Coversheet

This is the accepted manuscript (post-print version) of the article. Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication
Please cite the final published version:

Vosegaard T. Fast simulations of multidimensional NMR spectra of proteins and peptides. Magn. Reson. Chem. 2018;56:438–448. https://doi.org/10.1002/mrc.4663

Publication metadata

Title: Fast simulations of multidimensional NMR spectra of proteins and peptides
Author(s): Thomas Vosegaard
Journal: Magnetic Resonance in Chemistry
DOI/Link: https://doi.org/10.1002/mrc.4663
Document version: Accepted manuscript (post-print)

This is the peer reviewed version of the following article: [Vosegaard T. Fast simulations of multidimensional NMR spectra of proteins and peptides. Magn. Reson. Chem. 2018;56:438–448. https://doi.org/10.1002/mrc.4663], which has been published in final form at [https://doi.org/10.1002/mrc.4663]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

General Rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.
Fast simulations of multidimensional NMR spectra of proteins and peptides

Thomas Vosegaard

Danish Center for Ultrahigh-Field NMR Spectroscopy, Interdisciplinary Nanoscience Center and Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark

Correspondance to
Thomas Vosegaard
Danish Center for Ultrahigh-Field NMR Spectroscopy
Interdisciplinary Nanoscience Center and Department of Chemistry
Gustav Wieds vej 14
DK-8000 Aarhus C
Denmark
Email: tv@chem.au.dk
Tel: +45 87 15 99 29
Abstract

To simulate full multidimensional NMR spectra of peptides and proteins in a reasonable time frame, a strategy for separating the time-consuming full-density matrix calculations from the chemical shift prediction and calculation of coupling patterns is presented. The simulation setup uses SIMPSON to calculate TOCSY transfer amplitudes and average distances as a source for NOESY transfer amplitudes. Simulated $^1$H 1D, 2D TOCSY, and 2D NOESY NMR spectra of peptides with sequence PAGYN and NFGAIL and of ubiquitin are presented. In all cases, the simulations lasted from a few seconds to tens of seconds on a normal laptop computer.

Key-words:
SIMPSON; chemical shift prediction; TOCSY; NOESY; simulated spectra

Graphical abstract
Introduction

The nuclear spin interactions that we measure in NMR experiments carry important information on the molecular structure and dynamics. This offers a very diverse range of applications of NMR. For example, NMR experiments are routinely used to elucidate the structure of molecules in organic synthesis, in more detailed studies of protein structures, as finger-printing in metabolomics, in studies of polymers, or to characterise amorphous or crystalline materials. To obtain the desired information from NMR, the following approaches are the most prevalent. (i) Use of empirical knowledge based on experiments on similar compounds and (ii) simulations based on Hamiltonian theory of the nuclear spin interactions.

The first approach is typically used for interpretation of NMR spectra of small organic molecules, where researchers in the early days established important empirical knowledge on $^1$H and $^{13}$C chemical shifts, for example as gathered by Grant and Paul.[1] Such studies have paved the way for more recent establishments of databases for spectrum prediction.[2–9] For proteins, similar approaches have been developed, starting from the pioneering work[10] to programs predicting structural features of proteins from NMR spectra[11] or providing random-coil shifts based on the protein sequence.[12,13]

Simulations and theoretical descriptions of NMR pulse sequences by product operators[14] have been essential for virtually all pulse-sequence developments. While it is often possible to obtain insight in pulse sequences from analytical descriptions, it may be convenient to establish numerical simulations, e.g. for large spin systems or powders with orientational effects. To achieve this, simulation programs like SIMPSON,[15–18] SpinEvolution,[19] Spinach,[20] and SpinDynamica[21] have been developed. These versatile programs have been used to, for example, develop optimal control NMR experiments,[22] investigation of proton-driven spin diffusion,[23] and simulation of multidimensional NMR spectra of proteins.[24]

In this tutorial manuscript, focus will be on coupling information obtained from different sources to provide new insight, as previously outlined for simulations of so-called CHHC experiments[25] for proteins in the solid state.[26] Here, it will be demonstrated that simulated 1D and 2D TOCSY and NOESY
spectra of small proteins can easily be obtained using H and HA random-coil chemical shifts from a chemical-shift predictor\textsuperscript{[13]} or from the BioMagResBank,\textsuperscript{[27]} TOCSY transfer efficiencies calculated using SIMPSON,\textsuperscript{[15–18]} and NOESY cross peak intensities obtained from average interatomic distances reported for ensembles of intrinsically disordered proteins\textsuperscript{[28]} analysed using VMD,\textsuperscript{[29]} and that these bits can be put together in a simple manner to obtain simulated multidimensional NMR spectra of peptides and proteins.

It should be emphasized that the goal of this work is not to provide the most accurate reproduction of experimental spectra or to establish new methods for spectrum prediction. The goal is to combine various computational and empirical methods to achieve, in a simple manner, simulated spectra that would otherwise require tremendous efforts to obtain. We use this approach to generate liquid-state NMR spectra of small peptides that students can use to practice the procedures for protein NMR spectrum assignment.

**Simulation methods**

This section establishes the procedures for obtaining chemical shifts and J couplings defining the nuclear spin systems as well as procedures for estimation of TOCSY and NOESY cross peak intensities. The aim is that only the amino acid sequence should be added as input, then the scripts will return the simulated NMR spectra. As the focus is on small peptides, we assume that they adopt random-coil structures. However, the scripts are designed in a way that they can simulate the NMR spectra based on user-input chemical shift parameters.

We will use the Tcl scripting interface of SIMPSON\textsuperscript{[15–18]} to gather all information from different sources and generate the simulated spectra.

**Chemical shifts and J couplings**

The random-coil chemical shift predictor by Kjaergaard \textit{et al.}\textsuperscript{[13]} has been selected, as it has a JavaScript interface that has been easy to adopt to the current Tcl setup. We have done this by creating a Tcl package, which allows
the chemical shift predictor to be embedded in any Tcl script. The package is available for download from our web site.[30] In Tcl (and thereby in a SIMPSON input file), a package is included by the following commands

```tcl
lappend ::auto_path /path/to/package
package require randomcoil
```

In random coil structures, the $3J_{HNHA}$ vary among the amino acids, as reported by Dobson and co-workers.[31] We have included these $J$-coupling values in the random-coil package, so chemical shifts and $J$-couplings can be obtained for a specific sequence by the following commands

```tcl
set sequence "NFGAIL"
set shifts [randomcoil::shift $sequence 25 6.5]
set jcoupl [randomcoil::jhnha $sequence]
```

where the two additional inputs to the shift are temperature (in deg C) and pH. As the predictor does not distinguish between the two alpha protons of Gly, we manually add and subtract 0.03 ppm to the predicted HA chemical shift of Gly HA2 and HA3, respectively, as we know that there is a typical separation of 0.06 ppm between the two hydrogens.

The chemical shifts in the side chains are obtained from the BioMagResBank,[27] which reports statistics on all chemical shifts, so we can obtain the average chemical shifts and their standard deviations.[32] To establish the spin systems, three-bond $J$ couplings of $3J_{HH} = 7$ Hz are added between all relevant hydrogens, and in cases where geminal hydrogens are different (neighbor to chiral carbon, i.e., in general HB2 and HB3, Ile HG12 and HG13, and Gly HA2 and HA3), $2J_{HH} = 15$ Hz are added as well. The SIMPSON description of the spin systems for all 20 amino acids are given in the Supporting information. For fast generation of spin systems, we represent the spin system of alanine by the following definitions

```tcl
set atoms(ala) {{H 1} {HA 1} {HB 3}}
set coupling(ala) {{H HA 5.8} {HA HB 7}}
set shift(ala,H) {8.19 0.58}
```
The first line defines the three different atoms. For each atom is given the name and number of atoms, to for alanine there are three HB atoms. The second line lists the $J$ couplings in the amino acid, and the three last lines list the chemical shifts obtained from the BioMagResBank along with the standard deviations reported in the database. In SIMPSON format, the spin system will be defined as

```
spinsys {
  channels 1H
  nuclei 1H 1H 1H 1H 1H
  shift 1 8.19p 0 0 0 0 0
  shift 2 4.24p 0 0 0 0 0
  shift 3 1.36p 0 0 0 0 0
  shift 4 1.36p 0 0 0 0 0
  shift 5 1.36p 0 0 0 0 0
  jcoupling 1 2  5.8 0 0 0 0 0
  jcoupling 2 3  7.0 0 0 0 0 0
  jcoupling 2 4  7.0 0 0 0 0 0
  jcoupling 2 5  7.0 0 0 0 0 0
}
```

See Supporting information for a full list of spin systems for the amino acids.

**TOCSY cross peak intensities**

To establish the TOCSY cross peak intensities, SIMPSON is to calculate the transfer efficiencies between the different hydrogen atoms in each amino acid using the spin systems as explained above. We have used the DIPSI-2 pulse sequence,[33] which in SIMPSON may be implemented as (remember that pulse lengths are in µs)

```
proc dipsi2 {rf} {
  set tp [expr 0.25e6/$rf]

  foreach ph1 {x -x -x x} ph2 {-x x x -x} {
    pulse [expr $tp*3.556] $rf $ph1
    pulse [expr $tp*4.556] $rf $ph2
    pulse [expr $tp*3.222] $rf $ph1
    pulse [expr $tp*3.167] $rf $ph2
    pulse [expr $tp*0.333] $rf $ph1
    pulse [expr $tp*2.722] $rf $ph2
  }
}
```
pulse [expr $tp*4.167] $rf $ph1
pulse [expr $tp*2.944] $rf $ph2
pulse [expr $tp*4.111] $rf $ph1
}
}

The total length of this block is 115.11 times the 90-degree pulse length. Hence, to achieve a certain mixing time ($taum, here 80 ms), we make the following use of the SIMPSON propagator command to repeat the DIPSI-2 block

```
proc pulseq {} {
    ...
    set taum 80000
    set ndipsi [expr int($taum/(115.112*(0.25e6/$rf)))]
    dipsi2
    store 1
    prop 1 [expr $ndipsi-1]
    store 1
    ...
}
```

which provides a simple and efficient way to calculate the TOCSY pulse sequence, as the total mixing sequence is now stored as propagator 1 in SIMPSON.

The DIPSI-2 sequence transfers longitudinal magnetization. The goal is to figure out how much magnetization is transferred between the different hydrogen atoms in each amino acid. Therefore, we use SIMPSON to probe how this pulse sequence can transfer magnetization from one hydrogen to another, one by one. The most efficient way to implement this is by using the possibility to define the start and detect operators from within the pulse sequence

```
proc pulseq {} {
    ...
    for {set i 1} {$i <= $n} {incr i} {
        matrix set start operator I$i$z
        for {set j 1} {$j <= $i} {incr j} {
            ...
        }
    }
```
This pulse sequence runs through a double loop where it sets the start operator to $I_z$ on nucleus 1, 2, …, n, assuming that the spin system has n hydrogens. The detect operator is likewise looped, and the transfer amplitude is sampled. The transfer amplitudes will later be read out in the main section of the SIMPSON input file, for example by the code snippet

```
proc main {} {
    ...
    set f [fsimpson]
    set ii 1
    for {set i 1} {$i <= $n} {incr i} {
        for {set j 1} {$j <= $i} {incr j} {
            puts "$i ® $j: [findex $f $ii -re]"
            incr ii
        }
    }
}
```

Table 1 shows the result of the simulation for Ile (11 hydrogen spins). Although this is a large spin system, the simulation may be performed in a few minutes using a normal desktop computer. Transfer efficiencies for all amino acids are given in the Supporting information.

**NOESY cross peak intensities**

Ignoring longitudinal relaxation, NOESY cross peaks appear due to cross relaxation coupling nuclei I and S through this equation \[34\]
\[
\frac{dI_s}{dt} = \sigma_{IS}(S_2^0 - S_2),
\]

where \(\sigma_{IS}\) is the cross-relaxation rate, that depends on the spectra density function and is proportional to the squared dipolar coupling. It may be hard to precisely quantify the NOESY cross-peak intensities due to the many parameters influencing the NOESY spectra, so here we assume that the cross peak intensity can be described by

\[
I_{IS} = 1 - \exp\left(-r_0^6 \left(\frac{1}{r_{IS}^6}\right) \ln 2\right),
\]

and to model a NOESY spectrum with a relatively short mixing time favouring short distances, we here choose a value of \(r_0 = 3\) Å, which yields a cross peak intensity of 50% for \(r_{IS} = 3\) Å and less than 1% at \(r_{IS} = 6\) Å.

*Insert tables 2 and 3 here*

To establish average distances \(\langle r_{IS}^6 \rangle\), ensembles of structures for intrinsically disordered proteins have been used.\(^{[28]}\) As an example, Table 2 lists the average distances and transfer efficiencies based on Eq. (2) for isoleucine, and Table 3 lists the inter-residue distances required to do the sequential assignment. In the simulations, only the inter-residue distances in the first column are included, as they are by far the most intense, in line with normal expectations from protein NMR.\(^{[35]}\)

**Simulation details**

To simulate the 1D and 2D spectra of peptides and proteins, the calculation speed needs be considered. It is not convenient to perform full density-matrix calculations on large spin systems, although such efforts have been reported in a single case,\(^{[24]}\) so our approach relies only on such calculations to obtain the TOCSY transfer amplitudes. The actual simulation of the spectra is done simply by adding peaks at the positions of the chemical shifts of the different atoms. The coupling pattern is based on a simple splitting of the lines. As many couplings involve nuclei with similar chemical shifts (e.g. all the vicinal
couplings), all couplings are handled as strong couplings, i.e. a coupling between nuclei \( a \) and \( b \) gives rise to four peaks

\[
v = \frac{\nu_a + \nu_b}{2} \pm \frac{Q}{2} \pm \frac{J}{2},
\]

where \( \nu_a \) and \( \nu_b \) are the chemical shifts, \( J \) is the \( J \)-coupling, and

\[
Q = \sqrt{(\nu_a - \nu_b)^2 + J^2}.
\]

The strongly coupled spin system approaches a normal 1:1 doublet, when the chemical shift difference is large (i.e. \( Q \to |\nu_a - \nu_b| \)) and shows the well-known roofing when the chemical shift difference is small (i.e. \( Q \to J \), assuming a positive \( J \)-coupling). The intensity of the inner and outer peaks is given by

\[
\text{Inner} = \frac{Q + J}{2Q}, \quad \text{Outer} = \frac{Q - J}{2Q}.
\]

The coupling pattern for each nucleus is handled by nesting all its \( J \) couplings through the normal \( J \)-coupling tree, where the second coupling splits the lines of the first coupling according to the peak positions and intensities given in Eqs. (3) and (5).

Taking HA2 of Gly3 in the peptide PAGYN as an example, the HA2 chemical shift is assumed to be 3.95 ppm. We assume it has a \(^3J_{HNHA}\) coupling of 5.8 Hz and a \(^2J_{HAHA}\) coupling of 15 Hz. At 600 MHz the cascade of couplings will be as follows, here listed as frequency(intensity)

\[
2370.0(1.00) \rightarrow \{2367.1(0.50) \rightarrow \begin{cases} 2361.1(0.35) \\ 2366.9(0.35) \end{cases} \rightarrow \begin{cases} 2376.1(0.15) \\ 2381.9(0.15) \end{cases} \}
\]

and similarly for HA3. Each of these signals gives a double doublet with significant roofing because of the strong coupling with HA3, which has a chemical shift of 3.89 ppm.
Results and discussion

Focusing on small peptides, we will assume they adopt random-coil structures, but it would be straight forward to the present procedures with chemical shift predictors like SHIFTX,[36,37] SHIFTY,[38] or PROSHIFT,[39] if equipped with a known protein structure. In the current simulation setup, the flow is as follows

1. Establish the spin system including chemical shifts from random-coil predictions or by user-input and J couplings assuming standard values
2. Simulate the coupling patterns in 1D spectrum
3. Get TOCSY and NOESY transfer amplitudes by looking up the tables given in Supporting information
4. Simulate the 2D spectra by extending the couplings in two dimensions

The SIMPSON script that performs these tasks is available for download from our web site.[40] Alternatively, all files (the script and two Tcl libraries) are made available as supporting information to this paper.

*Insert figures 1 and 2 and table 4 here*

To illustrate the possibility to combine all this information, we will simulate \(^1\)H 1D and 2D spectra of a pentapeptide with sequence Pro-Ala-Gly-Tyr-Asn (PAGYN). The molecular structure of this pentapeptide is shown in Fig. 1, and the predicted chemical shifts are summarized in Table 4. The simulated 1D spectrum is shown in Fig. 2 along with the assignment of each peak. By inspection of the coupling patterns, each of them indeed shows the expected couplings and multiplicities. Only the peaks of 4Tyr HA and 5Asn HA are so overlapping that it is not possible to distinguish the individual coupling patterns.

*Insert figures 3 and 4 here*
The 2D TOCSY and NOESY spectra shown in Fig. 3 and with relevant expansions in Fig. 4 clearly demonstrate the expected intra-residue cross peaks, and in the case of NOESY, also inter-residue cross peaks according to the short distances listed in Table 3. We often use this kind of spectra when teaching undergraduate students to assign NMR spectra of proteins, as they provide simple examples for the basic principles of both peak assignment and sequential assignment.

*Insert figure 5 here*

To compare the simulations with experimental spectra, Figure 5a shows the experimental $^1$H NMR spectrum of the peptide NFGAIL, synthesized by solid-phase peptide synthesis as described elsewhere.$^{[41,42]}$ The spectrum was recorded on a Bruker spectrometer at 400.13 MHz in 9:1 H$_2$O:D$_2$O and using a pulse sequence with presaturation for suppression of the strong water signal. The assigned chemical shifts are listed in Table 5. Using these chemical shifts as input, our present simulation setup provided the simulated spectrum in Fig. 5b, which very much resembles the experimental spectrum. The major differences are varying intensities observed in the experimental spectrum, e.g. from water suppression. These are not modelled in our setup so the intensities differ from the experimental to the simulated spectrum for various signals.

As a final example of the present simulation setup, Fig. 5c shows the simulated 2D TOCSY spectrum of ubiquitin employing the chemical shifts reported in the BioMagResBank (BMRB id 68) from an early study.$^{[43]}$

Having demonstrated the possibility to simulate spectra of peptides and small proteins, it is relevant to evaluate the time it takes to perform such simulations. The computation time mainly consists of two contributions: (i) calculation of the peaks in the 2D spectra and (ii) writing the 2D datasets to the disk. The latter part takes a few seconds, and the first part depends on the number of amino acids in the sequence. In fact, the length of this part is proportional to the number of amino acids, as there are no long-range effects included in the simulations. Simple tests show that simulations of up to proteins with up to 100 amino acids can be done in less than a minute of
computation time on a normal laptop computer, while short peptides of less than 10 amino acids may be simulated in a few seconds.

The fact that the simulations are very fast is encouraging for making more precise estimates of transfer amplitudes in the multidimensional spectra and intensities in the 1D spectrum, since actual simulations of full multidimensional spectra as part of e.g. fitting a set of structural parameters is indeed within reach using the present concept of separating the computation-dense efforts from the full spectrum simulation.

Next project in line is to merge the present spectrum simulation into our new web-based spectrum analysis interface, for which we have already launched the spectrum viewer jsNMR\textsuperscript{[44]} and a spectroscopy game. A test site for the web-based version of the program may be found on our web site.\textsuperscript{[45]} Once the implementation is complete, it will be available together with all scripts employed in this work.

**Conclusions**

This paper demonstrates how information obtained from different sources (chemical shift prediction or BioMagResBank, density-matrix simulations, and average distances from PDB structures) may easily be put together to obtain simulated one- and multi-dimensional NMR spectra of peptides and proteins in solution. The outlined methods provide quite accurate simulations that may easily be improved by adding better chemical shift prediction and/or more empirical information, e.g. about typical TOCSY and NOESY transfer intensities.

**Acknowledgments**

Financial support from the Danish Ministry of Higher Education and Science (AU-2010-612-181), Carlsbergfondet, the European Commission (FET-OPEN grant 731475) and Aarhus University is acknowledged. Thanks to Drs. Jakob Toudahl Nielsen and Berit Paaske for useful discussions related to this project.
Supporting information

The Supporting information provides a step-by-step instruction on how to run the simulation using the present scripts. Furthermore, it provides detailed information about the TOCSY and NOESY transfer amplitudes and SIMPSON spin systems for each amino acid. The Supporting information also contains a ZIP file with the SIMPSON script and Tcl libraries used to simulate the 1D and 2D protein NMR spectra.

References

[1] D.M. Grant, E.G. Paul. J. Am. Chem. Soc. 1964, 86, 2984–2990.
[2] J. Aires-de-Sousa, M.C. Hemmer, J. Gasteiger. Anal. Chem. 2002, 74, 80–90.
[3] Y. Binev, M.M.B. Marques, J. Aires-de-Sousa. J. Chem. Inf. Model. 2007, 47, 2089–2097.
[4] A.M. Castillo, A. Bernal, R. Dieden, L. Patiny, J. Wist. J. Cheminformatics 2016, 8.
[5] ACD/HNMR Predictor; Advanced Chemistry Developments.
[6] SPINUS-WEB: Prediction of NMR spectra - 1H NMR Chemical Shifts, Coupling Constants and Full Spectrum from the Molecular Structure http://www2.ccc.uni-erlangen.de/services/spinus/ (accessed Jun 14, 2017).
[7] Modgraph NMR Prediction http://www.modgraph.co.uk/product_nmr.htm (accessed Jun 14, 2017).
[8] ChemDraw v16.0; PerkinElmer. http://www.cambridgesoft.com/software/overview.aspx.
[9] Predict 1H NMR spectra http://www.cheminfo.org/Spectra/NMR/Predictions/1H_Prediction/index.html (accessed Jun 14, 2017).
[10] K. Wüthrich. *NMR of proteins and nucleic acids*; John Wiley & Sons, 1986.
[11] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax. J. Biomol. NMR 2009, 44, 213–223.
[12] K. Tamiola, B. Acar, F.A.A. Mulder. J. Am. Chem. Soc. 2010, 132, 18000–18003.
[13] M. Kjaergaard, S. Brander, F.M. Poulsen. J. Biomol. NMR 2011, 49, 139–149.
[14] O.W. Sørensen, G.W. Eich, M.H. Levitt, G. Bodenhausen, R.R. Ernst. Prog. Nucl. Magn. Reson. Spectrosc. 1984, 16, 163–192.
[15] M. Bak, J.T. Rasmussen, N.C. Nielsen. J. Magn. Reson. 2000, 147, 296–330.
[16] T. Vosegaard, A. Malmendal, N.C. Nielsen. Monatsh Chem 2002, 133, 1555–1574.
[17] Z. Tosner, T. Vosegaard, C. Kehlet, N. Khaneja, S.J. Glaser, N.C. Nielsen. J Magn Reson 2009, 197, 120–134.
[18] Z. Tošner, R. Andersen, B. Stevensson, M. Edén, N.C. Nielsen, T. Vosegaard. *J. Magn. Reson.* **2014**, *246*, 79–93.

[19] M. Veshtort, R.G. Griffin. *J. Magn Reson* **2006**, *178*, 248–282.

[20] H.J. Hogben, M. Krzystyniak, G.T.P. Charnock, P.J. Hore, I. Kuprov. *J. Magn Reson* **2011**, *208*, 179–194.

[21] SpinDynamica » NMR Calculation and Simulation in Mathematica [http://www.spindynamica.soton.ac.uk/](http://www.spindynamica.soton.ac.uk/) (accessed Jun 13, 2017).

[22] C.T. Kehlet, A.C. Sivertsen, M. Bjerring, T.O. Reiss, N. Khaneja, S.J. Glaser, N.C. Nielsen. *J Am Chem Soc* **2004**, *126*, 10202–10203.

[23] M. Veshtort, R.G. Griffin. *J. Chem. Phys.* **2011**, *135*, 134509.

[24] D.V. Savostyanov, S.V. Dolgov, J.M. Werner, I. Kuprov. *Phys Rev B* **2014**, *90*, 085139.

[25] M. Aluas, C. Tripon, J.M. Griffin, X. Filip, V. Ladizhansky, R.G. Griffin, S.P. Brown, C. Filip. *J Magn Reson* **2009**, *199*, 173–187.

[26] T. Vosegaard. *Solid State Nucl Magn Reson* **2010**, *38*, 77–83.

[27] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Mazuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Kent Wenger, H. Yao, J.L. Markley. *Nucleic Acids Res.* **2007**, *36*, D402–D408.

[28] M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A.K. Dunker, I.C. Felli, J.D. Forman-Kay, R.W. Kriwacki, R. Pierattelli, J. Sussman, D.I. Svergun, V.N. Uversky, M. Vendruscolo, D. Wishart, P.E. Wright, P. Tompa. *Nucleic Acids Res.* **2014**, *42*, D326–D335.

[29] W. Humphrey, A. Dalke, K. Schulten. *J Mol Graph* **1996**, *14*, 33–38.

[30] Tcl libraries [http://nmr.au.dk/software/tcl-libraries/](http://nmr.au.dk/software/tcl-libraries/) (accessed Jul 24, 2017).

[31] L.J. Smith, K.A. Bolin, H. Schwalbe, M.W. MacArthur, J.M. Thornton, C.M. Dobson. *J. Mol. Biol.* **1996**, *255*, 494–506.

[32] Full Chemical Shift Statistics [http://www.bmrb.wisc.edu/ref_info/statsel.htm](http://www.bmrb.wisc.edu/ref_info/statsel.htm) (accessed Jun 15, 2017).

[33] J. Cavanagh, M. Rance. *J. Magn. Reson.* **1969** **1992**, *96*, 670–678.

[34] J. Keeler. *Understanding NMR Spectroscopy*, 2nd ed.; Wiley, 2010.

[35] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press, 1995.

[36] S. Neal, A.M. Nip, H. Zhang, D.S. Wishart. *J. Biomol. NMR* **2003**, *26*, 215–240.

[37] B. Han, Y. Liu, S.W. Ginzinger, D.S. Wishart. *J. Biomol. NMR* **2011**, *50*, 43.

[38] D.S. Wishart, M.S. Watson, R.F. Boyko, B.D. Sykes. *J. Biomol. NMR* **1997**, *10*, 329–336.

[39] J. Meiler. *J. Biomol. NMR* **2003**, *26*, 25–37.

[40] Instant simulation of protein NMR spectra [http://nmr.au.dk/software/simpson-code-examples?Program=SIMPSON&mode=show&id=33](http://nmr.au.dk/software/simpson-code-examples?Program=SIMPSON&mode=show&id=33) (accessed Jul 24, 2017).

[41] J.T. Nielsen, M. Bjerring, M.D. Jeppesen, R.O. Pedersen, J.M. Pedersen, K.L. Hein, T. Vosegaard, T. Skrydstrup, D.E. Otzen, N.C. Nielsen. *Angew Chem Int Ed* **2009**, *48*, 2118–2121.

[42] C.U. Hjørringgaard, J.M. Pedersen, T. Vosegaard, N.C. Nielsen, T. Skrydstrup. *J Org Chem* **2009**, *74*, 1329–1332.
[43] P.L. Weber, S.C. Brown, L. Mueller. *Biochemistry (Mosc.)* 1987, 26, 7282–7290.

[44] T. Vosegaard. *Magn. Reson. Chem.* 2015, 53, 285–290.

[45] Instant simulation of 2D protein NMR spectra. http://jsnmr3.pastis.dk/protein.html (accessed Jul 24, 2017).
Table 1. TOCSY transfer efficiency between hydrogens in isoleucine for a 80 ms DIPSI-2 mixing block at 600 MHz using an rf field strength of 10 kHz and rf carrier frequency of 4 ppm.

|     | H   | HA  | HB  | HG12 | HG13 | HG21 | HG22 | HG23 | HD11 | HD12 | HD13 |
|-----|-----|-----|-----|------|------|------|------|------|------|------|------|
| H   | 11  |     |     |      |      |      |      |      |      |      |      |
| HA  | 53  | 13  |     |      |      |      |      |      |      |      |      |
| HB  | 10  | 6   | 19  |      |      |      |      |      |      |      |      |
| HG12| 2   | 3   | 10  | 15   |      |      |      |      |      |      |      |
| HG13| 2   | 3   | 10  | 23   | 15   |      |      |      |      |      |      |
| HG21| 6   | 6   | 10  | 5    | 5    | 32   |      |      |      |      |      |
| HG22| 6   | 6   | 10  | 5    | 5    | 13   | 32   |      |      |      |      |
| HG23| 6   | 6   | 10  | 5    | 5    | 13   | 13   | 32   |      |      |      |
| HD11| 1   | 2   | 5   | 10   | 10   | 3    | 3    | 3    | 32   |      |      |
| HD12| 1   | 2   | 5   | 10   | 10   | 3    | 3    | 3    | 15   | 32   |      |
| HD13| 1   | 2   | 5   | 10   | 10   | 3    | 3    | 3    | 15   | 15   | 32   |
Table 2. Average distances \(\langle r_i^{-6}\rangle_{1/6}\) and resulting transfer efficiencies calculated by Eq. 2 with \(r_0 = 3\) Å for hydrogen atoms in isolucine based on 30 ensemble structures of intrinsically disordered proteins (see text). a)

|      | H   | HA  | HB  | HG12 | HG13 | HG2  | HD1  |
|------|-----|-----|-----|------|------|------|------|
| H    | 100 b) |     |     |      |      |      |      |
| HA   | 2.94 / 54 | 100 b) |     |      |      |      |      |
| HB   | 3.07 / 45 | 2.50 / 87 | 100 b) |      |      |      |      |
| HG12 | 2.74 / 70 | 3.10 / 44 | 2.60 / 80 | 100 b) |      |      |      |
| HG13 | 3.09 / 44 | 3.35 / 30 | 2.44 / 91 | 1.77 / 100 | 100 b) |      |      |
| HG2  | 3.30 / 33 | 2.73 / 71 | 2.57 / 83 | 3.16 / 40 | 3.04 / 47 | 100 b) |      |
| HD1  | 2.13 / 100 | 3.90 / 13 | 3.15 / 40 | 2.59 / 81 | 2.57 / 83 | 2.92 / 56 | 100 b) |

a) The data is represented as distance / transfer (in percent). b) The NOESY diagonal peak intensities are set to 100%.
Table 3. Average distances ($<r_{ij}^{-6}>^{-1/6}$) and resulting transfer efficiencies calculated by Eq. 2 with $r_0 = 3$ Å for inter-residue contacts between hydrogen atoms in 30 ensemble structures of intrinsically disordered proteins (see text). $^a$

|       | $H_{i+1}$ | $HA_{i+1}$ | $HB_{i+1}$ |
|-------|-----------|------------|------------|
| $H_i$ | 2.89 / 58 | 4.96 / 3   | 5.36 / 2   |
| $HA_i$| 2.55 / 84 | 4.51 / 6   | 5.14 / 3   |
| $HB_i$| 3.13 / 42 | 4.68 / 5   | 5.37 / 2   |

$^a$The data is represented as distance / transfer (in percent).
Table 4. Predicted $^1$H chemical shifts for the peptide PAGYN.

| Residue | Atom | Shift (ppm) |
|---------|------|-------------|
| 1 Pro   | HA   | 4.45        |
|         | HB2  | 2.07        |
|         | HB3  | 2.00        |
|         | HG   | 1.92        |
|         | HD   | 3.65        |
| 2 Ala   | H    | 8.49        |
|         | HA   | 4.25        |
|         | HB   | 1.36        |
| 3 Gly   | H    | 8.33        |
|         | HA2  | 3.95        |
|         | HA3  | 3.89        |
| 4 Tyr   | H    | 7.99        |
|         | HA   | 4.63        |
|         | HB2  | 2.90        |
|         | HB3  | 2.84        |
|         | HD   | 6.93        |
|         | HE   | 6.70        |
| 5 Asn   | H    | 8.23        |
|         | HA   | 4.61        |
|         | HB2  | 2.80        |
|         | HB3  | 2.75        |
|         | HD21 | 7.32        |
|         | HD22 | 7.15        |
Table 5. Experimentally determined \(^1\text{H}\) chemical shifts for the peptide NFGAIL.

| Residue | Atom | Shift (ppm) |
|---------|------|-------------|
| 1 Asn   | HA   | 4.314       |
|         | HB2  | 2.954       |
|         | HB3  | 2.885       |
|         | HD21 | 7.633       |
|         | HD22 | 6.994       |
| 2 Phe   | H    | 8.699       |
|         | HA   | 4.710       |
|         | HB2  | 3.207       |
|         | HB3  | 3.093       |
|         | HD   | 7.318       |
|         | HE   | 7.416       |
|         | HZ   | 7.355       |
| 3 Gly   | H    | 8.312       |
|         | HA2  | 3.924       |
|         | HA3  | 3.840       |
| 4 Ala   | H    | 7.991       |
|         | HA   | 4.314       |
|         | HB   | 1.380       |
| 5 Ile   | H    | 8.116       |
|         | HA   | 4.180       |
|         | HB   | 1.889       |
|         | HG12 | 1.526       |
|         | HG13 | 1.213       |
|         | HG2  | 0.937       |
|         | HD1  | 0.871       |
| 6 Leu   | H    | 8.267       |
|         | HA   | 4.358       |
|         | HB2  | 1.680       |
|         | HB3  | 1.640       |
|         | HG   | 1.660       |
|         | HD1  | 0.944       |
|         | HD2  | 0.891       |
Figure 1. Molecular structure of the pentapeptide PAGYN with indication of the naming convention for the different hydrogen atoms.
Figure 2. Simulated 1D spectrum of the peptide PAGYN. The top row shows the full spectrum with indication of integrals below the spectrum. The two bottom rows are expansions of the amide-, alpha-, and aliphatic regions of the spectrum.
Figure 3. Simulated 2D TOCSY and NOESY spectra of the peptide PAGYN.
Figure 4. Expansions of the HN-HA fingerprint and aliphatic regions of the simulated TOCSY and NOESY spectra of the peptide PAGYN (top) and of the aliphatic TOCSY region (bottom).
Figure 5. (a-b) Comparison between the experimental $^1$H NMR spectrum (400.13 MHz) of the peptide NFGAIL (a) and the corresponding simulated spectrum (b). (c) Simulated 600 MHz NOESY spectrum of human ubiquitin based on assignments from an early work.$^{[43]}$