alexandre M. J. J. Bonvin ·
Martin Engelhard · Marc Baldus

Received: 12 July 2012 / Accepted: 31 October 2012 / Published online: 10 November 2012
© Springer Science+Business Media Dordrecht 2012

Abstract We present a computational environment for Fast Analysis of multidimensional NMR Data Sets (FANDAS) that allows assembling multidimensional data sets from a variety of input parameters and facilitates comparing and modifying such “in silico” data sets during the various stages of the NMR data analysis. The input parameters can vary from (partial) NMR assignments directly obtained from experiments to values retrieved from in silico prediction programs. The resulting predicted data sets enable a rapid evaluation of sample labeling in light of spectral resolution and structural content, using standard NMR software such as Sparky. In addition, direct comparison to experimental data sets can be used to validate NMR assignments, distinguish different molecular components, refine structural models or other parameters derived from NMR data. The method is demonstrated in the context of solid-state NMR data obtained for the cyclic nucleotide binding domain of a bacterial cyclic nucleotide-gated channel and on membrane-embedded sensory rhodopsin II. FANDAS is freely available as web portal under WeNMR (<http://www.wenmr.eu/services/FANDAS>).

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9681-y) contains supplementary material, which is available to authorized users.

S. Gradmann · C. Ader · D. Nand · A. Cukkemane ·
M. van Dijk · A. M. J. J. Bonvin · M. Baldus (✉)
Faculty of Science, Bijvoet Center for Biomolecular Research,
Utrecht University, Padualaan 8, 3584 CH Utrecht,
The Netherlands
e-mail: m.baldus@uu.nl

I. Heinrich · M. Dittmann · M. Engelhard
Department of Physical Biochemistry, Max Planck Institute
for Molecular Physiology, Otto-Hahn-Strasse 11,
44227 Dortmund, Germany

Keywords NMR · Software · Chemical shift · Membrane · Protein · Solid-state NMR

Introduction

NMR spectroscopy has become increasingly powerful to provide insight into the structural or dynamical organization of large biomolecules that cannot be obtained using other structural biology methods (see, e.g., Lange et al. 2006; Gelis et al. 2007; Ader et al. 2010; Jehle et al. 2011; Kato et al. 2011; Renault et al. 2011; Loquet et al. 2012). Complexity in the resulting NMR spectra can not only result from the sheer molecular size giving rise to spectral crowding but it can also be related to the details of the experimental setup. For example, spinning sidebands are frequently observed in ^1H or ^{13}C edited solid-state NMR (ssNMR) experiments. Furthermore, the occurrence of polymorphism (see, e.g., Heise et al. 2005; Seidel et al. 2005) or the presence of several molecular components, such as in the case of cellular solid-state NMR studies (Renault et al. 2012a, b) can complicate the spectral analysis.

To deal with such increasing levels of spectral complexity, tailored NMR approaches have been developed in the last years that facilitate different aspects of an NMR-based analysis either in solution or in the solid state (see, e.g., Renault et al. 2010; Ruschak and Kay 2010 for recent reviews). These methods often employ a combination of advanced sample preparation routines including specific labeling and spectroscopic filtering methods to simplify the resulting NMR spectra. In addition, significant progress has been made to support and automate the NMR data analysis process. Initially such approaches were developed for solution-state NMR applications (see, e.g., Nelson et al.

1991; Zimmerman et al. 1997; Linge et al. 2003; Goddard and Kneller 2006; Guerry and Herrmann 2011). More recently, the underlying concepts have been tailored for ssNMR studies (Fossi et al. 2005; Loquet et al. 2008; Manolikas et al. 2008; Hu et al. 2011; Stevens et al. 2011). Dedicated algorithms have also been proposed to aid other aspects of NMR studies, for example in the context of sample preparation (Hefke et al. 2011) or homology modeling (Brothers et al. 2011).

Next to NMR approaches, there is also clear trend in combining information from various experimental sources and computations in order to tackle increasingly challenging systems. Methods such as cryo-electron microscopy (cryo-EM) or small angle X-ray scattering (SAXS) are generating a wealth of lower resolution structural information that can be used to complement more classical structural studies by NMR and X-ray crystallography (see, e.g., Jehle et al. 2011). All those data are being integrated with computational approaches for the modeling of large and complex systems (Melquiond et al. 2012). Examples of such integrative approaches are HADDOCK (Dominguez et al. 2003; de Vries et al. 2010), an information-driven docking program that can incorporate various sources of experimental data and bioinformatics prediction to model macromolecular assemblies, and the Integrative Modeling Platform (IMP) which was designed for the modeling of macromolecular assemblies (Alber et al. 2008).

With these aspects in mind, the ability to directly relate structural information gathered by other biophysical methods to NMR parameters opens up new avenues to speed up different steps of an NMR-based analysis of molecular structure and motion. In such studies, a rapid comparison between reference data and NMR results during data analysis is often highly desirable. Here, we introduce the computational environment FANDAS (Fast Analysis of multidimensional NMR data sets) that allows assembling multidimensional data sets from a variety of input data and to compare and modify such “in silico” spectra during various stages of the NMR data analysis. We first discuss the general modules of FANDAS. Subsequently, we illustrate the use of FANDAS for the analysis of ssNMR data sets obtained on a cyclic nucleotide binding domain (CNBD) of a bacterial cyclic nucleotide-gated channel (Cukkemane et al. 2012) and on membrane-embedded variants of sensory rhodopsin II (SRII). While the focus here is on solid-state NMR data, the underlying methodology can also be adapted for solution-state NMR studies. FANDAS is freely available as a web portal under WeNMR (<http://www.wenmr.eu/services/FANDAS>).

Materials and methods

Preparation of ^{13}C labelled lipids from *Natronomonas pharaonis*

Natronomonas pharaonis were grown in a minimal medium (MM9) supplemented with 3.4 M NaCl, pH 9.2 and ^{13}C -Na-acetate at 40 °C while illuminating with orange light. After harvesting, the cells were washed with a buffer containing 250 g/L NaCl, 20 g/L $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$, 3 mg/L sodium citrate and 2 g/L KCl. For lysis, 3 mg DNase A per L culture were added and the cell pellet was dialysed against double distilled H_2O (dd H_2O) over night. The lysate was sedimented at $170,000\times g$, washed three times with dd H_2O and lyophilised. Lipids were extracted by treatment with $\text{CHCl}_3/\text{MeOH}$ (1:1) under an argon atmosphere over night. The solvent was removed under reduced pressure; the extract was dissolved in $\text{CHCl}_3/\text{MeOH}$ (1:1) and applied on a silica gel column equilibrated with diethyl ether/petroleum ether (85:15). The column was washed with diethyl ether/petroleum ether (85:15) and with $\text{CHCl}_3/\text{EtOH}$ (2:1) containing 1 % (v/v) dd H_2O and 1 % (v/v) formic acid. $\text{CHCl}_3/\text{EtOH}$ (1:1) with 10 % (v/v) dd H_2O and 10 % (v/v) formic acid was used to elute polar lipids. Lipid containing fractions were collected and the solvent was evaporated. In order to remove protein impurities, $\text{CHCl}_3/\text{MeOH}$ (1:1) was added and the solution was centrifuged at $5,000\times g$, 10 min. This procedure was repeated until no protein impurities could be detected in a SDS-PAGE analysis. Freeze-dried ^{13}C labeled *N. pharaonis* lipids were resuspended in 50 mM NaP_i , pH 8 and after sonication stored at -80°C . ^{13}C enrichment 99 % as determined by ESI mass spectrometry.

Protein expression

Uniformly ^{13}C , ^{15}N labelled SRII was expressed in *Escherichia coli* and purified as described before (Hohenfeld et al. 1999; Etzkorn et al. 2007). Expression of SRII using *E. coli* cells grown in a minimal medium supplemented with 1,3- ^{13}C -glycerol was performed using the protocol of (LeMaster and Kushlan 1996). In both cases the yield was 3 mg/L medium. The N-terminally truncated transducer HtrII 157 was expressed in *E. coli* and purified as described in (Wegener et al. 2001). The proteins were reconstituted into unlabeled or ^{13}C labeled lipids from *N. pharaonis* according to (Klare et al. 2006). For the preparation of the CNBD domain, the reader is referred to (Cukkemane et al. 2012). The VLFY-reverse labeled sample was prepared as described in (Etzkorn et al. 2007).

Experimental setup

NMR experiments were performed using Bruker Avance III spectrometers equipped with 3.2 mm triple resonance (^1H , ^{13}C , ^{15}N) MAS probeheads at 700 MHz (Figs. 3, 4, 5), 500 MHz (Fig. 6) and 800 MHz (Fig. 7b) ^1H resonance frequency. The ^{13}C – ^{13}C spin diffusion spectrum with a 20 ms mixing time on SRII was acquired at 4 °C effective temperature and an MAS rate of 10.92 kHz. The 3D NCOX experiment conducted on CNBD employed SPECIFIC-CP conditions (Baldus et al. 1998) for sequential NCO transfer using radio frequency fields of 25 and 42.5 kHz on ^{15}N and ^{13}C channels, respectively. The phase-switched MIRROR (Scholz et al. 2008) scheme was applied for homonuclear (^{13}C , ^{13}C) mixing for a duration of 50 ms. The 3D data were acquired at 4 °C and a MAS rate of 17 kHz. The NMR measurements on the SRII/HtrII complex (Fig. 6) were performed at an effective temperature of 5 °C and an MAS rate of 7.5 kHz. The 2D NCOX experiment was acquired with a (^{13}C , ^{13}C) mixing time of 40 ms using proton-driven spin diffusion.

Results and discussion

FANDAS modules and program use

Data creation within FANDAS generally proceeds using the steps depicted in Fig. 1. The input data are generated in Step 1 of the program, based on the input amino acid sequence and secondary structure information if available. This procedure involves the creation of 31-tuples for each residue i of a given protein sequence that encodes proton chemical shifts (entries 1–15), backbone ^{15}N resonances (entry 16) and carbon chemical shifts (position 17–30). The nomenclature generally follows BMRB standards with some simplifications regarding the aliphatic CH_2 and CH_3 side-chain moieties (see supporting Table 1 for a complete list of definitions). NMR data sets of the protein of interest are hence represented by an $N \times 31$ dimensional matrix (with N corresponding to the amino acid sequence) of the form: $A_{i,j} = (\text{aa}_{i,0}, \text{cs}[\text{X}]_{i,j})$ with $i = 1, \dots, N$ and $j = 1, \dots, 30$. In this matrix, the amino acid sequence aa is denoted in standard single letter format. In addition, it is also possible to include secondary structure information using an additional vector with the dimension N in which for each residue the single letters α , β or c/u stand for α -helix, β -strand or random coil/undefined, respectively (Fig. 1, Step 1). Backbone chemical shift values (^1H , ^{15}N and ^{13}C) are taken from standard tables such as given in (Wang and Jardetzky 2002) corresponding to entries $\text{cs}[\text{X}]_{i,1-2}$, $\text{cs}[\text{X}]_{i,16}$ and $\text{cs}[\text{X}]_{i,17-19}$ (supporting Table 1)

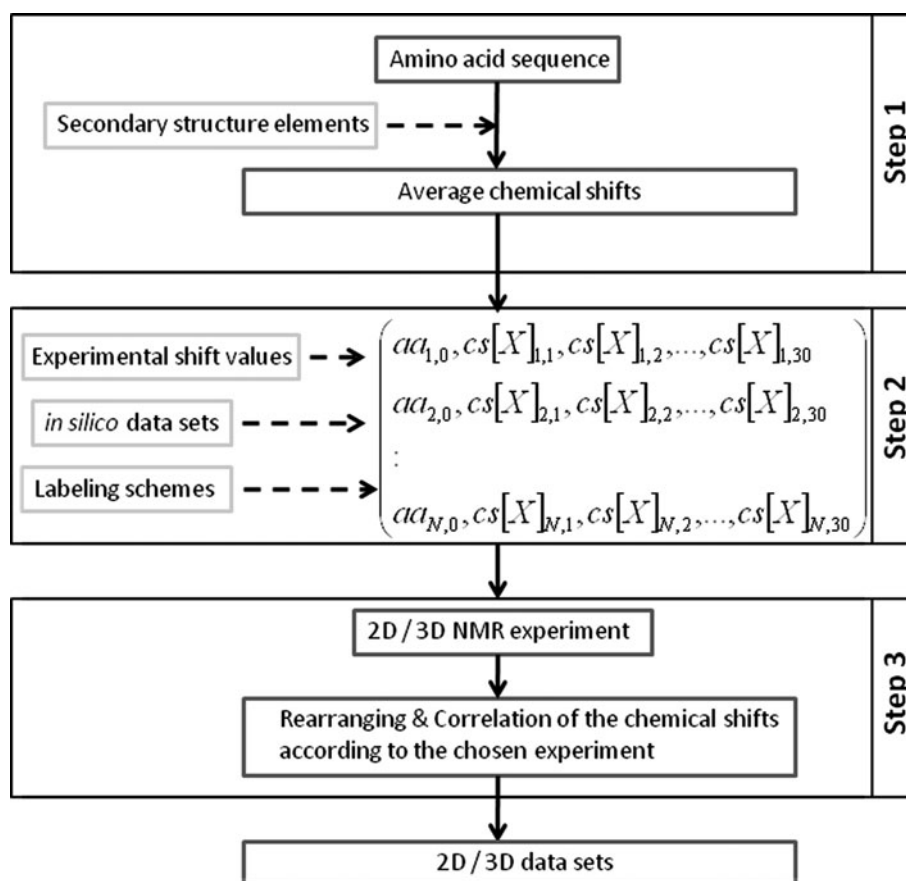
and for side chains from the BMRB data base (<http://www.bmrb.wisc.edu/>). As shown in Fig. 1, this information provides the input data that can subsequently be used for the prediction of NMR spectra.

In Step 2 (see Fig. 1), the input data can be further refined using experimental NMR chemical shift assignments deduced from reference samples or obtained under different experimental conditions. Furthermore, in silico chemical shift data predicted by computational procedures such as ShiftX (Neal et al. 2003), ShiftX2 (Han et al. 2011) or SPARTA+ (Shen and Bax 2010) can be included. Furthermore, tailored labeling schemes can be used at the input stage 2 of FANDAS. So far, (1,3- ^{13}C)-glycerol labeling and (2- ^{13}C)-glycerol labeling (LeMaster and Kushlan 1996; Hong and Jakes 1999; Castellani et al. 2002), fractional deuteration (Rosen et al. 1996; Nand et al. 2012), and ^{13}C , ^{15}N reverse labeling (Vuister et al. 1994; Etkorn et al. 2007) have been implemented. Scrambling can be included as exemplified below. Finally, Step 2 can be used to provide distance constraints on the basis of a given PDB file and a user-defined distance cut off.

The resulting input matrix can be retrieved at every stage of the input session consisting of the stages 1 (*Input Generation*) and stage 2 (*Optional Input Modification*) (Fig. 2). In the next stage (Step 3, *Spectrum prediction*), the user can select from a list of currently available 2D and 3D NMR experiments that can be predicted (Fig. 2; supporting Table 2). Figure 3 exemplifies the different stages of the input and spectrum modules for the case of the prediction of an intra-residue (^{13}C , ^{13}C) correlation spectrum assuming the peptide sequence Met–Ala–Thr–His–Glu (Fig. 3a, upper panel) and 1,3- ^{13}C glycerol labeling (Fig. 3a, lower segment). Since no information is given regarding the secondary structure, the program assumes statistical average values for all atoms under consideration. Note that scrambling of individual atomic positions can be specified by selecting *indicate relative labeling* within the Spectrum prediction module (Fig. 3b, upper segment). Furthermore, residue-specific probability limits can be defined allowing to adapt predictions to the expected amino-acid specific scrambling pattern. Depending on the experiment of interest, FANDAS rearranges and correlates the chemical shifts of the input file and generates predictions for spectra of intra-residue, sequential or distance-encoded data sets. For the later case, distance lists can be generated automatically from a PDB file under Step 2. In case of a multimer, FANDAS requires also the chain ID.

Finally, FANDAS returns a list of chemical shift predictions as an output file (Fig. 3b, lower panel) that can be downloaded and subsequently used directly as a peak list in NMR analysis programs. For example, when using the Sparky software (Goddard and Kneller 2006), the predictions are simply read in as a peak list via the command

Fig. 1 Flow chart of the different stages of FANDAS. *Solid arrows* denote compulsory steps and *dashed lines* reflect optional input modifications



“rp”. Note that question marks in the output file have an important function as they indicate the dimension of the experiment. Atom names and further additional information (such as scrambling parameters or interatomic distances) are stored in the columns named “notes” of the Sparky peak list.

Applications

In the following, we demonstrate how FANDAS can facilitate different aspects of a ssNMR-based data analysis. In Fig. 4a, a 2D ^{13}C – ^{13}C proton-driven spin diffusion (PDS) experiment is shown on sensory rhodopsin II (SRII) that was reconstituted using ^{13}C labeled lipids from *N. pharaonis*. Data were acquired with a mixing time of 20 ms, leading to the dominant contribution of one-bond contacts in the spectrum. Since both protein and lipids are ^{13}C labeled, the spectrum contains signal contributions of both molecular components. Protein chemical shift predictions by FANDAS help to distinguish lipid from the protein signals. Note that lipid signals are not only found in spectral regions well separated for aliphatic protein carbons, i.e., around 75–80 ppm but also extend into spectral segments between the 40 and 35 ppm region close to protein side-chain correlations. Figure 4b shows an

experimental ^{13}C – ^{13}C spin diffusion spectrum of the pure lipid sample with the same predictions confirming the FANDAS-based discrimination between protein and lipid signals on an experimental level.

In Fig. 4, we made use of Step 2 of the FANDAS input module by combining in silico and experimental data sets. Experimental SRII ssNMR chemical shift assignments (red) were taken from (Etzkorn et al. 2007). Additional chemical shift predictions were generated using ShiftX2 on the basis of the crystal structure (green, PDB file: 1GU8). Further modifications, for example by including solution-state NMR assignments are also possible. Average chemical shifts taken from the BMRB were included for side chain atoms using the simplified side chain topologies given in supporting table 1 whenever experimental assignments were unavailable. Focusing on the threonine/serine region (Fig. 4, insert) illustrates how FANDAS predictions can facilitate a further analysis of the experimental data. Here, C β –C α correlations corresponding to residue Thr79 as predicted from the X-ray structure (green) would be consistent with the assignment shown in the experimental spectrum of Fig. 4. Another example is Ser153 for which the FANDAS prediction (green) and the experimental assignment (red) nicely match.

In addition to the analysis of 2D data, FANDAS can also be used to predict three-dimensional data sets. In Fig. 5,

Fig. 2 Main page of the FANDAS program highlighting the different program steps as well as a list of currently available experiments under Step 3

home >> FANDAS

FANDAS

WeNMR web portal

Return to the WeNMR website

WELCOME TO THE WENMR FANDAS WEB PORTAL >>

THE FANDAS WEB SERVER

Some small introduction to FANDAS

You have not yet created an input table from sequence. Proceed with step 1 or **retrieve** a previously saved file by name

STEP 1: INPUT GENERATION

Tool	Generated File
Generate input table from sequence & Include average chemical shifts	

STEP 2: OPTIONAL INPUT MODIFICATION

Tool	Generated File
Read BMRB tables	
Read C-C and H-H distances from PDB	
Remove ^{13}C and ^{15}N shifts for reversely labeled amino acids	
Adapt input list to (1,3- ^{13}C)-glycerol labeling	
Adapt input list to (2- ^{13}C)-glycerol labeling	
Adapt input list to fractional deuteration	

STEP 3: SPECTRUM PREDICTION (SPARKY FORMAT)

Tool	Generated File
Predict intra-residue (^{13}C - ^{13}C) correlation spectrum	
Predict sequential (^{13}C - ^{13}C) correlation spectrum (i to i-1 and i+1)	
Predict (^{13}C - ^{13}C) correlation spectrum (distance encoded)	
Predict 2D DQ ^{13}C - ^{13}C spectrum (directly bonded C)	
Predict 2D ^{15}N - ^{13}C A spectrum	
Predict 2D ^{15}N - ^{13}C A- ^{13}C X spectrum (intra)	
Predict 2D ^{15}N - ^{13}C O spectrum	
Predict 2D ^{15}N - ^{13}C O- ^{13}C A/CB spectrum	
Predict 2D ^{15}N - ^{13}C O- ^{13}C X spectrum	
Predict backbone ^1H - ^{15}N spectrum	
Predict ^1H - ^{13}C spectrum (directly bonded)	
Predict ^{13}C - ^1H spectrum (directly bonded)	
Predict (^1H - ^1H) NOESY spectrum (intra)	
Predict (^1H - ^1H) NOESY spectrum (distance encoded)	
Predict CHHC spectrum (distance encoded)	
Predict NHHC spectrum (distance encoded)	
Predict 2D CHH spectrum (distance encoded)	
Predict 2D NHH spectrum (distance encoded)	
Predict 2D HHC spectrum (distance encoded)	
Predict 3D HHC spectrum (distance encoded)	
Predict 3D CHH spectrum (distance encoded)	
Predict 3D NHH spectrum (distance encoded)	
Predict 3D ^{15}N - ^{13}C A- ^{13}C X spectrum (intra)	
Predict 3D ^{15}N - ^{13}C A- ^{13}C X spectrum (inter)	
Predict 3D ^{15}N - ^{13}C A- ^{13}C X spectrum (^{13}C distance encoded)	
Predict 3D ^{15}N - ^{13}C O- ^{13}C X spectrum	
Predict 3D ^{15}N - ^{13}C O- ^{13}C X spectrum (^{13}C distance encoded)	
Predict 3D ^{13}C SQQSQ spectrum (inter)	
Predict 3D ^{13}C DQQSQ spectrum (intra)	
Predict 3D ^{13}C DQQSQ spectrum (inter)	

(a)

WELCOME TO THE WENMR FANDAS WEB PORTAL >>

THE FANDAS WEB SERVER

Provide a filename:

Protein sequence (single letter amino acid code). Required
Format: e.g. MATGPQR

Secondary structure specification (option)
For H,HA,N,CO,CA,CB you can demand secondary structure specific chemical shifts according to Wang and Jardetzky. Identifiers are b (beta-sheet), c (coil), and a (alpha-helix) e.g. cccccaccbbbbc etc.

(b)

WELCOME TO THE WENMR FANDAS WEB PORTAL >>

PREDICT INTRA-RESIDUE (^{13}C - ^{13}C) CORRELATION SPECTRUM VIA FILE

[Click here to download the generated file](#)

Choose a filename:

indicate relative labeling for 1,3-glycerol labeling
 indicate relative labeling for 2-glycerol labeling

Provide a probability limit:

FANDAS main menu

```

7-? 176.21 56.14 M1CO-CA int:0.5-0.67
7-? 176.21 33 M1CO-CB int:0.5-0.83
7-? 176.21 32.04 M1CO-CG1 int:0.5-0.33
7-? 176.21 17.14 M1CO-CE1 int:0.5-1
7-? 56.14 176.21 M1CA-CO int:0.67-0.5
7-? 56.14 33 M1CA-CB int:0.67-0.83
7-? 56.14 32.04 M1CA-CG1 int:0.67-0.33
7-? 56.14 17.14 M1CA-CE1 int:0.67-1
7-? 33 176.21 M1CB-CO int:0.83-0.5
7-? 33 56.14 M1CB-CA int:0.83-0.67
7-? 33 32.04 M1CB-CG1 int:0.83-0.33
7-? 33 17.14 M1CB-CE1 int:0.83-1
7-? 32.04 176.21 M1CG1-CO int:0.33-0.5
7-? 32.04 56.14 M1CG1-CA int:0.33-0.67
7-? 32.04 33 M1CG1-CB int:0.33-0.83
7-? 32.04 17.14 M1CG1-CE1 int:0.33-1
7-? 17.14 176.21 M1CE1-CO int:1-0.5
7-? 17.14 56.14 M1CE1-CA int:1-0.67
7-? 17.14 33 M1CE1-CB int:1-0.83
7-? 17.14 32.04 M1CE1-CG1 int:1-0.33
7-? 177.77 18.99 A2CO-CB int:1-1
7-? 18.99 177.77 A2CB-CO int:1-1

```

WELCOME TO THE WENMR FANDAS WEB PORTAL >>

ADAPT INPUT LIST TO (1,3- ^{13}C)-GLYCEROL LABELING

Choose a filename:

FANDAS main menu

Fig. 3 Interfaces reflecting the different stages of a FANDAS-based data production: In **a**, sequence input and definition of the file name (*upper segment*) as well as of the labeling pattern (1,3- ^{13}C glycerol labeling, *lower panel*) are shown. In **b** (*upper panel*), the experimental

data set to be predicted (intra-residue CC, with a specific file name) as well as the relative labeling pattern with the certain probability limit is given. Finally, the *lower panel* of **b** shows the resulting output file which can be read in as a peak list in NMR analysis software

results are shown for an NCOCX experiment obtained on a (^{13}C , ^{15}N labeled) sample of the cyclic nucleotide binding domain (CNBD) (Cukkemane et al. 2012) from a bacterial ion channel. FANDAS predictions were generated on the basis of solution-state NMR assignments (Schunke et al. 2009) (Fig. 5, red). In Fig. 5 a, b, consecutive planes in the 2–3 and 1–2 dimensions are shown, respectively. Comparison to the experimental data facilitated tentative assignments such as those indicated in green for a selected set of correlations. These were subsequently cross validated by additional ssNMR experiments leading to the assignments reported in (Cukkemane et al. 2012). Note that such studies can be performed interactively when scanning to the 3D experimental data sets. Moreover, FANDAS assembles 3D datasets within seconds after input parameters are defined.

Previously, sparse labeling has been used to obtain long-range distance constraints in several ssNMR studies (Castellani et al. 2002; Loquet et al. 2012). In such experiments, details of the protein labeling and potential scrambling effects can play an important role for the structural analysis. Again, FANDAS can aid such studies as demonstrated

in Fig. 6 for the case of the membrane-embedded SRII/HtrII complex. Here, only the receptor was isotope-labeled using 1,3- ^{13}C glycerol in the medium. The experimental results shown in Fig. 6 correspond to a ^{13}C - ^{13}C spin diffusion spectrum with a mixing time of 150 ms. Chemical shift predictions were generated for a distance-encoded ^{13}C - ^{13}C spectrum using 5 Å (red) and 3 Å (green) distance cut offs under the assumption of 1,3- ^{13}C glycerol labeling pattern. Chemical shift values as well as C–C distances were extracted from a structural model of SRII in complex with HtrII (1–157) obtained on the basis of crystallographic and NMR data (Etzkorn et al. 2007; Etzkorn et al. 2010). Comparison of the predictions to the experimental data indicates that several empirical correlations (such as cross peaks between 55 and 15 ppm) must stem from polarization transfer over distances larger than 3 Å. The remarkable differences between the FANDAS spectra in Fig. 6a, b originate from the applied probability limit which is computed by multiplying the relative degree of ^{13}C labeling of spin X and spin Y, giving rise to cross peak XY with an intensity $L(X) \cdot L(Y)$ (Higman et al. 2009). In Fig. 6a, all non-vanishing correlations (i.e., $L(X) \cdot L(Y) > 0$) are

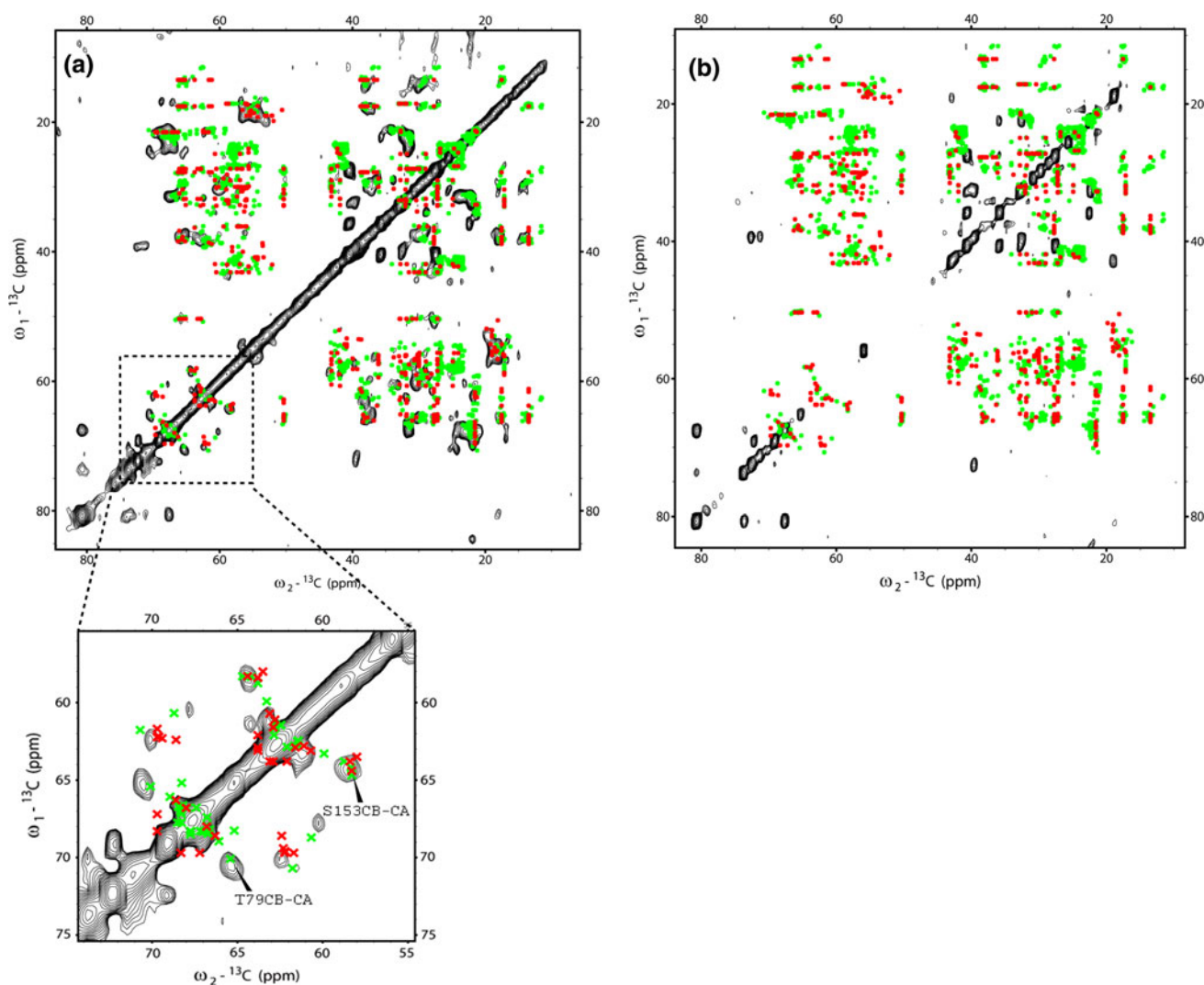


Fig. 4 Comparison of experimental (^{13}C , ^{13}C) PDSB experiments on ^{13}C -labeled (SRII, Lipids) **a** and ^{13}C liposomes **b**. For the given mixing time of 20 ms, only intra-residue contacts are expected. FANDAS predictions (*crosses*) of the SRII were based on previous

ssNMR assignments (Etzkorn et al. 2007) (*red*) and the crystal structure (*green*; PDB file: 1GU8). Combination of those data facilitates the further analysis of the spectral region comprising threonine and serines (extract)

included whereas Fig. 6b assumes a limit $L(X)\cdot L(Y) > 0.55$. Such sets of predictions allow assessing the level of scrambling in a particular sample as long as structural information as well as a distance classification of the polarization transfer dynamics for a given mixing time are available (Lange et al. 2003; Manolikas et al. 2008). In the case of the SRII/HtrII sample, the FANDAS data could be used to refine the structural model and could be compared to additional experimental data obtained using longer (^{13}C , ^{13}C) mixing times.

Finally, we demonstrate the utility of FANDAS in a stage where structural information available from other resources is used to decide on the optimal NMR labeling in terms of spectral dispersion. As an example, we again consider the case of SRII. In Fig. 7a, FANDAS predictions

of an NCOCA spectrum were obtained using the crystal structure (PDB file: 1GU8) and assuming uniformly ^{13}C , ^{15}N labeled SRII (green) and a VLFY reverse labeled sample (red). FANDAS predicts a sizable reduction in spectral crowding, for example for ^{13}C resonance frequencies between 65 and 70 ppm. Taking into account the intrinsic accuracy of chemical-shift predictions for ssNMR (see, e.g., Seidel et al. 2009) and the possibility of N–C β correlations for an experimental PDSB mixing of 40 ms, experimental and predicted results (Fig. 7b) are in good agreement. In a next stage, FANDAS could be used to aid an ssNMR-based analysis using complementary labeling schemes. As an example, Fig. 7c compares predictions for fully labeled SRII (green), a VLFY-reverse labeled sample (red) and a preparation of a TRIYM-reverse labeled sample

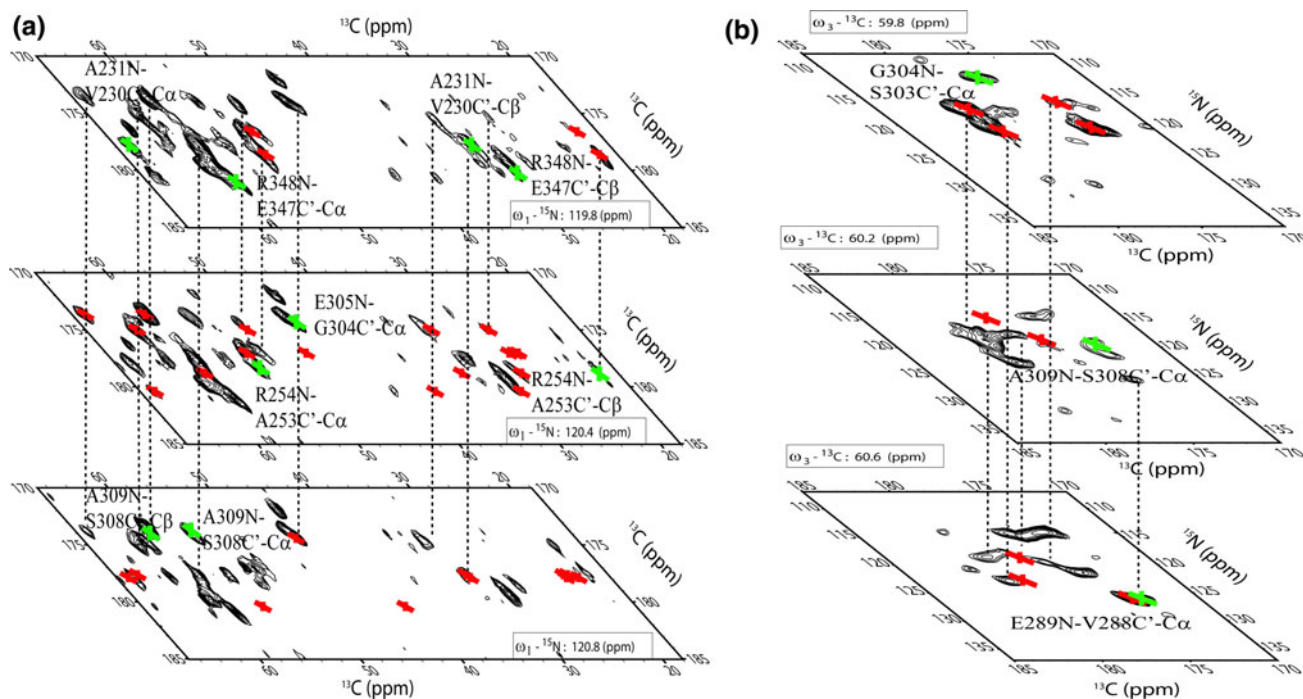


Fig. 5 Consecutive planes of a 3D NCOEX experiment of CNBD along different axes. 2D planes indicated in the *top* and the *lower* panel are cut along the ^{15}N and ^{13}C frequencies, respectively.

Chemical shift predictions (*red*) are based on solution state NMR assignments (PDB file: 2K0G) and allowed to obtain several tentative assignments (*green crosses*)

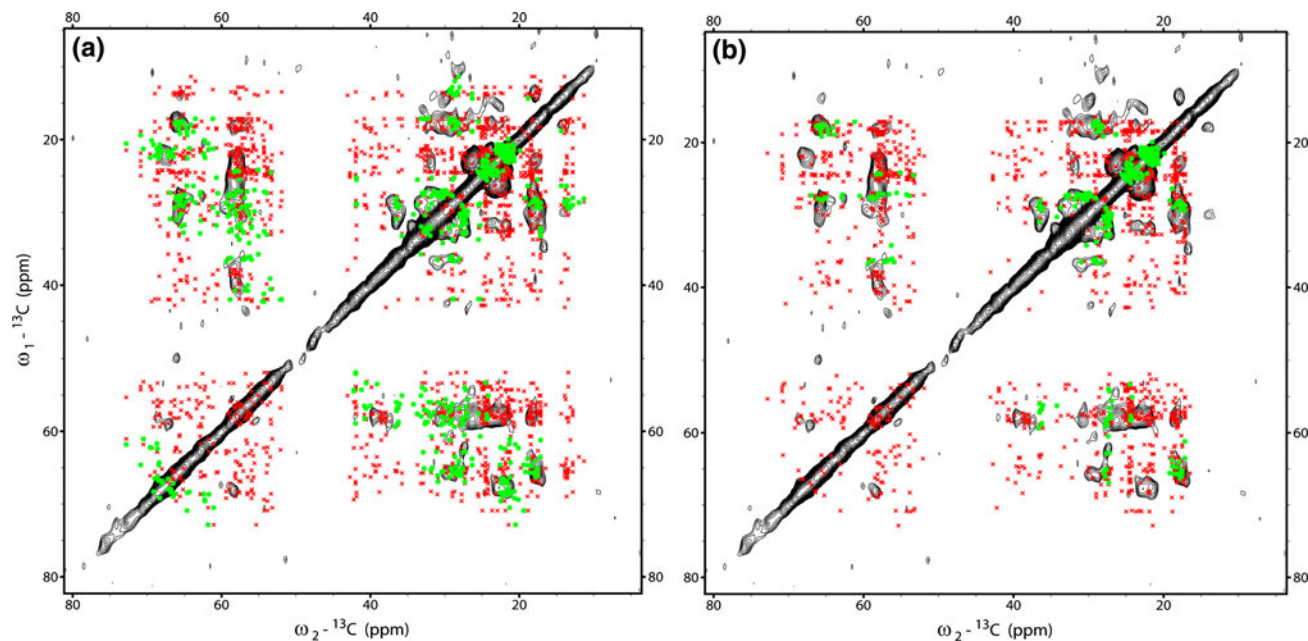


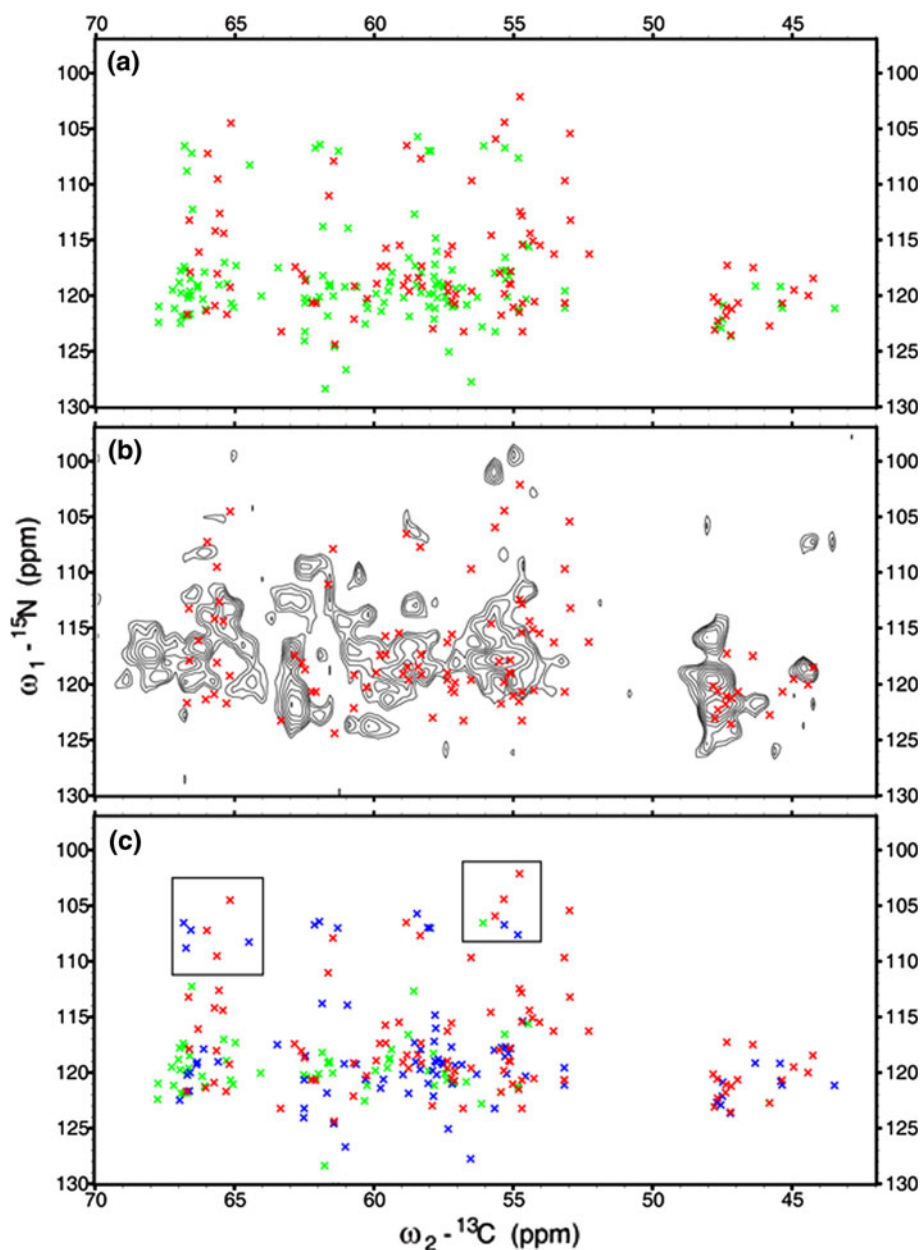
Fig. 6 ^{13}C - ^{13}C spin diffusion (mixing time 150 ms) spectrum of 1,3- ^{13}C glycerol labeled SR11 in complex with unlabeled HtrII in lipid bilayers. Predictions were generated for a distance encoded ^{13}C - ^{13}C spectrum with 5 Å (*red*) and 3 Å (*green*) distance cut offs,

respectively. Both, chemical shift values as well as distances were extracted from a homology model based on X-ray and NMR data (Etzkorn et al. 2010). In **a**, all possible correlations are present whereas in **b** only correlations with a probability $\geq 56\%$ are included

(blue). According to FANDAS, the combination of both labeling schemes would offer complementary spectral dispersion, for example in the indicated boxes referring to

sequential contacts involving glycine-valine (blue) and glycine-isoleucine correlations (red) (^{13}C resonance frequency around 65 ppm) as well as glycine-leucine (blue)

Fig. 7 FANDAS predictions for a 2D NCOCA experiment on **a** ^{13}C , ^{15}N uniformly labeled SRII sample (*green*) as well as a VLFY (*red*) and TRIYM (*blue*, **c**) reverse labeled sample. Predictions are based on the crystal structure (PDB file: 1GU8) suggesting a sizable reduction in spectral crowding and ambiguity for the reversed labeled preparations. In **b**, FANDAS predictions for an VLFY-labeled sample are compared to experimental results. In **c**, the two reverse labeling schemes are compared. Spectral regions in which FANDAS predicts complementary information, taking into account the intrinsic spectral resolution (Seidel et al. 2009) are indicated by *boxes*



and glycine-alanine (red) pairs close to the ^{13}C resonance frequency of 55 ppm.

Conclusions

We have introduced a computational environment that allows rapid prediction of multi-dimensional NMR data sets. These can be constructed even if empirical data are missing or incomplete and can be utilized at various stages of an NMR-based biomolecular study. Comparison of FANDAS sets to experimental results can facilitate the spectral assignment process as demonstrated for the case of 2D ^{13}C – ^{13}C spin diffusion experiments on SRII and a 3D

NCOCA experiment on CNBD. In addition, FANDAS readily allows distinguishing protein correlations from other molecular contributions such as lipids, peptidoglycans or nucleic acids (Renault et al. 2012a, b) and can be further extended to predict NMR correlations of such non-proteinaceous molecular components. Furthermore, labeling schemes and scrambling effects can be analyzed as illustrated here for the case of the membrane-embedded SRII/HtrII complex. Such considerations are important if distance restraints are to be deduced from multidimensional correlation experiments.

Especially in the early stage of an NMR study, the selection of the most suitable biomolecular labeling pattern or NMR experiment is often critical to maximize spectral

resolution and dispersion. FANDAS can be readily used for such considerations because the availability of experimental data is not critical for its application. Since FANDAS works from chemical-shift inputs derived from arbitrary PDB files, it also allows for an initial quality assessment of the NMR data if a reference structure is available. The modular character of FANDAS in principle allows further extensions in terms of (solid or liquid-state) NMR experiments, possibly including four-dimensional data sets or focusing on experiments that probe molecular motion on a residue-specific level. In summary, FANDAS provides a valuable tool to aid NMR-based studies of complex biomolecules at virtually all experimental stages starting from sample preparation to (supra)molecular structure elucidation.

Acknowledgments This work was supported by the Netherlands Organization for Scientific Research (NWO) (grants 700.26.121 and 700.10.443 to M.B.) and by the European Community's Seventh Framework Program, BioNMR project, contract number 211800, and the FP7 e-Infrastructure project WeNMR (contract no. 261572, www.wenmr.eu). We would like to thank Dieter Oesterhelt (Martinsried) for informing us that *N. pharaonis* can grow with acetate as sole carbon source. Initial work of J. Saueremann on the extraction of *N. pharaonis* lipids is gratefully acknowledged.

References

- Ader C, Frey S, Maas W, Schmidt HB, Goerlich D, Baldus M (2010) Amyloid-like interactions within nucleoporin FG hydrogels. *Proc Natl Acad Sci* 107:6281–6285
- Alber F, Forster F, Korkin D, Topf M, Sali A (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem Palo Alto Annu Rev Palo Alto Annu Rev* 77:443–477
- Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol Phys* 95:1197–1207
- Brothers MC, Nesbitt A, Hallock M, Rupasinghe S, Tang M, Harris J, Baudry J, Schuler M, Rienstra C (2012). VITAL NMR: using chemical shift derived secondary structure information for a limited set of amino acids to assess homology model accuracy. *J Biomol NMR* 52:41–56
- Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 420: 98–102
- Cukkemane A, Nand D, Gradmann S, Weingarth M, Kaupp B, Baldus M (2012) Solid-state NMR [^{13}C , ^{15}N] resonance assignments of the nucleotide-binding domain of a bacterial cyclic nucleotide-gated channel. *Biomol NMR Assign* 6:225
- de Vries SJ, van Dijk M, Bonvin AMJJ (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 5:883–897
- Domínguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125:1731–1737
- Etzkorn M, Martell S, Andronesi OC, Seidel K, Engelhard M, Baldus M (2007) Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew Chem Int Ed* 46:459–462
- Etzkorn M, Seidel K, Li L, Martell S, Geyer M, Engelhard M, Baldus M (2010) Complex formation and light activation in membrane-embedded sensory rhodopsin II as seen by solid-state NMR spectroscopy. *Structure* 18:293–300
- Fossi M, Castellani T, Nilges M, Oschkinat H, van Rossum BJ (2005) SOLARIA: a protocol for automated cross-peak assignment and structure calculation for solid-state magic-angle spinning NMR spectroscopy. *Angew Chem Int Ed* 44:6151–6154
- Gelis I, Bonvin AMJJ, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131:756–769
- Goddard TD, Kneller D (2006). SPARKY 3. University of California, California
- Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. *Q Rev Biophys* 44:257–309
- Han B, Liu Y, Ginzing S, Wishart D (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50:43–57
- Hefke F, Bagaria A, Reckel S, Ullrich S, Dötsch V, Glaubitz C, Güntert P (2011) Optimization of amino acid type-specific ^{13}C and ^{15}N labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. *J Biomol NMR* 49(2):75–84
- Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D, Baldus M (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length α -synuclein fibrils studied by solid-state NMR. *Proc Natl Acad Sci USA* 102(44):15871–15876
- Higman V, Flinders J, Hiller M, Jehle S, Markovic S, Fiedler S, van Rossum B-J, Oschkinat H (2009) Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively ^{13}C -labelled proteins. *J Biomol NMR* 44(4):245–260
- Hohenfeld IP, Wegener AA, Engelhard M (1999). Purification of histidine tagged bacteriorhodopsin, *pharaonis* halorhodopsin and *pharaonis* sensory rhodopsin II functionally expressed in *Escherichia coli*. *FEBS Lett* 442:198–202
- Hong M, Jakes K (1999) Selective and extensive C-13 labeling of a membrane protein for solid-state NMR investigations. *J Biomol NMR* 14:71–74
- Hu KN, Qiang W, Tycko R (2011) A general Monte Carlo/simulated annealing algorithm for resonance assignment in NMR of uniformly labeled biopolymers. *J Biomol NMR* 50:267–276
- Jehle S, Vollmar BS, Bardiaux B, Dove KK, Rajagopal P, Gonen T, Oschkinat H, Klevit RE (2011) N-terminal domain of αB -crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc Natl Acad Sci* 108:6409–6414
- Kato H, van Ingen H, Zhou B-R, Feng H, Bustin M, Kay LE, Bai Y (2011) Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci* 108:12283–12288
- Klare JP, Bordignon E, Doebber M, Fitter J, Kriegsmann J, Chizhov I, Steinhoff HJ, Engelhard M (2006) Effects of solubilization on the structure and function of the sensory rhodopsin II/transducer complex. *J Mol Biol* 356:1207–1221
- Lange A, Seidel K, Verdier L, Luca S, Baldus M (2003) Analysis of proton–proton transfer dynamics in rotating solids and their use for 3D structure determination. *J Am Chem Soc* 125: 12640–12648
- Lange A, Giller K, Hornig S, Martin-Eauclaire M-F, Pongs O, Becker S, Baldus M (2006) Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. *Nature* 440:959–962

- LeMaster DM, Kushlan DM (1996) Dynamical mapping of *E. coli* thioredoxin via C-13 NMR relaxation analysis. *J Am Chem Soc* 118:9255–9264
- Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19:315–316
- Loquet A, Bardiaux B, Gardiennet C, Blanchet C, Baldus M, Nilges M, Malliavin T, Bockmann A (2008) 3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints. *J Am Chem Soc* 130:3579–3589
- Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A (2012) Atomic model of the type III secretion system needle. *Nat Adv (online publication)*
- Manolikas T, Herrmann T, Meier BH (2008) Protein structure determination from ^{13}C spin-diffusion solid-state NMR spectroscopy. *J Am Chem Soc* 130:3959–3966
- Melquiond ASJ, Karaca E, Kastritis PL, Bonvin AMJJ (2012) Next challenges in protein–protein docking: from proteome to interactome and beyond. *WIREs Comput Mol Sci* 2:642–651
- Nand D, Cukkemane A, Becker S, Baldus M (2012) Fractional deuteration applied to biomolecular solid-state NMR spectroscopy. *J Biomol NMR* 52:91–101
- Neal S, Nip AM, Zhang HY, Wishart DS (2003) Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. *J Biomol NMR* 26:215–240
- Nelson SJ, Schneider DM, Wand AJ (1991) Implementation of the main chain directed assignment strategy—computer-assisted approach. *Biophys J* 59:1113–1122
- Renault M, Cukkemane A, Baldus M (2010) Solid-state NMR spectroscopy on complex biomolecules. *Angew Chem Int Ed* 49:8346–8357
- Renault M, Bos MP, Tommassen J, Baldus M (2011) Solid-state NMR on a large multidomain integral membrane protein: the outer membrane protein assembly factor BamA. *J Am Chem Soc* 133:4175–4177
- Renault M, Pawsey S, Bos MP, Koers EJ, Nand D, Tommassen-van Bortel R, Rosay M, Tommassen J, Maas WE, Baldus M (2012a) Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew Chem Int Ed* 51:2998–3001
- Renault M, Tommassen-van Bortel R, Bos MP, Post JA, Tommassen J, Baldus M (2012b) Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc Natl Acad Sci* 109:4863–4868
- Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* 263:627–636
- Ruschak AM, Kay LE (2010) Methyl groups as probes of supramolecular structure, dynamics and function. *J Biomol NMR* 46:75–87
- Scholz I, Huber M, Manolikas T, Meier BH, Ernst M (2008) MIRROR recoupling and its application to spin diffusion under fast magic-angle spinning. *Chem Phys Lett* 460:278–283
- Schunke S, Stoldt M, Novak K, Kaupp UB, Willbold D (2009) Solution structure of the *Mesorhizobium loti* K1 channel cyclic nucleotide-binding domain in complex with cAMP. *EMBO Rep* 10:729–735
- Seidel K, Etzkorn M, Heise H, Becker S, Baldus M (2005) High-resolution solid-state NMR studies on uniformly C-13, N-15 -labeled ubiquitin. *Chem BioChem* 6:1638–1647
- Seidel K, Etzkorn M, Schneider R, Ader C, Baldus M (2009) Comparative analysis of nmr chemical shift predictions for proteins in the solid phase. *Solid State NMR* 35:235–242
- Shen Y, Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* 48:13–22
- Stevens T, Fogh R, Boucher W, Higman V, Eisenmenger F, Bardiaux B, van Rossum B-J, Oschkinat H, Laue E (2011) A software framework for analysing solid-state MAS NMR data. *J Biomol NMR* 51:437–447
- Vuister GW, Kim SJ, Wu C, Bax A (1994) 2d and 3d NMR-study of phenylalanine residues in proteins by reverse isotopic labeling. *J Am Chem Soc* 116:9206–9210
- Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* 11:852–861
- Wegener AA, Klare JP, Engelhard M, Steinhoff HJ (2001) Structural insights into the early steps of receptor-transducer signal transfer in archaeal phototaxis. *EMBO J* 20:5312–5319
- Zimmerman DE, Kulikowski CA, Huang YP, Feng WQ, Tashiro M, Shimotakahara S, Chien CY, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. *J Mol Biol* 269:592–610