

## Peakr: simulating solid-state NMR spectra of proteins

Robert Schneider<sup>1,2,\*</sup>, Florian Odronitz<sup>1,†</sup>, Björn Hammesfahr<sup>1,†</sup>, Marcel Hellkamp<sup>1</sup> and Martin Kollmar<sup>1,\*</sup>

<sup>1</sup>Department of NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany and <sup>2</sup>Protein Dynamics and Flexibility by NMR Group, Institut de Biologie Structurale J.P. Ebel, 41 rue Jules Horowitz, F-38027 Grenoble Cedex 1, France

Associate Editor: Anna Tramontano

### ABSTRACT

**Motivation:** When analyzing solid-state nuclear magnetic resonance (NMR) spectra of proteins, assignment of resonances to nuclei and derivation of restraints for 3D structure calculations are challenging and time-consuming processes. Simulated spectra that have been calculated based on, for example, chemical shift predictions and structural models can be of considerable help. Existing solutions are typically limited in the type of experiment they can consider and difficult to adapt to different settings.

**Results:** Here, we present Peakr, a software to simulate solid-state NMR spectra of proteins. It can generate simulated spectra based on numerous common types of internuclear correlations relevant for assignment and structure elucidation, can compare simulated and experimental spectra and produces lists and visualizations useful for analyzing measured spectra. Compared with other solutions, it is fast, versatile and user friendly.

**Availability and implementation:** Peakr is maintained under the GPL license and can be accessed at <http://www.peakr.org>. The source code can be obtained on request from the authors.

**Contact:** robert.schneider@ibs.fr or mako@nmr.mpibpc.mpg.de

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on July 24, 2012; revised on January 15, 2013; accepted on February 10, 2013

### 1 INTRODUCTION

In recent years, solid-state nuclear magnetic resonance (NMR) has made significant progress in studying structure and function of biomolecules such as membrane proteins and protein fibrils (for reviews, see McDermott, 2009; Renault *et al.*, 2010; Judge and Watts, 2011; Tycko, 2011). However, especially for larger proteins, resonance assignment and determination of restraints for 3D structure calculations are still difficult and time-consuming processes owing to often limited spectral resolution and chemical shift ambiguity, as well as complex relationships

between internuclear distances and signal intensities. These problems especially apply to through-space correlations that cannot be traced along the chemical bond network (Manolikas *et al.*, 2008).

For the assignment of resonances and the extraction of restraints, it has been proven helpful to have simulated spectra (sometimes referred to as spectrum ‘predictions’) at hand. Simulated spectra can be calculated based on amino acid sequences, known or modeled 3D structures, chemical shift assignments or predictions from tools such as SHIFTX (Neal *et al.*, 2003) and the type of correlation probed in the respective experiment. This way, cross-peak assignments can be suggested, and, for example, it can be investigated whether an experimental spectrum can be explained by a given structural model (Wasmer *et al.*, 2008; Schneider, *et al.*, 2010b). Simulated spectra can also be used in an iterative process of obtaining shift assignments and refining the molecular structure at the same time (Matsuki *et al.*, 2007).

Existing software for the simulation of NMR spectra is usually tailored to calculating spectra of small molecules in solution (Golotvin *et al.*, 2007; Binev and Aires-de-Sousa, 2004; ACD/Labs NMR Predictors, 2007) or to specifically simulate NOESY spectra in the context of NMR protein structure determination (Gronwald and Kalbitzer, 2004, and references therein). Some NMR data analysis software packages such as Sparky (Goddard and Kneller), NMRPipe (Delaglio *et al.*, 1995) and CcpNmr (Vranken *et al.*, 2005) contain routines for simulation of protein NMR spectra, but these are often not straightforward to use or do not work in a stand-alone manner. Moreover, with the exception of recent additions to CcpNmr (Stevens *et al.*, 2011), they are usually more adapted to solution-state NMR correlation types. Thus far, spectrum simulation has typically been carried out using general-purpose tools like spreadsheet software or custom-made programs limited in flexibility and usability.

Thus, it would be desirable to be able to simulate a wide range of solid-state NMR experiments commonly used for resonance assignment and structure elucidation with a single software tool that is flexible enough to swiftly handle changes in input data such as chemical shift values, structural models or labeling patterns. We implemented Peakr, a software package that fulfills these requirements. Based on chemical shifts provided by the user or predicted by built-in third party software tools, spectra for 2D (<sup>15</sup>N, <sup>13</sup>C) and (<sup>13</sup>C, <sup>13</sup>C) intra- and inter-residue as well as through-space correlations can be computed quickly and can easily be adapted to specific needs. Using the calculated spectra,

\*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Presented in part at the 23rd International Conference on Magnetic Resonance in Biological Systems (ICMRBS), August 24–29, 2008, San Diego, CA, USA, and in Odronitz, F., Genomics and phylogeny of motor proteins: tools and analyses, PhD thesis, University of Göttingen, Germany, 2008 (<http://webdoc.sub.gwdg.de/diss/2008/odronitz/>).

visual and numerical comparisons between simulated and measured data are possible. Peakr is available through a web interface ([www.peakr.org](http://www.peakr.org)).

## 2 METHODS

The higher abstracted parts of Peakr are implemented in the object-oriented programming language Ruby (<http://www.ruby-lang.org>) using the BioRuby library (Goto *et al.*, 2010). The more data-intensive bookkeeping is done using a PostgreSQL database (<http://www.postgresql.org>). The Ruby on Rails framework (<http://rubyonrails.org>) is used for the web application, which drives the web interface. Calculated and experimental spectra are visualized by a Peakr-plugin to the PyBioMaps library (<http://http://pypi.python.org/pypi/PyBioMaps/>). The workflow of the software is outlined in Figure 1. Briefly, chemical shifts supplied by the user or predicted using one of several shift prediction programs (see below) are assigned to the nuclei of a user-provided protein sequence. Based on these shifts and an optionally provided protein structure, various 2D spectra can be calculated. For visualization, either all or a subset of the residues and nuclei of the protein can be selected. Simulated spectra can be superimposed on and compared with experimental spectra.

### 2.1 Protein sequences

In Peakr, proteins and their constituent amino acids are represented by protein object models. A protein is either created from a user-provided protein sequence or from a Protein Data Bank (PDB) structure file (Berman *et al.*, 2007).

### 2.2 Chemical shifts

Chemical shifts are added to the nuclei in the amino acids of a protein from user-provided shift lists, computer-generated shift estimations, database values or a combination of these methods. User-provided lists of shifts are currently accepted in Sparky (Goddard and Kneller), SHIFTX (Neal *et al.*, 2003) and comma-separated value formats. Output from other software can be converted to one of these formats using, for example, WeNMR (Vranken *et al.*, 2005). If a PDB file is provided, chemical shifts can directly be estimated using either SHIFTX (Neal *et al.*, 2003), SHIFTX2 (Han *et al.*, 2011), SPARTA (Shen and Bax, 2007), SPARTA+ (Shen and Bax, 2010), SHIFTS (Xu and Case, 2001), shAIC (Nielsen *et al.*, 2012) or CamShift (Kohlhoff *et al.*, 2009). Alternatively, average chemical shifts from the Biological Magnetic Resonance Database (BMRB, <http://www.bmrwisc.edu>; Markley *et al.*, 2008; Ulrich *et al.*, 2007) can be assigned to nuclei. These methods can also be combined, in which case user-provided data from a chemical shift list are given priority over chemical shift values predicted by third-party software, which in turn is prioritized over average database values. Thus, for example, user-provided shifts from a Sparky-format list are used where available; nuclei unassigned in this list the chemical shifts of which are predicted by SHIFTX are assigned SHIFTX prediction values; finally, nuclei unassigned in the Sparky-format list the chemical shifts of which are not predicted by SHIFTX, such as side-chain carbons beyond CB, are assigned BMRB database values. Each nucleus can have multiple shifts to account for possible cases of conformational polymorphism sometimes seen in solid-state preparations of proteins (Seidel *et al.*, 2005).

### 2.3 Conformations

Conformations represent a set of coordinates of the atoms of a given protein. Each protein can have several conformations. This way, an ensemble of structures or different conformations of the same protein under different experimental conditions can be represented. Conformations are obtained from models as included in PDB files. Protons have to be already present in the PDB file if they are required for the spectrum to be

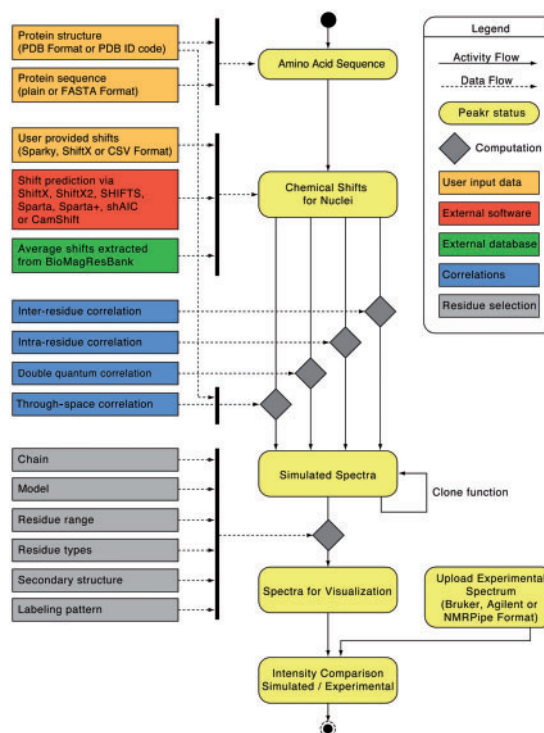


Fig. 1. Workflow of the Peakr web application

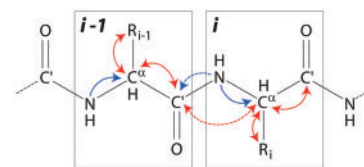


Fig. 2. Scheme of the protein backbone. Two amino acid residues  $i$  and  $i-1$  with side-chains symbolized by  $R_i$  and  $R_{i-1}$  are shown. Possible (solid-state) NMR correlations for sequential resonance assignment are indicated by arrows. Red: ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations; blue: ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) correlations; solid lines: denote intra-residue transfer; dashed lines: inter-residue transfer

simulated (see through-space correlations below). Protons can be added to a structure using standard software such as WHATIF (Vriend, 1990) or PyMOL (the PyMOL Molecular Graphics System).

### 2.4 Correlations

Correlations represent the type of experiment conducted to obtain a specific spectrum (see Fig. 2 for typical correlations used in resonance assignment). They are applied to all or selected residues of the protein sequence to yield a list of cross-peaks. The following six types of correlations are available.

**2.4.1 Intra-residue ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations** For defining ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations within residues, the user can select which nuclei should be considered, depending on their distance from the protein backbone (e.g. all carbons or only CO, CA and CB nuclei). It can also be specified by how many bonds two carbon nuclei may be separated to be included in the simulated spectrum (yielding, for example, only one-bond correlations

or correlations between all carbons separated by up to three bonds). This way, the user can select correlations of interest and avoid overcrowding of the simulated spectrum with peaks that might not be present in a measured spectrum owing to, for example, increased side-chain mobility or short mixing time.

**2.4.2 Double quantum correlations** ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) double quantum correlations represent intra-residue correlations seen in 2D double quantum–single quantum (INADEQUATE-type) correlation spectra (Bax *et al.*, 1980; Menger *et al.*, 1986), where the resonance frequency of a cross-peak in the indirect dimension corresponds to the sum of the chemical shifts of the two interacting nuclei. Double quantum correlations are created in the same way as regular intra-residue correlations; however, only one-bond correlations are considered, as only these are normally observed in experimental spectra.

**2.4.3 Inter-residue ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations** For ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations between neighboring residues, again, it can be specified which nuclei to consider, depending on their distance from the backbone. In addition, the desired unique or maximum residue number difference for nuclei to be correlated can be chosen.

**2.4.4 ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) correlations** ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) correlations can be defined as intra-residue N(i)–CA(i)–CX(i) or sequential N(i)–CO(i-1)–CX(i-1) correlations (with CX representing any other carbon nucleus in the respective residue). It can be specified whether all or only a subset of the  $^{13}\text{C}$  nuclei should be included in the simulation, depending on their distance from the backbone (yielding, for example, only N(i)–CA(i) or also N(i)–CA(i)–CB(i) correlations, etc.).

**2.4.5 Through-space correlations** Through-space correlations are created by specifying a set of residues, a set of conformations with atom coordinates and a pairwise distance cutoff up to which correlations should be taken into account. This type of correlation can act in different modes, allowing either direct distances between heteronuclei to be considered (yielding C–C or N–C through-space correlations) or the distances between protons directly attached to heteronuclei, yielding NHHC and CHHC correlations (Lange *et al.*, 2002; 2003). A minimum distance threshold and a minimal residue number difference can also be specified.

Peakr can also simulate intermolecular through-space correlation spectra if a PDB file with multiple protein chains is provided. For example, symmetry equivalents can be generated from known crystal structures using programs such as SwissPDBViewer (Guex and Peitsch, 1997) or PyMol. These can then be analyzed as different chains in Peakr. In this way, intermolecular cross-peaks arising because of crystal packing interactions can be identified. In the context of multimeric proteins or protein fibrils, the simulation of intermolecular correlations can be particularly useful (Wasmer *et al.*, 2008). Currently, the simulation of intermolecular through-space correlations requires the presence of two or more copies of the same protein in the PDB file; however, the generalization to hetero-multimeric protein complexes could be implemented as well.

## 2.5 Calculated spectra

Spectra represent a set of cross-peaks that are generated by applying a correlation object to a protein, using the experimental or estimated chemical shifts assigned to its nuclei as discussed in Section 2.2. It is worth noting that the resultant simulated spectrum is not a ‘prediction’ in the strict sense of the word. All internuclear correlations possible under the selected correlation type, subject to further selections (see below), give rise to cross-peaks in a Peakr-simulated spectrum, as long as chemical shift values have been assigned to the corresponding nuclei in the underlying protein object. In this sense, Peakr-simulated spectra are idealized representations of which peaks will be present in an experimental spectrum if all underlying data (chemical shift values, protein structure, etc.) are

correct, the experiment exhaustively and exclusively probes the desired internuclear correlations and no further effects affecting cross-peak appearance (signal to noise, local mobility, etc.) are present. It is thus up to the user to decide whether any differences between a simulated and an experimental spectrum are significant and what their origin might be (see Section 3 for an example).

By default, spectra are calculated for all amino acids in the protein. If a provided PDB file contains several chains and/or models, calculated spectra can be displayed for each of these separately. In addition, the user can select subsets of residues to be displayed in a spectrum. One option is to select a certain sequence range based on residue number. Another possibility is to select residues within particular secondary structure types. If a PDB file is provided, secondary structure is assigned to individual residues using Stride (Frishman and Argos, 1995); otherwise, it is predicted from the sequence using PsiPred v3.0 (Jones, 1999). Furthermore, any subset of amino acid types can be selected. This option can be used to simulate spectra of proteins that were expressed using forward or reverse labeling of certain amino acids (Vuister *et al.*, 1994; Heise *et al.*, 2005).

In addition, the user can select one of several more complex isotope-labeling schemes. In the current implementation, Peakr provides  $^{13}\text{C}$ -labeling patterns as obtained from protein expression using 1,3- $^{13}\text{C}$ -glycerol, 2- $^{13}\text{C}$ -glycerol, 1- $^{13}\text{C}$ -glucose and 2- $^{13}\text{C}$ -glucose as sole carbon sources (Hong, 1999; Castellani *et al.*, 2002; Lundström *et al.*, 2007). The probabilities of individual carbon nuclei to be isotope labeled as listed in these publications are translated into opacity values in Peakr’s spectrum display. Thus, a correlation between two nuclei that are less likely to be both  $^{13}\text{C}$  labeled in the selected labeling scheme appears correspondingly less intense in the simulated spectrum.

It is particularly helpful for analysis and interpretation of spectra if subsets of the same protein sequence are displayed at the same time, but with different colors or markers. Therefore, we provide a ‘Clone’ button to copy every spectrum as often as needed. For each of these identical spectra, the user can then, for example, select different sets of residues and update the displayed spectra accordingly.

## 2.6 Experimental spectra

Processed experimental spectra can be read in and displayed, alone or overlaid with simulated spectra. Peakr accepts processed spectra in Bruker (XWinNMR or Topspin; Bruker Biospin, Karlsruhe, Germany), Varian/Agilent (Vnmr or VnmrJ; Agilent Technologies, Santa Clara, CA) and NMRPipe (Delaglio *et al.*, 1995) formats.

## 2.7 Output

**2.7.1 Data storage** The generated simulated and experimental spectra can be downloaded as file archives referenced by a checksum string for further external investigation. Unmodified file archives can again be uploaded to Peakr for further analysis, including the generation of additional simulated spectra based on the previous settings. The individual checksum provides for security of the user’s data. In addition, all data are automatically deleted from the server 24 hours after creation, to avoid long-term storage of potentially sensitive research data.

**2.7.2 Lists** The cross-peak lists of simulated spectra can be retrieved as tab-delimited files that can be directly read into the Sparky program. When comparing with an experimental spectrum, lists are generated that contain the intensity of the measured spectrum at the positions of the simulated cross-peaks or in a defined region around them. This provides for a straightforward numerical comparison between simulation and experiment, which can be useful for model validation or generation of restraint lists for structure calculation. These peak lists are searchable, and sub-selections of peaks can be made based on, for example, spectral intensity or interatomic distance. Then, the spectrum display can be modified to show only the selected peaks.

**2.7.3 Graphics** Peakr generates spectra with cross-peaks as PNG files that can be zoomed and browsed, providing the functionality known from applications like Google Maps. Visualized spectra (covering all simulated peaks or a region selected by the user) can be downloaded in PNG, PDF or SVG format. Simulated and experimental spectra can be combined in one plot. Cross-peaks from different spectra are distinguished by color. If a specific  $^{13}\text{C}$ -labeling scheme is chosen, the probabilities of the correlations are displayed as opacities. Tooltips on every peak indicate the contributing nuclei and corresponding chemical shifts.

### 3 RESULTS AND DISCUSSION

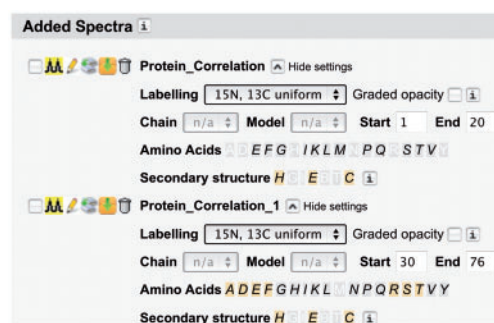
Peakr is available via a web interface (Fig. 3). The workflow has been designed to provide the highest possible flexibility in terms of correlation types as well as protein regions and conformations that can be chosen for comparison and analysis. The user can provide a protein sequence or a structural model; can choose between providing chemical shift lists, predicting shifts using different third-party software or combining these data; and can select various types of correlations that are used in experimental setups. During spectrum simulation, Peakr first calculates correlations for all residues. Subsequently, any combination of residues can be selected for display, also from different chains or models that may be present in a PDB file (Fig. 4). Here, the 'clone' function provides an easy way to visualize various selections of residues and nuclei while using the same settings for shifts and correlations.

Typical spectrum calculations with Peakr are relatively fast, on the order of a few seconds to about half a minute for intra- and inter-residue correlation spectra for small- to medium-sized proteins used in NMR studies (<300 residues, Intel Xeon E5405 2.0 GHz, Supplementary Data). Through-space correlation spectra are more computationally intensive and can take a minute or more to calculate for proteins of this size, depending on the upper internuclear distance threshold chosen. However, these values include the time spent by the interfaced third-party software tools to calculate chemical shift predictions, as well as the graphical output generation. Also, given the ease of use of Peakr, the times mentioned here nearly correspond to the total time to arrive at a simulated spectrum and will thus nevertheless be considerably shorter than with any other existing or custom-written method.

As an example for using Peakr, we demonstrate the simulation of intra-residue ( $^{13}\text{C},^{13}\text{C}$ ) correlation spectra of solid ubiquitin and their comparison with experimental data. The experimental dataset consists of a ( $^{13}\text{C},^{13}\text{C}$ ) correlation spectrum of microcrystalline uniformly [ $^{13}\text{C},^{15}\text{N}$ ] isotope-labeled ubiquitin prepared as described (Seidel *et al.*, 2005). It was recorded on a 700 MHz spectrometer (Bruker Biospin, Karlsruhe, Germany) using 7.8 ms of dipolar-assisted rotational resonance (DARR)  $^{13}\text{C}$ - $^{13}\text{C}$  mixing (Takegoshi *et al.*, 2001). Simulated spectra were generated in Peakr based on the ubiquitin amino acid sequence in plain one-letter code format and using chemical shift assignments for this solid-phase ubiquitin preparation as reported (Seidel *et al.*, 2005) in Sparky list format. We generated three different ( $^{13}\text{C},^{13}\text{C}$ ) spectrum simulations based on these data using Peakr's intra-residue ( $^{13}\text{C},^{13}\text{C}$ ) correlation option, including side-chain nuclei up to CD and allowing for one-



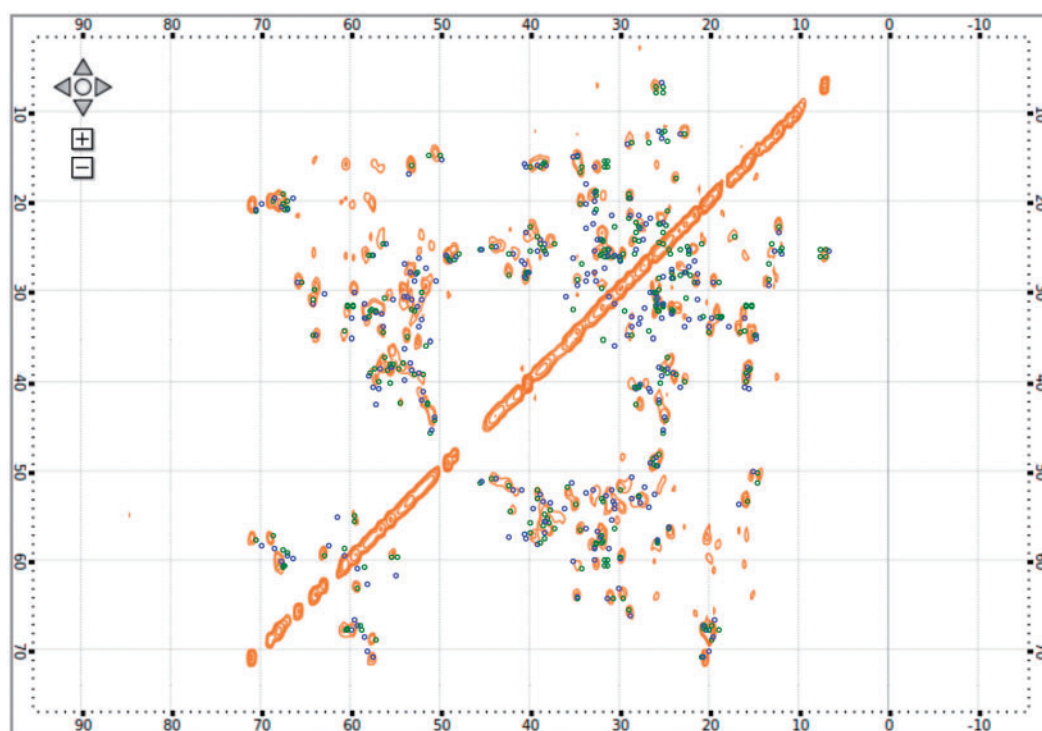
**Fig. 3.** Screenshot of the Peakr web application. The part in which spectra can be simulated is shown divided into the sections: Protein, Chemical shifts and Correlations. The example dataset provided on the website, with solid-state chemical shifts of ubiquitin as published in Seidel *et al.* (2005), is shown here as input



**Fig. 4.** The screenshot shows the list of simulated spectra after cloning the first spectrum using the Clone function (third icon from left). Different parts of the protein sequence were selected (see Start and End values). The spectra show resonances from the first 20 residues and from residues 30 to 76. For the second spectrum, only some of the amino acid types present in the sequence were selected, which are thus highlighted in orange

two- or three-bond correlations. Then, we compared these simulated spectra with the experimental data by listing the intensities at the positions of the simulated peaks in the experimental spectrum. Based on visual inspection of the experimental spectrum, an intensity cutoff of 3000 was chosen to differentiate between signal and noise. Based on this cutoff value, 53.4% of one-bond, 17.9% of two-bond, and 2.4% of three-bond simulated correlations were found to be present in the spectrum, showing that the experimental spectrum yielded mainly one-bond and to some extent two-bond correlations under the conditions chosen.

With this approach for comparing simulation with experiment, spectral intensities at exactly the position of simulated peaks are returned. It can be useful to allow for more variability in peak



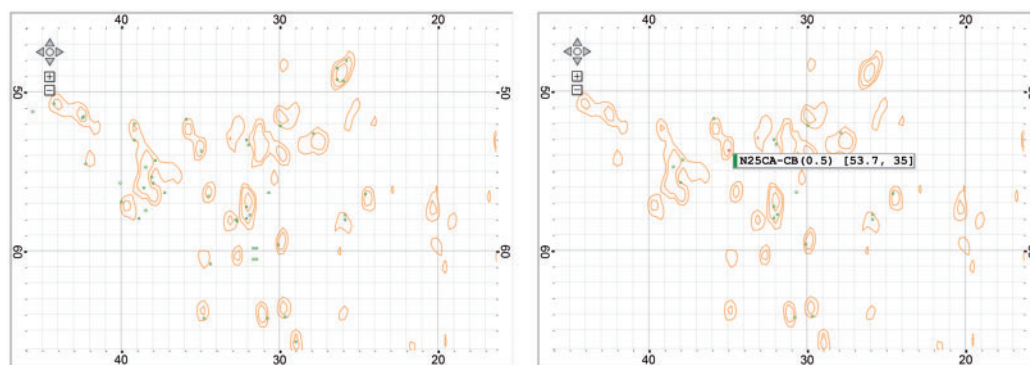
**Fig. 5.** Screenshot of the Peakr spectrum display window showing the example discussed in Section 3. Light grey (orange): Orange: experimental DARR ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) spectrum of microcrystalline ubiquitin; Light grey and dark grey circles (green and blue, respectively): green and blue: simulated one-bond ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations based on assignments reported in Seidel *et al.* (2005) (light grey; green) (green) and based on assignments from Igumenova *et al.* (2004) (dark grey; blue) (blue)

positions to account for, for example, small variations in sample conditions such as temperature or pH. To do so, Peakr can return the maximum intensity in a defined region of the experimental spectrum around each simulated peak. Allowing for  $\pm 0.2$  parts per million (p.p.m.) variation in peak position when comparing the above one-bond ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) simulated spectrum with experiment, 74.5% of simulated one-bond correlations are found above the selected threshold, indicating good agreement between simulation and experiment (Fig. 5, green dots).

For comparison, we simulated the same one-bond ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlation spectrum of ubiquitin using assignments reported for a different microcrystalline ubiquitin preparation (Igumenova *et al.*, 2004), adjusted for the referencing offset of 2.01 p.p.m. between these assignments and the values used above. For this ubiquitin preparation using 2-methyl-2,4-pentanediol as precipitant, rather than poly-(ethylene glycol; Seidel *et al.*, 2005), significant chemical shift differences were reported in some regions of the protein (Seidel *et al.*, 2005; Schneider *et al.*, 2010a). Correspondingly, Peakr only finds 66.0% of all simulated one-bond peaks within a range of  $\pm 0.2$  p.p.m. of a spectral intensity above the selected threshold (Fig. 5, blue dots). Whether such a difference in the percentage of simulated peaks above an intensity threshold is significant is of course up to the user to decide, as there is no universal way to judge whether a peak can be considered present in an experimental spectrum and what percentage of simulated peaks needs to be present in an experimental spectrum to constitute good agreement between them. Such judgements always also depend on, for example, accuracy of

third-party chemical shift prediction algorithms used as well as signal to noise and resolution in the experimental spectrum and are beyond the scope of Peakr. However, this example illustrates the utility of Peakr to quickly make qualitative assessments about quality and state of a protein sample as well as to compare different alternative hypotheses to explain a spectrum. Expected peaks that are absent from an experimental spectrum, weaker than expected or shifted may hint at conformational differences or local motion (Schneider *et al.*, 2010a). Thus, Peakr can directly point the user to spectral regions that merit further investigation.

Several specific  $^{13}\text{C}$ -labeling schemes that have been used in solid-state NMR studies in recent years are also implemented in Peakr. Labeling patterns obtained from using 1,3- $^{13}\text{C}$ - or 2- $^{13}\text{C}$ -glycerol (Castellani *et al.*, 2002) as well as 1- $^{13}\text{C}$ - or 2- $^{13}\text{C}$ -glucose as sole carbon sources (Hong, 1999; Lundström *et al.*, 2007) can be selected for spectrum simulation. This feature is demonstrated in Figure 6 for the same one-bond ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlation spectrum of ubiquitin as shown in green in Figure 5, using the 1,3- $^{13}\text{C}$ -glycerol-labeling scheme. Peakr calculates opacity values of individual peaks according to the  $^{13}\text{C}$ -labeling probabilities of the nuclei that give rise to the correlation. For the glycerol-based schemes, detailed labeling probabilities and isotopomer patterns are available (Castellani *et al.*, 2002) and implemented in Peakr, while the simplified scheme as shown in (Lundström *et al.*, 2007) is used for simulating spectra with 1- $^{13}\text{C}$ - and 2- $^{13}\text{C}$ -glucose-based labeling. Such simulated spectra should be very helpful in assigning spectra of proteins expressed with one of these



**Fig. 6.** Screenshots of the Peakr spectrum display window demonstrating different labeling schemes. Shown is a region from the ubiquitin spectrum shown in Figure 5, with simulated one-bond ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) correlations based on the assignments reported in Seidel *et al.* (2005). **(A)** Simulated spectrum based on a uniformly  $^{13}\text{C}$ -labeled sample. **(B)** Simulated spectrum based on the labeling scheme expected from using 1,3- $^{13}\text{C}$ -glycerol as sole carbon source during protein expression. For one cross-peak, a tooltip shows assignment, chemical shifts and opacity value (corresponding to the probability that the corresponding nuclei are both isotope labeled)

labeling patterns, effectively circumventing the need for the user to consult tables of labeling schemes manually. In addition to the option to select only certain amino acid types for spectrum simulation, the selective  $^{13}\text{C}$ -labeling option in Peakr allows to assess which labeling method would best reduce spectral crowding for larger proteins with sizable spectral overlap. Complementary to approaches such as the UPLABEL algorithm (Hefke *et al.*, 2011), this offers a fast and convenient way to guide protein expression strategies for further experiments.

#### 4 CONCLUSIONS

The Peakr software presented here can be of considerable help when analyzing solid-state NMR spectra of proteins. It can simulate 2D spectra for many common experimental setups. The simulated spectra can be helpful for guiding the resonance assignment process and for deriving restraints for 3D structure calculations. As demonstrated in the case study, basic assumptions about a measured spectrum can be made in a matter of seconds, which can be useful in quality control of samples. In contrast to existing solutions, Peakr is very flexible and can use subsets of residues or nuclei to define spectra. This is especially valuable when reverse or selective labeling methods are used or when only a portion of the protein, for example, the N-terminus, is of interest. Here, Peakr spectrum simulations can, for example, be used to assess which isotope-labeling patterns would be optimal for a given protein to reduce spectral crowding. Peakr's ability to rapidly simulate intra- and intermolecular through-space correlation spectra, with the same flexibility in choosing protein regions as well as upper distance limits to be considered, should be especially valuable in solid-state NMR structural studies. The option to compare simulated with measured spectra allows for estimating the degree of agreement between simulation and measurement. In this context, the percentage of simulated cross-peaks with a measured intensity above a given threshold can be seen as a simple figure of merit.

The Peakr framework is itself highly flexible and can accommodate extensions desired by its users. Future versions may thus, for example, be extended to simulate 3D correlation spectra or

proton-detected experiments, which are increasingly used in solid-state NMR (Habenstein *et al.*, 2011; Linser *et al.*, 2011), as well as to incorporate solution-state NMR correlation types.

In summary, Peakr has the power and flexibility to become a useful tool for routine analysis of solid-state NMR spectra. It is thus hoped that the community will adopt it and provide active feedback for further improvement and extension.

#### ACKNOWLEDGEMENTS

The ubiquitin spectrum was kindly provided by Dr Hans Förster and Dr Stefan Steuernagel (Bruker Biospin, Karlsruhe). The authors thank Prof. Christian Griesinger for discussions and continuous support and Dr Paul Schanda (IBS, Grenoble) for providing test spectra.

*Funding:* We thank the Max-Planck Society and especially the Department of NMR based Structural Biology, headed by Prof. Christian Griesinger, at the MPI for Biophysical Chemistry for generous financial support.

*Conflict of interest:* none declared.

#### REFERENCES

- Advanced Chemistry Development. (2007) *ACD/Labs NMR Predictors*. [computer programme]. Toronto, ON, Canada.
- Bax, A.D. *et al.* (1980) Natural abundance carbon-13-carbon-13 coupling observed via double-quantum coherence. *J. Am. Chem. Soc.*, **102**, 4849–4851.
- Berman, H. *et al.* (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
- Binev, Y. and Aires-de-Sousa, J. (2004) Structure-based predictions of  $^1\text{H}$  NMR chemical shifts using feed-forward neural networks. *J. Chem. Inf. Comput. Sci.*, **44**, 940–945.
- Castellani, F. *et al.* (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*, **420**, 98–102.
- Delaglio, F. *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
- Frishman, D. and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.
- Goddard, T.D. and Kneller, D.G. *SPARKY3*. University of California, San Francisco.

- Golovtin, S.S. et al. (2007) Automated structure verification based on a combination of 1D ( $^1\text{H}$ ) NMR and 2D ( $^1\text{H}$  - ( $^{13}\text{C}$ ) HSQC) spectra. *Magn. Reson. Chem.*, **45**, 803–813.
- Goto, N. et al. (2010) BioRuby: bioinformatics software for the ruby programming language. *Bioinformatics*, **26**, 2617–2619.
- Gronwald, W. and Kalbitzer, H.R. (2004) Automated structure determination of proteins by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectr.*, **44**, 33–96.
- Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
- Habenstein, B. et al. (2011) Extensive de novo solid-state NMR assignments of the 33 kDa C-terminal domain of the Ure2 prion. *J. Biomol. NMR*, **51**, 235–243.
- Han, B. et al. (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR*, **50**, 43–57.
- Hefke, F. et al. (2011) Optimization of amino acid type-specific  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling for the backbone assignment of membrane proteins by solution- and solid-state NMR with the UPLABEL algorithm. *J. Biomol. NMR*, **49**, 75–84.
- Heise, H. et al. (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proc. Natl Acad. Sci. USA.*, **102**, 15871–15876.
- Hong, M. (1999) Determination of multiple  $\phi$ -torsion angles in proteins by selective and extensive  $^{13}\text{C}$  labeling and two-dimensional solid-state NMR. *J. Magn. Reson.*, **139**, 389–401.
- Igumenova, T.I. et al. (2004) Assignments of carbon NMR resonances for microcrystalline ubiquitin. *J. Am. Chem. Soc.*, **126**, 6720–6727.
- Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
- Judge, P.J. and Watts, A. (2011) Recent contributions from solid-state NMR to the understanding of membrane protein structure and function. *Curr. Opin. Chem. Biol.*, **15**, 690–695.
- Kohlhoff, K.J. et al. (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.*, **131**, 13894–13895.
- Lange, A. et al. (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. et al. (2002) Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J. Am. Chem. Soc.*, **124**, 9704–9705.
- Linsler, R. et al. (2011) Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angew. Chem. Int. Ed. Engl.*, **50**, 4508–4512.
- Lundström, P. et al. (2007) Fractional  $^{13}\text{C}$  enrichment of isolated carbons using [ $1-^{13}\text{C}$ ] or [ $2-^{13}\text{C}$ ] glucose facilitates the accurate measurement of dynamics at backbone  $\alpha$  and side-chain methyl positions in proteins. *J. Biomol. NMR*, **38**, 199–212.
- Manolikas, T. et al. (2008) Protein structure determination from  $^{13}\text{C}$  spin-diffusion solid-state NMR spectroscopy. *J. Am. Chem. Soc.*, **130**, 3959–3966.
- Markley, J. et al. (2008) BioMagResBank (BMRB) as a partner in the worldwide protein data bank (wwPDB): new policies affecting biomolecular NMR depositions. *J. Biomol. NMR*, **40**, 153–155.
- Matsuki, Y. et al. (2007) Spectral fitting for signal assignment and structural analysis of uniformly  $^{13}\text{C}$ -labeled solid proteins by simulated annealing based on chemical shifts and spin dynamics. *J. Biomol. NMR*, **38**, 325–339.
- McDermott, A. (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu. Rev. Biophys.*, **38**, 385–403.
- Menger, E.M. et al. (1986) Observation of carbon-carbon connectivities in rotating solids. *J. Am. Chem. Soc.*, **108**, 2215–2218.
- Neal, S. et al. (2003) Rapid and accurate calculation of protein  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts. *J. Biomol. NMR*, **26**, 215–240.
- Nielsen, J.T. et al. (2012) Chemical shift prediction for protein structure calculation and quality assessment using an optimally parameterized force field. *Prog. Nucl. Magn. Reson. Spectrosc.*, **60**, 1–28.
- Renault, M. et al. (2010) Solid-state NMR spectroscopy on complex biomolecules. *Angew. Chem. Int. Ed. Engl.*, **49**, 8346–8357.
- Schneider, R. et al. (2010a) Probing molecular motion by double-quantum ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) solid-state NMR spectroscopy: application to ubiquitin. *J. Am. Chem. Soc.*, **132**, 223–233.
- Schneider, R. et al. (2010b) The native conformation of the human VDAC1 N terminus. *Angew. Chem. Int. Ed. Engl.*, **49**, 1882–1885.
- Seidel, K. et al. (2005) High-resolution solid-state NMR studies on uniformly [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-labeled Ubiquitin. *Chem. Bio. Chem.*, **6**, 1638–1647.
- Shen, Y. and Bax, A.D. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR*, **38**, 289–302.
- Shen, Y. and Bax, A.D. (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.J. et al. (2011) A software framework for analysing solid-state MAS NMR data. *J. Biomol. NMR*, **51**, 437–447.
- Takegoshi, K. et al. (2001)  $^{13}\text{C}$ - $^1\text{H}$  dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.*, **344**, 631–637.
- The PyMOL Molecular Graphics System Schrödinger, LLC.
- Tycko, R. (2011) Solid-State NMR Studies of Amyloid Fibril Structure. *Annu. Rev. Phys. Chem.*, **62**, 279–299.
- Ulrich, E.L. et al. (2007) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
- Vranken, W.F. et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
- Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56, 29.
- Vuister, G.W. et al. (1994) 2D and 3D NMR Study of phenylalanine residues in proteins by reverse isotopic labeling. *J. Am. Chem. Soc.*, **116**, 9206–9210.
- Wasmer, C. et al. (2008) Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science*, **319**, 1523–1526.
- Xu, X.P. and Case, D.A. (2001) Automated prediction of  $^{15}\text{N}$ ,  $^{13}\text{C}$   $\alpha$ ,  $^{13}\text{C}$   $\beta$  and  $^{13}\text{C}$   $\gamma$  chemical shifts in proteins using a density functional database. *J. Biomol. NMR*, **21**, 321–333.