

A software framework for analysing solid-state MAS NMR data

Tim J. Stevens · Rasmus H. Fogh · Wayne Boucher · Victoria A. Higman ·
Frank Eisenmenger · Benjamin Bardiaux · Barth-Jan van Rossum ·
Hartmut Oschkinat · Ernest D. Laue

Received: 15 April 2011 / Accepted: 5 September 2011 / Published online: 28 September 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Solid-state magic-angle-spinning (MAS) NMR of proteins has undergone many rapid methodological developments in recent years, enabling detailed studies of protein structure, function and dynamics. Software development, however, has not kept pace with these advances and data analysis is mostly performed using tools developed for solution NMR which do not directly address solid-state specific issues. Here we present additions to the CcpNmr Analysis software package which enable easier identification of spinning side bands, straightforward analysis of double quantum spectra, automatic consideration of non-uniform labelling schemes, as well as extension of other existing features to the needs of solid-state MAS data. To underpin this, we have updated and extended the CCPN data model and experiment descriptions to include transfer types and nomenclature appropriate for solid-state NMR experiments, as well as a set of experiment prototypes covering the experiments commonly employed by solid-state MAS protein NMR spectroscopists.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9569-2) contains supplementary material, which is available to authorized users.

T. J. Stevens · R. H. Fogh · W. Boucher · E. D. Laue (✉)
Department of Biochemistry, University of Cambridge,
Cambridge CB2 1GA, UK
e-mail: e.d.laue@bioc.cam.ac.uk

V. A. Higman
Department of Biochemistry, University of Oxford,
South Parks Road, Oxford OX1 3QU, UK

F. Eisenmenger · B. Bardiaux · B.-J. van Rossum ·
H. Oschkinat
Department of Structural Biology, Leibniz-Institut für
Molekulare Pharmakologie, Robert-Roessle-Str. 10,
13125 Berlin, Germany

This work not only improves solid-state MAS NMR data analysis but provides a platform for anyone who uses the CCPN data model for programming, data transfer, or data archival involving solid-state MAS NMR data.

Keywords Solid-state MAS NMR · CCPN · Experiment nomenclature · Software · Data model

Introduction

During the past decade solid-state magic-angle-spinning (MAS) NMR has developed into a useful tool for the study of protein structure and dynamics. With no intrinsic size limit on the structure studied and no requirement for long range order, it forms a highly complementary technique to solution NMR and X-ray crystallography. It has enabled structure determination in cases which are difficult to tackle by other methods, such as amyloid fibrils (Jaroniec et al. 2004; Petkova et al. 2004; Ferguson et al. 2006; Wasmer et al. 2008), dynamic oligomeric protein complexes (Jehle et al. 2010) as well as membrane proteins in their native lipid environment (Andronesi et al. 2005; Cady et al. 2010). So far, data analysis and structure calculations have mainly been carried out using software tools developed for solution NMR. These include spectrum visualisation programs used for resonance and distance restraint assignments (Goddard and Kneller; Johnson and Blevins 1994), structure calculation programs (Brünger 1992; Güntert et al. 1997; Brünger et al. 1998; Schwieters et al. 2003; Rieping et al. 2007) as well as tools to extract torsion angle restraints from chemical shifts (Shen et al. 2009; Cheung et al. 2010) and to validate structures (Hooft et al. 1996; Bhattacharya et al. 2007) (<http://nmr.cmbi.ru.nl/cing/>). Over the years, increasingly sophisticated software tools

have become available to solution NMR spectroscopists which enable (semi-) automated data analysis and significantly speed up the overall process of protein structure determination (Zimmerman et al. 1997; Herrmann et al. 2002; Jung and Zweckstetter 2004; Huang et al. 2005; Rieping et al. 2007; Lemak et al. 2011). The Collaborative Computation Project for the NMR community (CCPN) has the aim of streamlining these processes even further. To this end a data model describing all the information used during NMR analysis and in related disciplines was developed (Vranken et al. 2005). Based on this data model, the CCPN group has further developed a spectrum analysis program (CcpNmr Analysis) and a format converter. Importantly, the data model forms an informatics system which other programs can build on and which facilitates data storage and exchange: it enables movement from one software package to another without encountering difficulties with data conversion or loss of information. Initiatives to connect a wide range of NMR software packages with the CCPN data model have already interfaced several programs (www.extend-nmr.eu) and more are currently ongoing.

To date, few software packages include modules which have been designed specifically with solid-state MAS NMR data analysis in mind (Fossi et al. 2005; Loquet et al. 2008; Moseley et al. 2010; Tycko and Hu 2010). However, this is a definite requirement if solid-state NMR spectroscopists are to move towards an efficient, user-friendly and (semi-) automated data analysis pipeline such as that available for solution NMR. Solid-state MAS NMR has several requirements which are not met by the current solution-NMR based software tools. Spectral features such as side bands, the frequent use of experiments with double-quantum chemical shift axes or complicated isotope labelling schemes, such as those obtained when using 1,3-¹³C- or 2-¹³C-labelled glycerol as the sole carbon source in the bacterial growth medium, call for further software features to enable the spectral information to be captured efficiently, and for the spectral analysis to be made user-friendly. Finally, solid-state NMR uses different experiments from those used in solution, with a different set of magnetisation/coherence transfer types, different preferences for excitation and measurement, and a separate vocabulary to describe them. Descriptions of these experiments are useful because resonance assignment routines and data filtering facilities require knowledge of magnetisation transfer in the NMR experiments used; while the information has been available for some time within the CCPN data model for solution NMR experiments, solid-state MAS NMR experiments have not been included until now. Here we present extensions to the CCPN data model and CcpNmr Analysis software to address the needs of solid-state MAS NMR spectroscopists.

We provide a formal description of solid-state MAS NMR experiment types which concentrates on spectral features and downstream data analysis requirements, rather than on detailed transfer mechanisms. Descriptions of user-specified labelling patterns are introduced into the data model and these can be used as a filter within a variety of operations in CcpNmr Analysis. Furthermore, we have incorporated a double quantum chemical shift axis type into CcpNmr Analysis which enables automatic calculation of chemical shifts in the double quantum dimension and improved visual comparison with other spectra. In addition, a number of smaller features make the program especially user-friendly for solid-state NMR spectroscopists, and many features already available for solution-NMR data analysis have been extended to deal with solid-state MAS NMR data. Importantly, the extensions which describe MAS experiments and isotope labelling schemes in the CCPN data model have not only allowed new features to be introduced to CcpNmr Analysis, but are of benefit to anybody who uses the CCPN data model for programming, data transfer, or data archival involving solid-state MAS NMR data.

Reference data descriptions

Classification of NMR experiment types

The CCPN data model contains a description of NMR experiments at the magnetisation transfer level, with an associated naming convention. An earlier version of the system has been described previously (Fogh et al. 2006); this paper serves as a reference for a new and improved version of the naming convention. The data model and naming convention are described in detail in the supplementary material (Figure S1). The description abstracts the key information required for resonance assignment and other downstream data analysis from the physics of the NMR experiment. Thus magnetisation transfer by HSQC, HMQC or SPECIFIC cross-polarisation is uniformly described as ‘one-bond’, and NOESY, ROESY and long mixing time DARR transfer is uniformly described as ‘through-space’ (any deviations from this due to the isotopic labelling pattern used are dealt with downstream). The core of the data model represents the atom sites over which the magnetisation passes, the magnetisation transfers connecting these atoms, and the NMR measurements being recorded (which potentially arise from more than one atom site). The description of the 2D HH J-resolved experiment provides the simplest example: there are two atom sites, H1 and H2, connected by a J coupling magnetisation transfer. Here the NMR parameters being measured are the H1 chemical shift, and the H1H2 J coupling value. NMR

experiments descriptions are organised by magnetisation/coherence transfer pathway. For example the HCA(ONNH) experiment is potentially a 5D experiment, although spectroscopists usually acquire this experiment with fewer observed dimensions, e.g. a 4D HCA(CO)NNH, 3D H(CACO)NNH or 3D (H)CA(CO)NNH (where the frequencies of nuclei in brackets are not measured). In the data model the magnetisation transfer for this experiment is described only once in terms of the 5D experiment with five atom sites and the lower dimensional experiments are all described by reference to the top-level description.

The model corresponds to a convention for systematically naming experiments that is a more precise extension of the traditional NMR experiment names. The names reflect the underlying experiment descriptions in the data model and the general magnetisation transfer pathways found in the actual experiments. In brief, the atoms involved in the magnetisation transfer are listed in order; sites whose chemical shifts are not evolved are written in lower case; out-and-back transfers are indicated by square brackets; when the transfer is neither a one-bond transfer nor mediated by J-couplings, an underscore is inserted between the atom names, and the transfer types (e.g. ‘relayed’ or ‘through-space’) are listed in order after a full stop. For instance, a ^1H - ^{15}N HSQC, an H(CACO)NNH and a ^{15}N -NOESY-HSQC are represented as H[N], HcacoNH and H_H[N].through-space, respectively.

Extension of the data model and the naming convention to include solid-state MAS NMR experiments requires several considerations. Firstly, experiments which operate using very different physical mechanisms may produce very similar spectra e.g. ^{13}C - ^{13}C DARR (Takegoshi et al. 2001), RFDR (Bennett et al. 1992) and PAR (De Paëpe et al. 2008) experiments recorded at short mixing times will all result in ^{13}C - ^{13}C correlation spectra containing only intra-residue cross peaks, although they each exploit very different mechanisms in order to achieve this. Secondly, the same experiment recorded with different mixing times can produce very different results, e.g. a ^{13}C - ^{13}C DARR with a short mixing time will give rise to intra-residue cross peaks only (such as a TOCSY transfer would), but with a long mixing time inter-residue cross peaks can be seen too (such as observed with a NOESY transfer). Furthermore, the labelling scheme employed can also affect the transfers seen at particular mixing times. The purpose of the data model experiment types is to encode which information is contained within the spectrum so that this can be used downstream to weight or filter assignment options, i.e. which atoms can give rise to resonances in which dimension and what their bonding relationships are. This limited purpose can be achieved with a small number of transfer types, avoiding the need for maintaining a large number of different descriptions for

experiments with identical assignment information. The use of an increased number of transfer mechanisms (often with equivalent effect) in solid-state MAS compared to solution NMR experiments suggests that the terms NOESY or TOCSY should be avoided in a unified nomenclature. Thus the transfer types used by CCPN henceforth are the following:

- ‘onebond’—transfer between directly bonded atoms, regardless of transfer mechanism
- ‘relayed’—multiple stepwise transfer along chemical bonds (replaces ‘TOCSY’)
- ‘relayed-alternate’—like ‘relayed’, but peaks from alternating atoms along the bond network have alternating sign
- ‘through-space’—transfer through space, not limited to chemical bonds (replaces ‘NOESY’)
- ‘Jcoupling’—J coupling transfer, not limited to a particular number of bonds
- ‘Jmultibond’—J coupling transfer, excluding one-bond couplings

The default transfer type is generally ‘Jcoupling’ for H–H, H–F, and F–F transfers and ‘onebond’ for all other transfers (regardless of mechanism), but exceptions are made in order to follow normal spectroscopic usage. The HNCA experiment is commonly understood to involve J coupling transfer between the protein backbone N and either $C\alpha_i$ or $C\alpha_{i-1}$. To remain consistent with this usage transfer between atoms named N and CA defaults to ‘Jcoupling’. The solid state experiment known as ‘NCA’, which actually involves a SPECIFIC CP transfer between the N and $C\alpha$, is therefore given the official name N_CA.onebond (whereas the official name for the ‘NCO’ experiment remains simply NCO). Three of the transfer types are new to CCPN: (1) ‘relayed’ refers to intra-residue transfers and replaces the previous ‘TOCSY’ transfer type. (2) ‘through-space’ refers to any through-space transfer, regardless of the mechanism (e.g. TEDOR, cross polarisation, NOE, coupling through hydrogen bonds etc.) and replaces the previous ‘NOESY’ transfer type. (3) ‘relayed-alternate’ is a transfer type which has been created to deal with double-quantum transfers which produce ‘relayed’ magnetisation transfers, but with the added information that $n + 1$ and n bond transfers yield cross peaks of alternate sign.

The CCPN experiment types provide input to assignment and validation routines, but neither data model nor assignment software insists on absolute consistency. A few peaks that are not compatible with the description in the experiment type can be accommodated in the assignment process. In problematic cases it is up to the spectroscopist to choose the experiment description that best fits the data, e.g. by replacing a ‘relayed-alternate’ experiment type by a less restrictive ‘relayed’ or even ‘through-space’ alternative.

The transfer types have been used to build experiment prototypes for the most commonly used solid-state MAS NMR experiments using the CCPN Data Model description. The experiment descriptions are available from the CCPN web site and are distributed as reference information in all CcpNmr Analysis software releases from version 2.2.1 onwards. Table 1 shows a list of the new experiments, together with their new systematic name and a synonym where there are several common names. In addition, there are many other experiments which have originated in solution NMR, but which can also be used by solid-state MAS NMR spectroscopists; these can be found in the complete experiment prototype list in the Supplementary Information (Table S1).

Users who develop new pulse sequences requiring new experiment descriptions can add these using the graphical Experiment Prototypes module within CcpNmr Analysis. It is hoped that users will then share these experiment descriptions with the wider NMR community by making them available for future distributions of CCPN.

Once an experiment has been allocated an experiment prototype within CcpNmr Analysis, the information contained within the prototype is used by the program at several different stages. For instance, in an NCACX spectrum recorded with a long mixing time (N_CA_C.onebond,through-space), the experiment prototype description indicates that the N and C α must derive from the same spin system; when looking at assignment options for a peak the program can filter the assignment options accordingly. Similarly, the descriptions of a DARR and CHHC experiment (C_C.through-space and Ch_hC.through-space, respectively) allow the software to distinguish automatically whether distance restraints need to be generated between carbon atoms (in the case of a DARR spectrum) or their bonded hydrogen atoms (in the case of a CHHC spectrum), despite the fact that the peak lists will in both cases simply consist of pairs of carbon atom chemical shifts.

Isotope labelling schemes

Solid-state MAS NMR studies often benefit from a variety of non-uniform labelling schemes (van Gammeren et al. 2004; Etzkorn et al. 2007; Hiller et al. 2008; Higman et al. 2009; Schneider et al. 2010; Hefke et al. 2010). Selective and extensive isotope labelling with 1,3- ^{13}C or 2- ^{13}C -labelled glucose, for instance, results in few neighbouring atom sites being simultaneously spin-labelled. The advantages of these labelling schemes include the narrowing of signals, a reduction in signal overlap, an increase in observable long-range correlations and the possibility of accurately identifying intermolecular contacts (Hong and Jakes 1999; Castellani et al. 2002; Loquet et al. 2010). However, the labelling schemes are generally not straight forward, especially for amino acids synthesised via the

citric acid cycle for which several different isotopomers are produced. The labelling scheme contains significant amounts of useful information for resonance and distance constraint assignment, but harnessing this information is not trivial. The program SOLARIA was developed to filter for the labelling scheme when assigning distance constraints for iterative automated structure calculations (Fossi et al. 2005) and this approach has also been used in ARIA (Rieping et al. 2007) from version 2.3 onwards. The labelling information can also be useful at many other stages of data analysis.

We have introduced a comprehensive Isotopic Labelling module into the data model and a corresponding graphical interface in CcpNmr Analysis which allows users to specify any type of macromolecular labelling. This will benefit both solid-state MAS and solution NMR applications. The system can describe uniform labelling (e.g. [U- ^{13}C , ^{15}N] labelling), amino-acid specific labelling (e.g. from cell-free protein synthesis or reverse labelling), labelling which results in a mixture of isotopomers for some amino acids (e.g. labelling with [1,3- ^{13}C]-glycerol), or segmental labelling which results in only part of the molecule being labelled (e.g. obtained by using solid-phase peptide synthesis or ligation techniques).

Reference isotopic labelling schemes are provided for many commonly used isotopic labelling strategies including natural abundance labelling, [U- ^{13}C , ^{15}N], [U- ^2H , ^{13}C , ^{15}N], [U- ^{15}N] with [1,3- ^{13}C]-glycerol labelling, SAIL labelling (Kainosho et al. 2006) and others. Copies of these standard schemes can be altered for any amino acid type on an atom by atom basis, or new schemes can be created from scratch. It is possible to introduce multiple isotopomers for any amino acid type and, if known, the relative residue isotopomer populations can be specified (this is required e.g. for the glycerol-based labelling schemes).

In CcpNmr Analysis, a separate dialog box allows users to apply labelling schemes to molecules (in effect describing ‘samples’) and associate these with particular NMR experiments (Fig. 1). Each ‘labelled sample’ may contain one or more ‘isotopic labelling patterns’. In the simplest case, spectra may have been recorded on a sample composed of 100% [U- ^{13}C , ^{15}N]-labelled protein and thus the sample’s molecule is associated with a single labelling pattern (Fig. 1a, b). However, in order to detect intermolecular contacts in an oligomeric protein, half the protein in the sample may have been [U- ^{13}C]- and the other half [U- ^{15}N]-labelled. In this case the molecule will be associated with two labelling patterns in a 1:1 ratio (Fig. 1a, c). Both these scenarios are easily set up using the Isotope Labelling panel (Fig. 1a). If different regions or residues of a molecule have different labelling patterns (e.g. due to labelling via solid-phase peptide synthesis or expressed protein ligation methods), the labelling pattern

Table 1 New solid-state MAS NMR Experiment Prototypes included in CcpNmr Analysis showing for each experiment prototype the systematic name, the maximum possible number of dimensions, a synonym and examples of experiments included in the prototype

Systematic experiment name	Max dim	Experiment synonym ^a	Examples ^b
H	1	¹ H 1D	
C	1	¹³ C 1D	
N	1	¹⁵ N 1D	
F	1	¹⁹ F 1D	
P	1	³¹ P 1D	
CC	2	CC (onebond); CC COSY ^c	DOAM, CMAR, CMR, CTUC
C_C.relayed	2	CC (relayed); CC TOCSY ^c	TOBSY, short mixing time PDSO, DARR, RFDR, PAR
C_C.through-space	2	CC (through-space); CC NOESY ^c	Long mixing time PDSO, DARR, RFDR, PAR
C_C.relayed-alternate	2	CC (relayed-alternate)	
C[DQ(CC)]	2	¹³ C DQ	POST-C7, INADEQUATE etc.
N_N.through-space	2	NN (through-space)	
NC	2	NC (onebond)	Short mixing time TEDOR, PAIN
N_C.through-space	2	NC (through-space)	Long mixing time TEDOR, PAIN
N_CA.onebond	2	NCA	
NCO	2	NCO	
NC_C.relayed	3	NCC (onebond,relayed)	
NC_C.through-space	3	NCC (onebond,through-space)	
N_CA_C.onebond,relayed	3	NCACX (relayed)	
NCO_C.relayed	3	NCOCX (relayed)	
N_CA_C.onebond,through-space	3	NCACX (through-space)	
NCO_C.through-space	3	NCOCX (through-space)	
N_CA_CB.onebond,relayed-alternate	3	NCACB	
N_CA_CB_C.onebond,relayed-alternate,relayed	4	NCACBCX (relayed)	
NCOCA	3	NCOCA	
NCOCA_CB.relayed-alternate	4	NCOCACB	
N_CACO.onebond	3	NCACO	
CA_NCO.onebond	3	CANCO	
CA_NCO_C.onebond,relayed	4	CANCOCX (relayed)	CANCOCA
CA_NCO_C.onebond,through-space	4	CANCOCX (through-space)	
CBCACO	3	CBCACO	
CACBCG	3	CACBCG	
C_C_C.relayed,through-space	3	CCC (relayed,through-space)	
Cali_Cali_C.relayed,through-space	3	CaliCaliC (relayed,through-space)	
Cali_Cali_C.relayed-alternate,relayed	3	CaliCaliC (DREAM,relayed)	
Cali_Cali_C.relayed-alternate,through-space	3	CaliCaliC (DREAM,through-space)	
CaliCali_C.relayed	3	CaliCaliC (onebond,relayed)	
CaliCali_C.through-space	3	CaliCaliC (onebond,through-space)	
CH_HC.through-space	4	CHHC	
CH_HC_C.through-space,through-space	5	CHHCC	
NH_HC.through-space	4	NHHC	
NH_HC_C.through-space,through-space	5	NHHCC	
H[N]_HC.through-space	4	HNHHC	
H_HC.through-space	3	HHC NOESY	
H_HN.through-space	3	HHN NOESY	
HC_C.relayed	3	HCC (relayed); HCC TOCSY	

Table 1 continued

Systematic experiment name	Max dim	Experiment synonym ^a	Examples ^b
HC_C.through-space	3	HCC (through-space)	
HNC	3	HNC (onebond)	
HN_C.through-space	3	HNC (through-space)	
HNCA_C.relayed	4	HNCACX	HNCACB
HNCO_C.relayed	4	HNCOCACX	HNCOCACB

^a In several cases synonyms have been created to reflect the experimental conditions under which an experiment is recorded (e.g. ‘NCACX (relayed)’ would be appropriate for an NCACX recorded with a short mixing time and ‘NCACX (through-space)’ for one recorded with a long mixing time). In other cases synonyms have been created which cover all experiments that are appropriate for that experiment prototype

^b Examples for some of the spectra covered by the experiment prototype

^c In cases where the experiment type is also commonly used in solution NMR, a synonym reflecting the solution NMR nomenclature is included

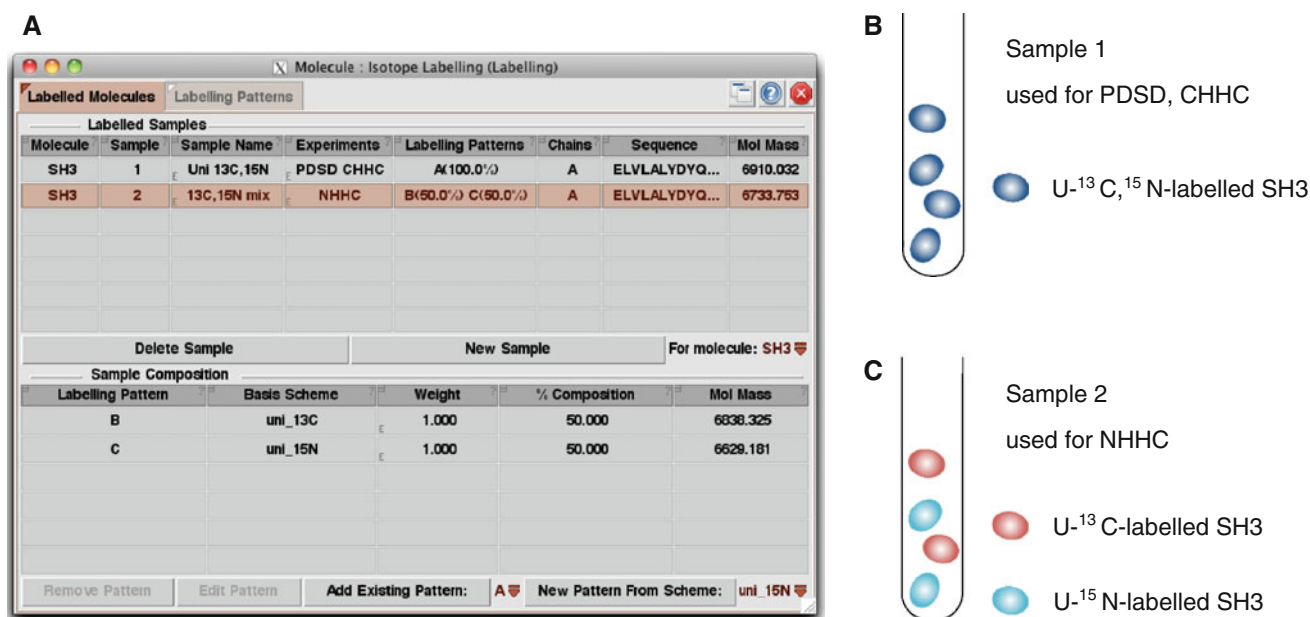


Fig. 1 **a** Isotope Labelling panel showing the way in which a protein sample (in this case an SH3 domain) can be designated as being uniformly [¹³C, ¹⁵N] labelled (*upper panel*, Sample 1) or composed of 50% [U-¹³C] and 50% [U-¹⁵N] labelled protein (*upper panel*, Sample

2 and *lower panel*). The ‘Experiments’ column in the *upper panel* shows which experiments are associated with each sample. **b** Diagram illustrating the composition of Sample 1. **c** Diagram illustrating the composition of Sample 2

for any residue, or group of residues, in the molecule can be modified to that of an alternative reference labelling scheme. Once the labelled samples are linked to experiments, the labelling information is available for data analysis with CcpNmr Analysis (where its use is automatic, but may be disabled), and for database deposition with CcpNmr ECI.

Analysis of solid-state MAS NMR spectra

Double quantum chemical shift axes

Double quantum experiments may be performed or processed in such a way as to result in two single quantum

chemical shift axes or in one single and one double quantum chemical shift axis. In the latter type of spectra two neighbouring atoms A and B with chemical shifts a and b will give rise to peaks at the chemical shift positions $(a, a + b)$ and $(b, a + b)$, respectively (Fig. 2, blue spectrum). Such spectra can provide complementary information to conventional ¹³C–¹³C correlation spectra e.g. by removing spectral overlap where neighbouring nuclei give rise to identical chemical shifts (as may be the case for Ser C α /C β , Met C β /C γ and other pairs of resonances). However, the process of converting the double quantum chemical shifts into the two component chemical shifts usually has to be carried out by hand, and visual comparisons of chemical shifts with other spectra, e.g. using marker lines, is not straight forward. Until

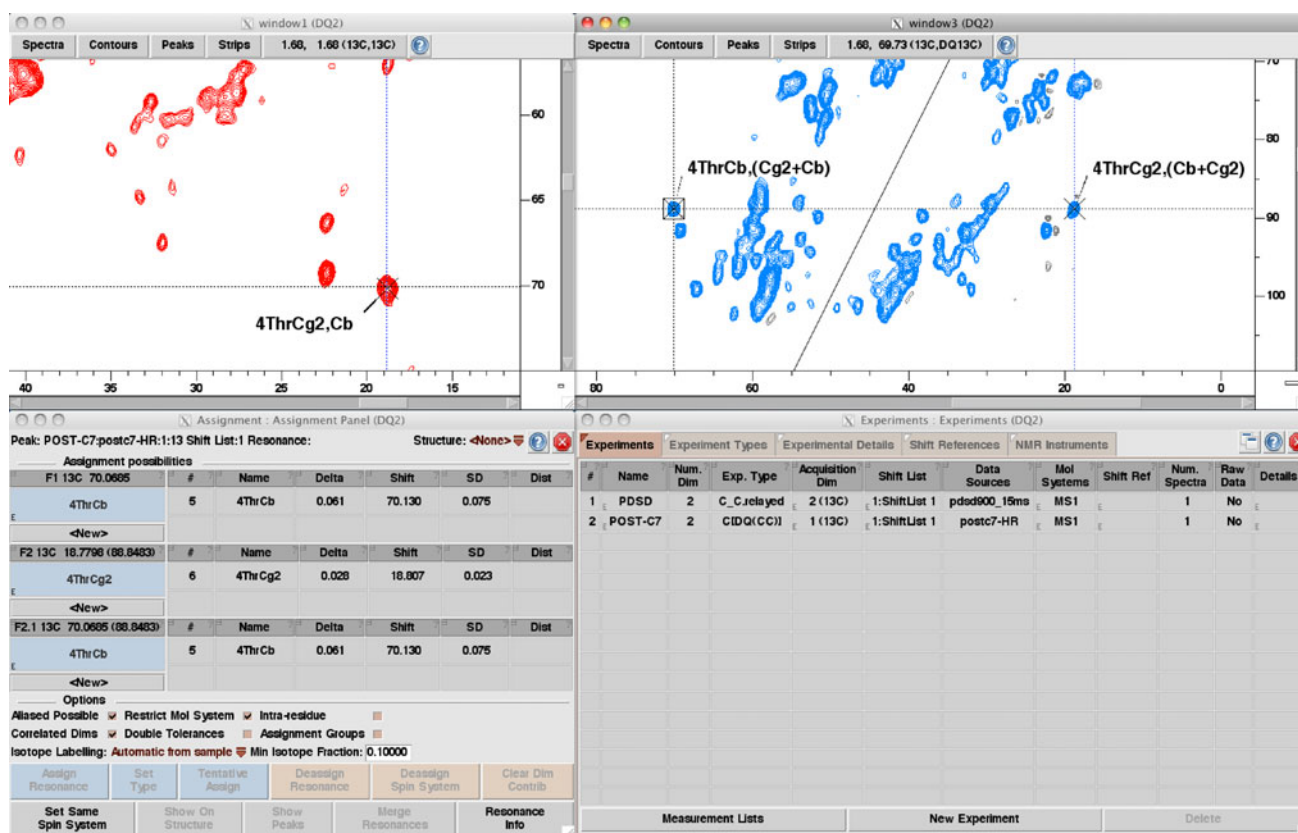


Fig. 2 Upper left panel PDSF spectrum (red) of a small $^{13}\text{C},^{15}\text{N}$ -labelled membrane protein showing a Thr $\text{C}\beta$ - $\text{C}\gamma 2$ cross peak. Upper right panel double quantum POST-C7 spectrum (blue) recorded on the same sample showing the two peaks arising from the Thr $\text{C}\beta$ and Thr $\text{C}\gamma 2$ resonances. A diagonal is visible at $(\delta, 2\delta)$. The left hand peak has been selected (surrounded by black box) and a mark has been placed at this position which results in all the remaining visible lines being drawn at the Thr $\text{C}\beta$ and Thr $\text{C}\gamma 2$ chemical shifts in both spectra and the (Thr

$\text{C}\beta + \text{Thr } \text{C}\gamma 2$) chemical shift along the double quantum axis of the POST-C7 spectrum. Lower left panel The assignment panel for the selected peak shows the assignments and assignment options in both dimensions. The chemical shift in the double quantum dimension (88.85 ppm) has been split into the two component chemical shifts (70.07 ppm and 18.78 ppm) corresponding to the Thr $\text{C}\beta$ and Thr $\text{C}\gamma 2$ resonances. Lower right panel Experiments panel showing the Experiment Prototypes selected for the PDSD (15 ms mixing time) and POST-C7 spectra

recently CcpNmr Analysis was limited to chemical shift axis types corresponding to a single nucleus. We have now added double quantum chemical shift axes to the program (Fig. 2, lower right hand panel). These generate a modified diagonal at $(\delta, 2\delta)$ as well as modified cross-hairs at positions $(\delta 1 + \delta 2)$ along the indirect double quantum dimension and $\delta 1$ and $\delta 2$ in the direct dimension. The $\delta 1$ and $\delta 2$ positions are reflected in the crosshairs of other spectral windows, improving the way in which peak positions can be compared between double and single quantum experiments (Fig. 2). When assigning the double quantum chemical shift, the value is automatically deconvoluted into the two constituent chemical shifts. Assignment options are then presented and can be chosen in the conventional manner (Fig. 2, lower left hand panel).

Spinning side bands

Solid-state MAS NMR spectra contain peaks arising from spinning side bands as well as from the main centre band. Usually the side-band peaks are not used during spectral analysis but it is important to identify them in order to prevent the misinterpretation of side-band peaks as centre-band peaks. Several new features have been added to CcpNmr Analysis in order to aid the identification of side-band peaks. Once the spinning frequency has been entered for an experiment, for instance, the side-band diagonals are marked with dashed lines (Fig. 3). Furthermore, the ability to draw lines at particular ppm positions which are propagated through all spectral windows (using 'm' for cross marks, 'h' for horizontal rulers and 'v' for vertical rulers) has been

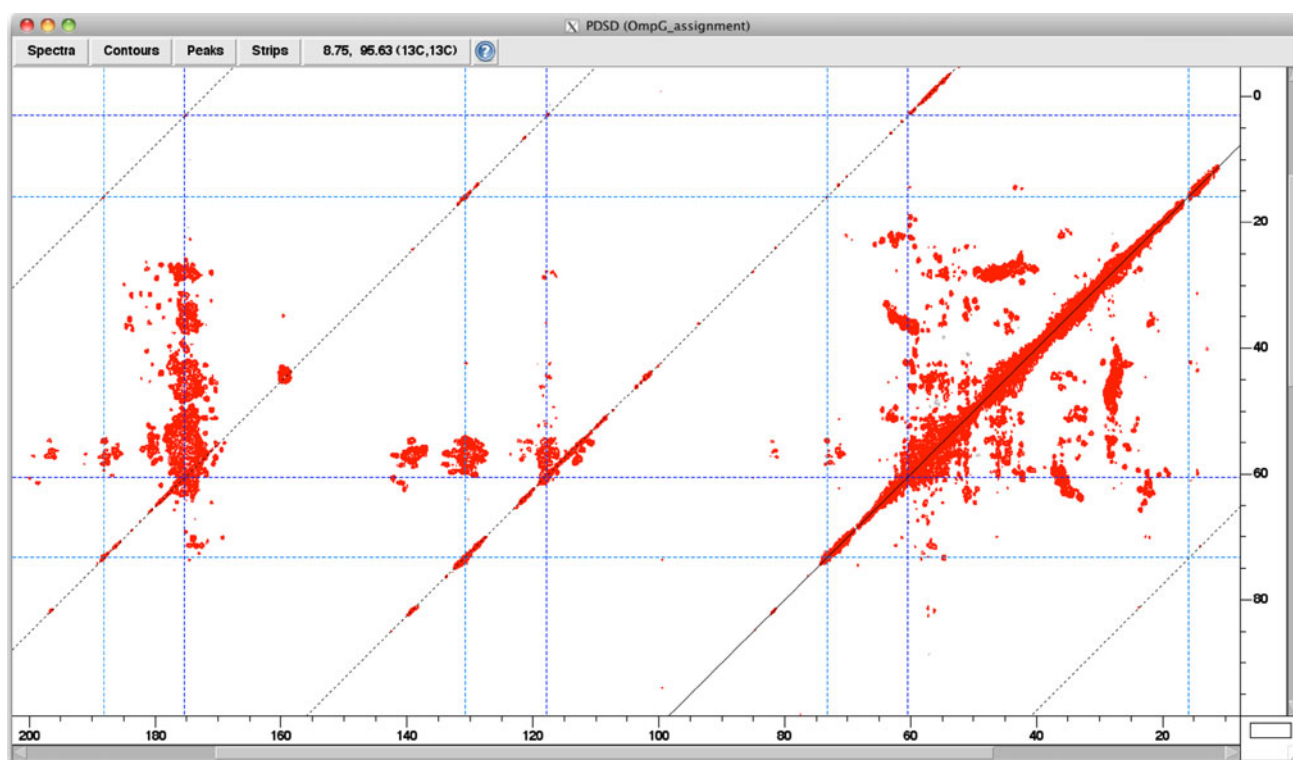


Fig. 3 DARR spectrum of [2- ^{13}C]-glycerol labelled Outer membrane protein G (OmpG) recorded at 900 MHz using 13 kHz MAS and a mixing time of 200 ms. Once the spinning frequency has been entered, the diagonal is automatically repeated at 13 kHz intervals using *dashed lines*, thus highlighting side-band diagonals. The *dark blue rulers* drawn (using the 'V' key command) at 118 ppm and

automatically repeated at 13 kHz intervals indicate that the peaks at 118 ppm are likely to be side bands of the carbonyl peaks at 176 ppm. The *light blue rulers* drawn at 73 ppm and automatically repeated at 13 kHz intervals indicate that these peaks are likely to be side bands of the aromatic signals at 132 ppm and not inter-residue cross peaks involving Thr $C\beta$ resonances

extended to draw cross marks and rulers at all side-band positions relative to the current mouse position (using 'M' for cross-marks, 'H' for horizontal rulers and 'V' for vertical rulers). This aids the identification of side-band peaks (Fig. 3). A module allowing the explicit linking of two peaks as being the centre- and side-band peaks of the same resonance is beyond the scope of the current work. However, this will become possible when a general module for peak components (e.g. for peaks split by couplings) is incorporated into CcpNmr Analysis. Meanwhile, it is possible to label side-band peaks using the peak's 'Details' field or to assign the peak a low 'Merit' score. In this way, peaks can be visually highlighted with a symbol (e.g. '*') and automatically excluded when generating distance restraints or performing other analyses.

Isotope abundance and correlation filtering

CcpNmr Analysis (and ARIA to which it is has been coupled) makes use of the isotope labelling information associated with a particular NMR experiment in order to generate label-aware assignments and distance restraints

from through-space correlations. The labelling information is specified simply in the graphical interface by allocation of an experiment to a 'labelled sample' (with isotopic labels derived from a standard scheme, a modified scheme or a mixture of different schemes, see above). The abundances of magnetically active nuclei that give rise to a peak are then automatically calculated according to the experiment from which the peak derives. Any operation that requires the coincidence of two magnetically active nuclei is aware of both the average isotopic incorporation at each atom site and the proportion of situations where both sites are labelled at the same time; for example $C\alpha$ and $C\beta$ atoms in a given residue may both be labelled, but never at the same time, within the same isotopomer. The incorporation of isotopic labelling information allows easier assignment within CcpNmr Analysis since inappropriate resonance assignment possibilities are removed. Importantly, when generating distance restraints by matching chemical shifts to peak positions, ambiguity is reduced. Knowledge of the particular isotopic abundance also allows for a calibration of through-space peak intensities, such that the underlying intensity of the peak, as if it were 100% spin labelled, can be calculated prior to the

application of an intensity-to-distance relation. This method is also used for NMR experiments where the through-space correlated resonances are not observed directly, for example in a CHHC experiment where only ^{13}C resonances are observed, but the actual magnetisation transfer (and therefore the generated distance restraints) are between the unobserved hydrogen atoms. In this case the labelling of both the carbon and hydrogen nuclei are considered.

Synthetic peak lists

The use of the experiment types and labelling schemes becomes very powerful when combined with the ability of CcpNmr Analysis to create synthetic peak lists. One possibility, for instance, is to import chemical shift lists from the BMRB or shift prediction programs such as SHIFTX (Neal et al. 2003) and SPARTA (Shen and Bax 2007) using the CcpNmr Format Converter. These shift lists can then be used to create peaks for any type of spectrum with any type of isotopic labelling in order to compare predicted and actual peak positions. Alternatively, once a protein has been assigned, it is possible to create a synthetic peak list which contains only intra-residue peaks or peaks between neighbouring residues. This allows rapid identification of long-range peaks required for structure calculations in spectra containing both long- and short-range correlations. Such peak lists can also be used as an efficient tool to check whether all assignments have been made in a spectrum.

Additional features

In addition to the features described above which have been developed specifically for solid-state MAS NMR data analysis, several other features in CcpNmr Analysis are particularly useful in the analysis of solid-state MAS NMR data. The easy way in which 2D and 3D spectra can be overlaid is well suited to making comparisons between spectra recorded with different mixing times or on samples with different labelling schemes. The option to arrange strips horizontally rather than vertically suits many solid-state MAS NMR spectra, such as NCACX or NCOCX spectra, whose peaks conventionally occur in horizontal rather than vertical strips. The ability to make tentative assignments is a helpful way of keeping track electronically of multiple assignment possibilities in ambiguous situations. Many other routines within CcpNmr Analysis have also been updated and tested to ensure their compatibility with solid-state MAS NMR data, enabling the easy display of build-up curves, the creation of distance restraints involving carbon or nitrogen atoms or assignment statistics in the quality reports which exclude hydrogen atoms.

Conclusions

In conclusion, we present additions to the CCPN data model and new features within the CcpNmr Analysis program which significantly improve solid-state MAS NMR data analysis and streamline the resonance assignment and structure calculation processes. The structure calculation program, ARIA (Rieping et al. 2007), for example, uses the CCPN data model and movement between ARIA and CcpNmr Analysis is almost seamless. This is now also the case when using solid-state MAS NMR data. The information about labelling schemes and experiment types which has been entered into CcpNmr Analysis can be accessed by ARIA and exploited during its structure calculation routines. The expanded data model also provides a platform for the development of new, interconnected software geared towards solid-state MAS NMR data analysis, e.g. automated assignment programs (Moseley et al. 2010; Tycko and Hu 2010) or line-shape fitting routines for the extraction of dipolar couplings to produce angular structure restraints (Franks et al. 2008).

Several of the features presented here also provide improvements for solution NMR data analysis. The double quantum axes, for instance, can be used to analyse reduced dimensionality experiments and filtering according to isotope labelling schemes is useful when using SAIL-labelled proteins (Kainosho et al. 2006). Overall, the expanded data model and new features in CcpNmr Analysis provide an important contribution towards the analysis of protein solid-state MAS NMR data to match the rapid methodological developments that have occurred over the past decade.

Versions 2.2.1 and above of CcpNmr Analysis contain all the described solid-state MAS NMR extensions, together with the latest CCPN data model. APIs for Python, C and Java languages are available, storing data in the form of the updated data model, which includes the new experiment descriptions and isotope labelling schemes. Subsequent CCPN software releases will include further experiment descriptions and labelling schemes provided by the user community. The software can be downloaded free of charge for non-profit institutions at <http://www.ccpn.ac.uk>. Detailed documentation is available at the CCPN web site via <http://www.ccpn.ac.uk/documentation/> and <http://www.ccpn.ac.uk/wiki/>.

Acknowledgments We thank Johanna Becker-Baldus, Patrick van der Wel and Trent Franks for helpful discussions. We acknowledge the Deutsche Forschungsgemeinschaft and the Biotechnology and Biological Sciences Research Council (UK) for funding.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Andronesi OC, Becker S, Seidel K, Heise H, Young HS, Baldus M (2005) Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *J Am Chem Soc* 127:12965–12974
- Bennett AE, Ok JH, Griffin RG, Vega S (1992) Chemical-shift correlation spectroscopy in rotating solids—radio frequency-driven dipolar recoupling and longitudinal exchange. *J Chem Phys* 96:8624–8627
- Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins Struct Func Bioinformatics* 66:778–795
- Brünger AT (1992) XPLOR version 3.1: a system for X-ray crystallography and NMR. Yale University Press, New Haven
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Gross-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst D* 54:905–921
- Cady SD, Schmidt-Rohr K, Wang J, Soto CS, DeGrado WF, Hong M (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* 463:689–692
- 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
- Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Res* 202:223–233
- De Paëpe G, Lewandowski JR, Loquet A, Böckmann A (2008) Proton assisted recoupling and protein structure determination. *J Chem Phys* 129:245101
- 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 Edit* 46:459–462
- Ferguson N, Becker J, Tidow H, Tremmel S, Sharpe TD, Krause G, Flinders J, Petrovich M, Berriman J, Oschkinat H, Fersht AR (2006) General structural motifs of amyloid protofilaments. *Proc Natl Acad Sci USA* 103:16248–16253
- Fogh RH, Vranken WF, Boucher W, Stevens TJ, Laue ED (2006) A nomenclature and data model to describe NMR experiments. *J Biomol NMR* 36:147–155
- 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 Edit* 44:6151–6154
- Franks WT, Wylie BJ, Schmidt HLF, Nieuwkoop AJ, Mayrhofer RM, Shah GJ, Graesser DT, Rienstra CM (2008) Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR. *Proc Natl Acad Sci USA* 105:4621–4626
- Goddard TD, Kneller DG. SPARKY 3. University of California, San Francisco. <http://www.cgl.ucsf.edu/home/sparky/>
- Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program D. *J Mol Biol* 273:283–298
- Hefke F, Bagaria A, Reckel S, Ullrich S, Dötsch V, Glaubitz C and Güntert P (2010) Optimization of amino acid type-specific ¹³C and ¹⁵N labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. *J Biomol NMR*. doi:10.1007/s10858-010-9462-4
- Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24(3):171–189
- Higman VA, Flinders J, Hiller M, Jehle S, Markovic S, Fiedler S, van Rossum BJ, Oschkinat H (2009) Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively C-13-labelled proteins. *J Biomol NMR* 44:245–260
- Hiller M, Higman VA, Jehle S, van Rossum BJ, Kühlbrandt W, Oschkinat H (2008) 2, 3-C-¹³-labeling of aromatic residues—getting a head start in the magic-angle-spinning NMR assignment of membrane proteins. *J Am Chem Soc* 130:408–409
- 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
- Hoof RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272
- Huang YJ, Moseley HN, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperki T, Montelione GT (2005) An integrated platform for automated analysis of protein NMR structures. *Methods Enzymol* 394:111–141
- Jaroniec CP, MacPhee CE, Bajaj VS, McMahon MT, Dobson CM, Griffin RG (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci USA* 101:711–716
- Jehle S, Rajagopal P, Bardiaux B, Markovic S, Kühne R, Stout JR, Higman VA, Klevit RE, van Rossum BJ, Oschkinat H (2010) Solid-state NMR and SAXS studies provide a structural basis for the activation of alpha B-crystallin oligomers. *Nature Struct Mol Biol* 17:1027–1042
- Johnson BA, Blevins RA (1994) NMR view: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614
- Jung YS, Zweckstetter M (2004) Mars—robust automatic backbone assignment of proteins. *J Biomol NMR* 30:11–23
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57
- Lemak A, Gutmanas A, Chitayat S, Karra M, Farès C, Sunnerhagen M, Arrowsmith CH (2011) A novel strategy for NMR resonance assignment and protein structure determination. *J Biomol NMR* 49:27–38
- Loquet A, Bardiaux B, Gardiennet C, Blanchet C, Baldus M, Nilges M, Malliavin T, Boeckmann 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, Giller K, Becker S, Lange A (2010) Supramolecular interactions probed by C-13-C-13 solid-state NMR spectroscopy. *J Am Chem Soc* 132:15164–15166
- Moseley H, Sperling L, Rienstra C (2010) Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of β 1 immunoglobulin binding domain of protein G (GB1). *J Biomol NMR* 48:123–128
- Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts. *J Biomol NMR* 26:215–240
- Petkova AT, Buntkowsky G, Dyda F, Leapman RD, Yau WM, Tycko R (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J Mol Biol* 335:247–260
- Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23:381–382

- Schneider R, Etzkorn M, Giller K, Daebel V, Eisfeld J, Zweckstetter M, Griesinger C, Becker S, Lange A (2010) The native conformation of the human VDAC1 N terminus. *Angew Chem Int Edit* 49:1882–1885
- Schwieters CD, Kuszewski J, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Res* 160:65–73
- Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR* 38:289–302
- Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223
- Takegoshi K, Nakamura S, Terao T (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem Phys Lett* 344:631–637
- Tycko R, Hu KN (2010) A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning. *J Magn Res* 205:304–314
- van Gammeren AJ, Hulsbergen FB, Hollander JG, de Groot HJM (2004) Biosynthetic site-specific C-13 labeling of the light-harvesting 2 protein complex: a model for solid state NMR structure determination of transmembrane proteins. *J Biomol NMR* 30:267–274
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct Funct Bioinformatics* 59:687–696
- Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH (2008) Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* 319:1523–1526
- 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