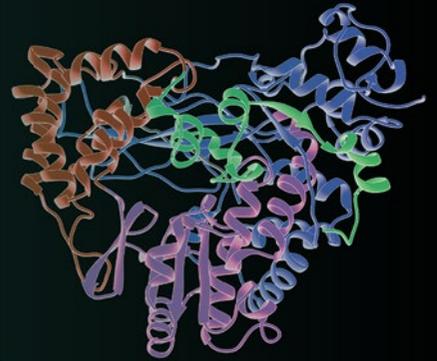


Methods in
Molecular Biology 1688

Springer Protocols



Ranajeet Ghose *Editor*

Protein NMR

Methods and Protocols

 Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:

<http://www.springer.com/series/7651>

Protein NMR

Methods and Protocols

Edited by

Ranajeet Ghose

Department of Chemistry and Biochemistry, The City College of New York, New York, NY, USA

 Humana Press

Editor

Ranajeet Ghose
Department of Chemistry and Biochemistry
The City College of New York
New York, NY, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7385-9 ISBN 978-1-4939-7386-6 (eBook)
DOI 10.1007/978-1-4939-7386-6

Library of Congress Control Number: 2017955368

© Springer Science+Business Media LLC 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media, LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Ever since its discovery in the work of Rabi [1], Purcell [2], and Bloch [3], NMR, now into its eighth decade, has found widespread use in the determination of the structure, dynamics, and interactions of molecules their assemblies. The applications of modern NMR spectroscopy range from the study of biomolecules to the analyses of the properties of batteries; from the solution state, through the semi-solid to the solid state; from the structural analysis of an end product of a chemical synthesis comprising only a few atoms, to the determination of the structural organization of an intact viral capsid with tens of thousands of atoms. The path spanning more than five decades of protein NMR from the one-dimensional spectrum of ribonuclease by Saunders et al. in 1957 [4] to a detailed NMR analysis of gating in a 670 kDa 20S proteasomal core by Sprangers et al. in 2007 [5] required multiple layers of development, in terms of technology (60 MHz in 1957 to 800 MHz in 2007), in fundamental concepts (continuous wave, CW NMR in 1957 to multidimensional relaxation optimized Fourier spectroscopy in 2007), and in sample preparation (unlabeled protein bought from a commercial source in 1957 to bacterially expressed protein specifically ^{13}C , ^1H enriched on Ile, Leu, and Val methyl groups in an otherwise ^{12}C , ^2H background in 2007). The success of any NMR experiment depends on four basic stages of equal importance: (1) sample preparation, (2) spectroscopy, (3) data processing, and (4) analysis. Improper implementation of any of these steps has the potential to doom any NMR-based research project. Many approaches have emerged over the years to optimize each of these stages of an NMR experiment. These include advances in magnet/spectrometer/probe design, novel biochemical methodology to enable the production of proteins or their assemblies optimally labeled with NMR-active nuclei, new pulse sequences to manipulate spin dynamics, and new computational approaches for data collection and analysis. In the 20 chapters of this volume we will cover state-of-the-art approaches detailing many of these innovations with a focus on their application to proteins and protein complexes in solution and in the solid state.

As we embark on the eighth decade of NMR spectroscopy, the magnetic fields that are now available to NMR spectroscopists have increased by more than an order of magnitude enabling the discovery of new spin physics as well as the study of more complex systems. While—as recently as the 1980s—an NMR spectrometer operating at 60 MHz (1.4 T) was considered state of the art, chemistry departments even in the smallest research universities these days have access to NMR spectrometers 500 MHz (11.7 T) or higher. Spectrometers operating at 800 MHz (18.8 T) have become staples for most medium-to-large research universities; a 1000 MHz (23.5 T) spectrometer has been operational at Lyon since 2009 and several new 1200 MHz (28.2 T) spectrometers, which are currently being developed, are on order to be placed at multiple sites. A report detailing the development of a hybrid magnet that allows the performance of NMR experiments at 1500 MHz (35.2 T) in the solid-state has recently appeared in the literature [6]. Chapter 1 looks into the next decade providing an overview of emerging magnet technologies that enable the generation of homogenous magnetic fields higher than 23.5 T and the great advances that these ultra-high-field spectrometers are likely to provide for biomolecular NMR in solids and in the solution state.

It was recently realized that residual magnetization generated during the preparation phase of an NMR experiment but discarded during the detection phase, the so-called afterglow, could be further manipulated post acquisition and detected during subsequent periods. Thus, advances in modern spectrometer design have allowed the deployment of multiple acquisition periods during the course of a single experiment. Application of this approach in protein NMR in the solid state has been provided in Chapter 2 and a detailed protocol for the optimization of these afterglow experiments has been provided in Chapter 3.

Solid-state NMR provides the advantage of being applicable to large biomolecular assemblies such as intact viruses. Chapter 4 provides detailed approaches in optimizing intact virus particles for solid-state NMR studies. Chapter 5 provides a description of novel membranous assemblies based on inert nanoparticles that enable solid-state NMR studies of membrane proteins especially those whose conformations are sensitive to membrane curvature. Solid-state NMR spectra of large molecular assemblies are complicated and benefit from computational tools not only to simplify analyses but also to aid in the choice of optimal labeling schemes for spectral simplification. Chapter 6 provides an example of such a tool. Additionally, solid-state NMR studies of many large assemblies including viruses are generally hindered by sensitivity issues due to sample concentration and/or stability, thus preventing the application of several experiments in the NMR toolbox in an optimal fashion. Dynamic nuclear polarization (DNP) that utilizes electrons to polarize nuclear spins provides an excellent means to obtain large improvements in sensitivity, thus overcoming the drawbacks of low sample concentrations or speeding up the acquisition of data with optimal sensitivity in rapidly deteriorating samples. DNP has found application both in the solid state (Chapter 7) and in solution (so-called “dissolution” DNP, Chapter 8).

A great strength of NMR spectroscopy that sets it apart from a variety of other structural techniques is its ability to quantify the dynamics of proteins on timescales that range from the fast (picoseconds to nanoseconds) to the slow (microsecond to millisecond) to the very slow (seconds and slower). Over the years, many methods have emerged to study these dynamics using the measurement of NMR spin-relaxation rates. Generally, in order to accurately quantify motional modes in the fast regime, relaxation rates have to be measured at multiple static field strengths in order to accurately interpret the dynamics in a site-specific fashion. In Chapter 9, the authors describe an approach that utilizes the fringe field of a standard high-field NMR magnet together with some novel engineering to measure relaxation rates that extend over three orders of magnitude in magnetic field strength. Chapter 10 provides detailed protocols on measuring dynamics that occur on the slow regime and generally involve large-scale conformational changes. These conformational changes involve activation barriers that lead to sampling of so-called “excited states.” These states, which often have relevance to folding and function, are sparsely populated (often less than 1%) and are thus invisible to conventional techniques. Chapter 11 describes how the structures of these excited states may be accessed using specialized relaxation measurements. Additional effects such as the enhancement of relaxation rates due to close albeit transient encounters with a paramagnetic center (paramagnetic relaxation enhancement, PRE) may also provide a means to probe these lowly populated states. This approach is described in Chapter 12.

An essential part of modern NMR spectroscopy is the processing of raw data obtained from the spectrometer using a variety of mathematical operations, e.g., digital filtering of

solvent, apodization etc. prior to Fourier transformation, and the subsequent analysis of the processed data, e.g., resonance assignment and often the extraction of constraints for the generation of atomic models. Chapters 13 and 14 describe various aspects of data analysis and structure calculation, respectively, using two of the most widely used software platforms. While most common NMR experiments still utilize the Fourier transform that requires digital sampling of the time-domain signal in linear fashion, i.e., in equally spaced intervals, it has been realized that nonlinear sampling often provides significant advantages in some cases. Many functional aspects of nonlinear sampling approaches are described in Chapter 15. Chapter 16 describes a novel computational means that relies on specific correlations obtained in protein NMR spectra to simplify the resonance assignment procedure in large proteins with a significant amount of spectral crowding where conventional approaches fail to resolve ambiguities.

As NMR studies are extended to larger systems, the complexity of NMR spectra, the suboptimal performance of many experiments, and loss of information due to the use of certain essential labeling schemes, e.g., replacement of the information-rich ^1H nuclei by ^2H , lead to incomplete or sparse data that are insufficient to determine structures *de novo*. In such cases NMR data has to be integrated with that from orthogonal techniques, e.g., X-Ray or neutron scattering techniques (SAXS, SANS), to obtain structural information using a hybrid strategy. Chapter 17 describes a computational methodology that integrates NMR data with those obtained from SAXS to obtain atomic detail information on macromolecular complexes.

Ever since its discovery in the 1950s from the work of Proctor and Yu [7], Dickinson [8], Hahn [9], and Arnold [10], the chemical shift has played a central role in chemical, and indeed in biomolecular NMR as an excellent indicator of local structural environment and its changes upon molecular interaction. It was recently realized that a survey of correlated chemical shift changes in response to discrete perturbations, e.g., binding of specific ligands, provides insight into pathways of connectivity and/or allosteric networks in proteins. A detailed method to perform such an analysis is described in Chapter 18.

The final two chapters of this volume deal with two major areas of the recent application of biomolecular NMR spectroscopy—Chapter 19 provides a step-by-step guide to the expression, purification, and preparation of highly homogenous samples of a G-protein coupled receptor (GPCR) for studies using ^{19}F NMR. GPCRs are of immense medical importance and NMR provides an avenue to probe their regulation by agonists and antagonists given that many relevant conformational states are dynamic and/or transient and therefore difficult to characterize by other means. Traditionally, most biomolecular NMR has relied on proteins or protein complexes purified to homogeneity and studied in a suitable buffer. This scenario represents a significant departure from the heterogeneous cellular milieu which may have a substantial influence on physicochemical properties of the protein or protein complex in question by altering its “quinary” structure. Chapter 20 describes recent efforts to extend NMR studies to a more physiological setting using in-cell methodology.

Given the broad applications of modern biomolecular NMR spectroscopy of proteins, it is not possible to compile an exhaustive compendium of all approaches and methodologies. However, based on the selection here, I have attempted to provide the reader a flavor of various aspects of modern protein NMR spectroscopy. I expect that the detailed step-by-step protocols provided will be of great use to specialists and also to nonspecialists who would like to add NMR spectroscopy to their repertoire of experimental tools.

I would like to express my heartfelt gratitude to the authors of the volume for taking the time out of their busy schedules to contribute and to the series editor Dr. John Walker and Springer for giving me the opportunity to put this volume together. I would also like to acknowledge research support from the National Science Foundation (MCB 1412007) and the National Institutes of Health (GM 084278, AI 101792) for research support.

New York, NY, USA

Ranajeet Ghose

References

1. Rabi II (1937) Space quantization in a gyrating magnetic field. *Phys Rev* 51:652–654
2. Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. *Phys Rev* 69:37–38
3. Bloch F (1946) Nuclear induction. *Phys Rev* 70:460–474
4. Saunders M, Wishnia A, Kirkwood JG (1957) The nuclear magnetic resonance spectrum of ribonuclease. *J Am Chem Soc* 79:3289–3290
5. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445(7128):618–622
6. Gan Z et al (2017) NMR spectroscopy up to 35.2 T using a 35.2 T using a series-connected hybrid magnet. *J Magn Reson* 284:125–136
7. Proctor WG, Yu FC (1950) The dependence of a nuclear magnetic resonance frequency upon chemical compound. *Phys Rev* 77:717
8. Dickinson WC (1950) Dependence of the F19 nuclear resonance position on chemical compound. *Phys Rev* 77:736–737
9. Hahn EL (1950) Spin echoes. *Phys Rev* 80:580–594
10. Arnold JT, Dharmatti SS, Packard ME (1951) Chemical effects on nuclear induction signals from organic compounds. *J Chem Phys* 19:507

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>xi</i>
1 NMR of Macromolecular Assemblies and Machines at 1 GHz and Beyond: New Transformative Opportunities for Molecular Structural Biology	1
<i>Caitlin M. Quinn, Mingzhang Wang, and Tatyana Polenova</i>	
2 Experimental Aspects of Polarization Optimized Experiments (POE) for Magic Angle Spinning Solid-State NMR of Microcrystalline and Membrane-Bound Proteins	37
<i>T. Gopinath and Gianluigi Veglia</i>	
3 Afterglow Solid-State NMR Spectroscopy	55
<i>Gili Abramov and Nathaniel J. Traaseth</i>	
4 Filamentous Bacteriophage Viruses: Preparation, Magic-Angle Spinning Solid-State NMR Experiments, and Structure Determination	67
<i>Omry Morag, Nikolaos G. Sgourakis, Gili Abramov, and Amir Goldbourt</i>	
5 Spherical Nanoparticle Supported Lipid Bilayers: A Tool for Modeling Protein Interactions with Curved Membranes	99
<i>Erin R. Tyndall and Fang Tian</i>	
6 Rapid Prediction of Multi-dimensional NMR Data Sets Using FANDAS	111
<i>Siddarth Narasimhan, Deni Mance, Cecilia Pinto, Markus Weingarth, Alexandre M.J. J. Bonvin, and Marc Baldus</i>	
7 Strategies for Efficient Sample Preparation for Dynamic Nuclear Polarization Solid-State NMR of Biological Macromolecules	133
<i>Boris Itin and Ivan V. Sergeev</i>	
8 In-Vitro Dissolution Dynamic Nuclear Polarization for Sensitivity Enhancement of NMR with Biological Molecules	155
<i>Yaewon Kim, Yunyi Wang, Hsueh-Ying Chen, and Christian Hilty</i>	
9 Determination of Protein ps-ns Motions by High-Resolution Relaxometry	169
<i>Samuel F. Cousin, Pavel Kadeřávek, Nicolas Bolik-Coulon, and Fabien Ferrage</i>	
10 Characterizing Protein Dynamics with NMR $R_{1\rho}$ Relaxation Experiments	205
<i>Francesca Massi and Jeffrey W. Peng</i>	
11 CPMG Experiments for Protein Minor Conformer Structure Determination	223
<i>Anusha B. Gopalan, D. Flemming Hansen, and Pramodh Vallurupalli</i>	
12 Probing the Atomic Structure of Transient Protein Contacts by Paramagnetic Relaxation Enhancement Solution NMR	243
<i>Vincenzo Venditti and Nicolas L. Fawzi</i>	

13	From Raw Data to Protein Backbone Chemical Shifts Using NMRFX Processing and NMRViewJ Analysis.....	257
	<i>Bruce A. Johnson</i>	
14	Protein Structure Elucidation from NMR Data with the Program Xplor-NIH.....	311
	<i>Guillermo A. Bermejo and Charles D. Schwieters</i>	
15	Practical Nonuniform Sampling and Non-Fourier Spectral Reconstruction for Multidimensional NMR.....	341
	<i>Mark W. Maciejewski, Adam D. Schuyler, and Jeffrey C. Hoch</i>	
16	Covariance NMR Processing and Analysis for Protein Assignment	353
	<i>Bradley J. Harden and Dominique P. Frueh</i>	
17	Structures of Dynamic Protein Complexes: Hybrid Techniques to Study MAP Kinase Complexes and the ESCRT System	375
	<i>Wolfgang Peti, Rebecca Page, Evzen Boura, and Bartosz Różycki</i>	
18	Implementation of the NMR CHEMical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery.....	391
	<i>Stephen Boulton, Rajeevan Selvaratnam, Rashik Ahmed, and Giuseppe Melacini</i>	
19	High-Efficiency Expression of Yeast-Derived G-Protein Coupled Receptors and ¹⁹ F Labeling for Dynamical Studies.....	407
	<i>Libin Ye, Alexander P. Oraziotti, Aditya Pandey, and R. Scott Prosser</i>	
20	Quantitative Determination of Interacting Protein Surfaces in Prokaryotes and Eukaryotes by Using In-Cell NMR Spectroscopy.....	423
	<i>David S. Burz, Christopher M. DeMott, Asma Aldousary, Stephen Dansereau, and Alexander Shekhtman</i>	
	<i>Index</i>	445

Contributors

- GILI ABRAMOV • *Department of Chemistry, New York University, New York, NY, USA*
- RASHIK AHMED • *Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada*
- ASMA ALDOUSARY • *Department of Chemistry, University at Albany, SUNY, Albany, NY, USA*
- MARC BALDUS • *NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- GUILLERMO A. BERMEJO • *Center for Information Technology, National Institutes of Health, Bethesda, MD, USA*
- NICOLAS BOLIK-COULON • *Laboratoire des Biomolécules, Département de Chimie, École Normale Supérieure - PSL Research University, UPMC Univ Paris 06, CNRS, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, Ecole Normale Supérieure, CNRS, Laboratoire des Biomolécules (LBM), Paris, France*
- ALEXANDRE M. J. J. BONVIN • *Computational Structural Biology, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- STEPHEN BOULTON • *Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada*
- EVZEN BOURA • *Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic*
- DAVID S. BURZ • *Department of Chemistry, University at Albany, SUNY, Albany, NY, USA*
- HSUEH-YING CHEN • *Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD, USA*
- SAMUEL F. COUSIN • *Laboratoire des Biomolécules, Département de Chimie, École Normale Supérieure - PSL Research University, UPMC Univ Paris 06, CNRS, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, Ecole Normale Supérieure, CNRS, Laboratoire des Biomolécules (LBM), Paris, France*
- STEPHEN DANSEREAU • *Department of Chemistry, University at Albany, SUNY, Albany, NY, USA*
- CHRISTOPHER M. DEMOTT • *Department of Chemistry, University at Albany, SUNY, Albany, NY, USA*
- NICOLAS L. FAWZI • *Department of Molecular Pharmacology, Physiology, and Biotechnology, Brown University, Providence, RI, USA*
- FABIEN FERRAGE • *Laboratoire des Biomolécules, Département de Chimie, École Normale Supérieure - PSL Research University, UPMC Univ Paris 06, CNRS, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, Ecole Normale Supérieure, CNRS, Laboratoire des Biomolécules (LBM), Paris, France*
- DOMINIQUE P. FRUEH • *Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
- AMIR GOLDBOURT • *School of Chemistry, Tel Aviv University, Tel Aviv, Israel*
- ANUSHA B. GOPALAN • *TIFR Centre for Interdisciplinary Sciences, Narsingi, Hyderabad, India*
- T. GOPINATH • *Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, Minneapolis, MN, USA*

- D. FLEMMING HANSEN • *Division of Biosciences, Institute of Structural and Molecular Biology, University College London, London, UK*
- BRADLEY J. HARDEN • *Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
- CHRISTIAN HILTY • *Chemistry Department, Texas A&M University, College Station, TX, USA*
- JEFFREY C. HOCH • *Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, USA*
- BORIS ITIN • *New York Structural Biology Center, New York, NY, USA*
- BRUCE A. JOHNSON • *Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY, USA; One Moon Scientific, Inc., Westfield, NJ, USA*
- PAVEL KADERÁVEK • *Laboratoire des Biomolécules, Département de Chimie, Ecole Normale Supérieure - PSL Research University, UPMC Univ Paris 06, CNRS, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, Ecole Normale Supérieure, CNRS, Laboratoire des Biomolécules (LBM), Paris, France*
- YAEWON KIM • *Chemistry Department, Texas A&M University, College Station, TX, USA*
- MARK W. MACIEJEWSKI • *Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, USA*
- DENI MANCE • *NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- FRANCESCA MASSI • *Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA*
- GIUSEPPE MELACINI • *Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada; Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada*
- OMRY MORAG • *School of Chemistry, Tel Aviv University, Tel Aviv, Israel*
- SIDDARTH NARASIMHAN • *NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- ALEXANDER P. ORAZIETTI • *Department of Chemistry, University of Toronto, UTM, Mississauga, ON, Canada; Department of Biochemistry, University of Toronto, Toronto, ON, Canada*
- REBECCA PAGE • *Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA*
- ADITYA PANDEY • *Department of Chemistry, University of Toronto, UTM, Mississauga, ON, Canada; Department of Biochemistry, University of Toronto, Toronto, ON, Canada*
- JEFFREY W. PENG • *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN, USA*
- WOLFGANG PETI • *Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ, USA*
- CECILIA PINTO • *NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- TATYANA POLENOVA • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA*
- R. SCOTT PROSSER • *Department of Chemistry, University of Toronto, UTM, Mississauga, ON, Canada; Department of Biochemistry, University of Toronto, Toronto, ON, Canada*

- CAITLIN M. QUINN • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA*
- BARTOSZ RÓZYCKI • *Institute of Physics, Polish Academy of Sciences, Warsaw, Poland*
- ADAM D. SCHUYLER • *Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, USA*
- CHARLES D. SCHWIETERS • *Center for Information Technology, National Institutes of Health, Bethesda, MD, USA*
- RAJEEVAN SELVARATNAM • *Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada*
- IVAN V. SERGEYEV • *Bruker Biospin Corporation, Billerica, MA, USA*
- NIKOLAOS G. SGOURAKIS • *Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, CA, USA*
- ALEXANDER SHEKHTMAN • *Department of Chemistry, University at Albany, SUNY, Albany, NY, USA*
- FANG TIAN • *Department of Biochemistry and Molecular Biology, The Pennsylvania State University, Hershey, PA, USA*
- NATHANIEL J. TRAASETH • *Department of Chemistry, New York University, New York, NY, USA*
- ERIN R. TYNDALL • *Department of Biochemistry and Molecular Biology, The Pennsylvania State University, Hershey, PA, USA*
- PRAMODH VALLURUPALLI • *TIFR Centre for Interdisciplinary Sciences, Narsingi, Hyderabad, India*
- GIANLUIGI VEGLIA • *Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, Minneapolis, MN, USA; Department of Chemistry, University of Minnesota, Minneapolis, MN, USA*
- VINCENZO VENDITTI • *Department of Chemistry, Iowa State University, Ames, IA, USA; Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, USA*
- MINGZHANG WANG • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA; Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA*
- YUNYI WANG • *Chemistry Department, Texas A&M University, College Station, TX, USA*
- MARKUS WEINGARTH • *NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
- LIBIN YE • *Department of Chemistry, University of Toronto, UTM, Mississauga, ON, Canada; Department of Biochemistry, University of Toronto, Toronto, ON, Canada*

Chapter 1

NMR of Macromolecular Assemblies and Machines at 1 GHz and Beyond: New Transformative Opportunities for Molecular Structural Biology

Caitlin M. Quinn, Mingzhang Wang, and Tatyana Polenova

Abstract

As a result of profound gains in sensitivity and resolution afforded by ultrahigh magnetic fields, transformative applications in the fields of structural biology and materials science are being realized. The development of dual low temperature superconducting (LTS)/high-temperature superconducting (HTS) magnets has enabled the achievement of magnetic fields above 1 GHz (23.5 T), which will open doors to an unprecedented new range of applications. In this contribution, we discuss the promise of ultrahigh field magnetic resonance. We highlight several methodological developments pertinent at high-magnetic fields including measurement of ^1H - ^1H distances and ^1H chemical shift anisotropy in the solid state as well as studies of quadrupolar nuclei such as ^{17}O . Higher magnetic fields have advanced heteronuclear detection in solution NMR, valuable for applications including metabolomics and disordered proteins, as well as expanded use of proton detection in the solid state in conjunction with ultrafast magic angle spinning. We also present several recent applications to structural studies of the AP205 bacteriophage, the M2 channel from Influenza A, and biomaterials such as human bone. Gains in sensitivity and resolution from increased field strengths will enable advanced applications of NMR spectroscopy including *in vivo* studies of whole cells and intact virions.

Key words Ultrahigh magnetic fields, Quadrupolar nuclei, Proton detection, Fast magic angle spinning, Biomaterials, Structure determination, Whole cell NMR, TROSY, Nonuniform sampling

1 Perspective

In the past two decades, the field of magnetic resonance has realized astounding advancements. One critical achievement toward progress in the fields of structural biology, materials science, and clinical imaging (to name just a few) has been the development of higher magnetic field strengths [1–4]. Higher magnetic fields yield profound improvements in both sensitivity and resolution.

The greatest drawback of nuclear magnetic resonance (NMR) relative to other analytical methods is relatively low sensitivity. Increasing magnetic field strength is a compelling way to increase

sensitivity, as the signal-to-noise ratio (SNR, i.e., sensitivity) scales approximately as the power of $3/2$ with the magnetic field strength [5]. The implications of improved sensitivity are numerous and include the capability to study increasingly large and complex biological systems (e.g., whole cells and viral particles), the expansion of applications to physiologically relevant nuclei beyond ^1H , ^{13}C , and ^{15}N , such as ^{17}O , ^{23}Na , and ^{35}Cl , as well as practical considerations, for instance shorter experiment times for multi-dimensional experiments [6].

Increased magnetic fields also yield improvements in resolution:

$$\Delta\nu \propto \frac{1}{\gamma B_0} \quad (1)$$

where $\Delta\nu$ is the linewidth. As with improvements in sensitivity, increased resolution allows for the study of more complex systems and visualization of finer structural details.

Beyond significant gains in sensitivity and resolution, increased magnetic field strengths are particularly powerful for the study of quadrupolar nuclei ($I > 1/2$). Quadrupolar nuclei yield very broad spectra (on the order of MHz) due to the strong coupling (expressed by the coupling constant, C_Q) between the nuclear quadrupolar moment and the electric field gradient (EFG). At higher magnetic field strengths, the contributions of the second-order quadrupolar interaction are reduced, leading to narrower spectra [7].

Both NMR and clinical magnetic resonance imaging (MRI) benefit from increased magnetic field strengths. (For reviews of applications of high-magnetic fields to MRI, *see* refs. 8, 9.) Here, we focus on advancements in technology and methodology as applied in the burgeoning field of NMR at field strengths above 1 GHz. We also discuss several recent applications of high-magnetic fields to the study of biological systems.

1.1 Current State of the Art

Numerous technological gains have promoted the advancement of NMR to current levels, where field strengths above 1 GHz (23.5 T) have recently been achieved (Fig. 1). Such advancements have included improvements in superconducting materials, cryo-cooling, and resistive coils. To date, NMR spectrometers up to 1 GHz have utilized low temperature superconducting (LTS) materials such as Nb_3Sn and NbTi . However, 1–1.1 GHz is believed to be the upper limit for magnets produced with only LTS materials due to limitations such as current density and temperature [1]. Recent advancements have focused on the development of dual LTS/HTS (high-temperature superconducting) magnets [1, 3, 10–13], wherein the inner HTS coil(s) are surrounded by outer LTS coil(s) (Fig. 1). The development of HTS

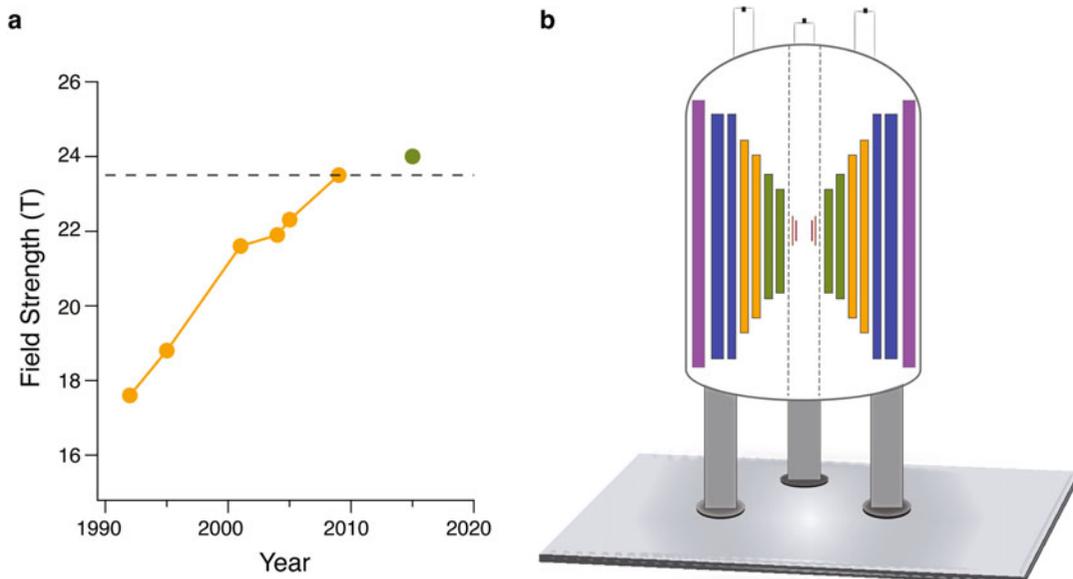


Fig. 1 (a) Advancements in magnetic field strengths for NMR over the last few decades. *Yellow* indicates LTS magnets while *green* indicates dual LTS/HTS magnets. The *dashed line* indicates a frequency of 1 GHz (23.5 T). (b) General schematic of dual LTS/HTS magnet including shim coils and superconducting magnet coils. The precise number of LTS and HTS coils may vary. *Purple* represents superconducting shim coils, while *blue* represents outer NbTi LTS coils, *orange* represents Nb₃Sn inner LTS coils, and *green* represents HTS coils. Within the magnet bore, *maroon* (outer) and *pink* (inner) bars represent the ferromagnetic and room temperature shim coils respectively

materials [14], such as Rare Earth BaCo oxides (REBCO, e.g., YBCO [15, 16]) and Bi-2223 (Bi₂Sr₂Ca₂Cu₃O_{10-x} [17]), has faced challenges including mechanical properties of the materials and manufacturing [18, 19]. HTS technology has very recently advanced to the stage where these materials can be used to generate reasonably stable and homogeneous magnetic fields for NMR magnets. Beyond the development of HTS materials, challenges of the magnet technology have included jointing between the HTS and LTS coils [20] and field homogeneity, further discussed below. Furthermore, the addition of HTS coils creates a spatial challenge. A sufficiently wide bore is needed for biomolecular NMR measurements to be possible. (Bore size is also one of the critical limitations in applying MRI at higher magnetic fields.) One approach to overcoming this issue has been modifications in the geometry of LTS coils so that these take up less space [21].

To date, work on ultrahigh field LTS/HTS magnets for biomolecular NMR applications has focused on proof of concept, with the bulk of published work done on the 24 T magnet at the National Institute for Materials Science (NIMS) developed in collaboration with RIKEN, Kobe Steel, and JEOL (Japan) [1, 22]. The magnet reached 1020 GHz in 2015 using a Bi-2223 inner

HTS coil with outer NbTi and Nb₃Sn LTS coils. With this magnet, they have demonstrated the ability to acquire high-quality natural abundance spectra for insensitive nuclei such as ¹⁷O (0.04% natural abundance [1, 23]). ¹H-¹H-¹H 3D data indicate the field homogeneity and stability achieved at 24 T [24]. As discussed below, a significant advantage of magnetic field strengths over 1 GHz is in the measurement of chemical shift anisotropies (CSAs). The CSA scales with the B_0 field, thus at higher fields CSAs are larger and can be measured more accurately. This is particularly advantageous for small CSAs, such as those of the aliphatic protons. Pandey et al. demonstrated the measurement of ¹H CSAs for L-histidine with the dual LTS/HTS 24 T spectrometer [24].

Beyond superconducting coil technology, for successful operation of ultrahigh field spectrometers advancements have needed to be made with respect to other spectrometer components, concurrent with the development of LTS/HTS magnets. Magnetic field stability and homogeneity are critical to successful experiments. A significant issue in the use of HTS technology for NMR is the inhomogeneous magnetic fields generated by the HTS materials. To compensate for these inhomogeneities in the LTS/HTS systems requires the use of ferromagnetic shim coils in addition to the standard superconducting and room temperature shims used in LTS magnets [25, 26]. Due to residual resistance in the HTS coils and further advancements still needed in HTS/LTS jointing, for stable operation of the 1.02 GHz NMR spectrometer at NIMS, the magnet needed be to run in external current or driven mode [27]. Field fluctuations generated by the DC power supply had to be corrected with an external ²H lock, not typically required for solid-state NMR experiments [28]. Further, use of a DC power supply required development of a new safety mechanism in case of power failure [29]. Additional developments for the use of HTS materials in ultrahigh field magnets have included design of active quench protection [30] and advancements in cryo-cooling and refrigeration [31, 32].

The next stage in the evolution of increased magnetic field strengths for applications in NMR may be hybrid magnets, where an inner resistive coil is surrounded by an outer superconducting coil. NMR spectrometers currently in use utilize only superconducting coils, with 27 T the highest field achieved to date (the instrument setup for this system is not compatible with biomolecular NMR measurements due to the narrow bore size [33]). Challenges in the development of hybrid magnets for applications in NMR include field stability and homogeneity, cost of operation, and size considerations. The highest persistent magnetic field strength from a hybrid magnet to date is the 45 T spectrometer at the National High Magnetic Field Laboratory (NHMFL); however, due to the small bore size, applications are limited to condensed matter physics. The first-in-class hybrid magnet

intended for applications including biomolecular NMR and chemistry is the 36 T Series Connected Hybrid (SCH) magnet being developed at NHMFL [34], which makes use of Florida-Bitter resistive coils [35] and Nb₃Sn superconducting coils. Designing the resistive and superconducting coils to be powered in series rather than in parallel was a significant achievement that greatly reduced the size, in stability, and operating costs of the 36 T magnet relative to the 45 T hybrid magnet. An additional significant achievement of the 36 T magnet is the 40 mm bore size, large enough to accommodate the instrumentation needed for biomolecular NMR experiments. 1 GHz magnets currently produced by Bruker have a standard 54 mm bore. At present efforts are concentrated on achieving the high-field homogeneity needed for NMR experiments.

As mentioned above, increased magnetic field strengths in NMR will have a major impact in a significant number of fields including structural biology, materials science, and imaging. NMR studies of inorganic materials will greatly benefit from increased magnetic field strengths as many inorganic nuclei have large quadrupolar coupling constants and/or low natural abundance. Nuclei of particular interest for inorganic applications that have been shown to benefit from higher magnetic field strengths include ⁶Li ($I = 1$, 7.6% natural abundance (NA) [36, 37]), ¹¹B ($I = 3/2$, 80.1% NA [36, 38]), ¹⁷O ($I = 5/2$, 0.04% NA [38, 39]), ²⁵Mg ($I = 5/2$, 10.0% NA [36, 38]), ²⁷Al ($I = 5/2$, 100% NA [38]), ⁵⁹Co ($I = 7/2$, 100% NA [40, 41]), ⁷³Ge ($I = 9/2$, 7.7% NA [42]), and ¹²⁷I ($I = 5/2$, 100% NA [43, 44]). These elements are essential components of materials in many fields of inorganic chemistry research, including batteries and fuel cells, semiconductors, optical materials, metal organic frameworks (MOFs), catalysts, and glasses. NMR studies of these materials can yield valuable information often inaccessible by other methods including structure, oxidation state, hydrogen bonding environment, molecular motions, and effects of impurities.

Key applications in the field of structural biology that will benefit from development of ultrahigh magnetic fields include membrane proteins in native-like environments, intrinsically disordered proteins such as amyloids and other proteins implicated in neurodegenerative diseases, as well as emergent properties associated with complex environments, such as whole cells and intact viral particles, and bodily fluids (in metabolomics). Increased field strengths will yield enhanced detection of low concentration species with increased chemical shift dispersion. In addition, the vast majority of structural biology work in NMR at present utilizes ¹H, ¹³C, and ¹⁵N nuclei. However, there are many additional NMR active nuclei with critical roles in biology for which work has been limited to date due to sensitivity- and resolution-related challenges, such as ^{35/37}Cl and ²³Na. The group at NIMS in Japan has

demonstrated measurements at 24 T on the spin 3/2 nuclei ^{35}Cl and ^{37}Cl [44, 45], as well as ^{17}O [1, 23]. One of the fields with greatest potential for growth with the advent of magnetic fields over 1GHz is ^{17}O NMR [39, 46], a critical element in biology which could serve as an exquisite probe for a variety of mechanistic and structural questions, including enzyme catalysis, pharmaceuticals analysis, protein structure and hydration [47, 48], as well as materials/MOF applications mentioned above. To date, ^{17}O has not been extensively exploited due to its extremely low natural abundance and the prohibitive cost of isotopic enrichment. In addition to expanding the repertoire of available nuclei, greater sensitivity will improve the ease of studying ^{13}C and ^{15}N at natural abundance (1% and 0.3% respectively). This is particularly advantageous for systems that cannot be readily isotopically labeled, as well as for in vivo studies. In the following sections, we discuss several methods with potentially powerful applications in structural biology at ultrahigh magnetic fields, with regards to achieving the highest quality data, as well as methods to address pressing questions related to protein structure and function. We also highlight several recent notable applications of ultrahigh fields to the study of increasingly complex biological problems.

2 Methods for the Study of Biomolecules at Ultrahigh Magnetic Fields

Ultrahigh magnetic fields benefit the study of macromolecules both by solution NMR and solid-state NMR. At higher fields, the chemical shift dispersion increases linearly. As discussed above, increased field strengths improve both sensitivity and resolution. Narrower linewidths facilitate the atomic resolution structural analysis of large biomolecules, ranging from soluble proteins, to protein assemblies, biological complexes, and membrane-embedded and disordered proteins.

2.1 Sensitivity and Resolution Enhancement

2.1.1 Detection Methods: Solid-State and Solution NMR

With increasing magnetic field strengths and concomitant improvements in sensitivity and resolution, additional detection methods have become available to both solution and solid-state NMR spectroscopists. To obtain maximum sensitivity it is preferable to detect on the highest γ nucleus. Given the large ^1H gyromagnetic ratio, for many biological systems, this means ^1H -detection. Another consequence of proton's high γ is the very strong ^1H - ^1H homonuclear dipolar coupling, which is on the order of 100 kHz:

$$D = -\frac{\mu_0}{4\pi} \frac{\gamma_1 \gamma_2 \hbar}{r_{12}^3} \quad (2)$$

where r_{12} is the distance between the two nuclei. In the absence of motional averaging, the strong ^1H - ^1H dipolar coupling causes severe line broadening. Molecular tumbling in solution averages

the dipolar coupling, yielding narrow proton lines. Thus, in solution NMR, ^1H -detection is the standard approach. However, in the solid state, ^1H lines are very broad. Therefore, magic angle spinning (MAS) NMR experiments traditionally have relied on the detection of ^{13}C and ^{15}N to obtain significantly higher resolution [49–51]. However, this approach sacrifices sensitivity given the smaller gyromagnetic ratios of ^{13}C and ^{15}N . With heteronuclear detection in the solid state a large amount of isotopically labeled sample (typically 10–20 mg) and long data acquisition times are required to compensate for the low sensitivity [49, 52].

Given the resolution enhancement obtained at ultrahigh magnetic fields and the development of fast MAS probes (discussed below), it has become feasible to obtain high-sensitivity solution-like ^1H -detected spectra in the solid state. At MAS rates below ~ 100 kHz a degree of ^1H dilution is typically still required to eliminate broadening from residual ^1H - ^1H couplings (*vide infra*). A great deal of work toward the execution of ^1H detection in the solid state has focused on achieving the optimum level of deuteration for the best combination of sensitivity and resolution. Early work indicated that at moderate MAS rates (~ 10 – 30 kHz) relatively high levels of deuteration were required [53–55]. Oschkinat and coworkers demonstrated that there is no difference in ^1H linewidths for samples with 10 or 100% $^1\text{H}^{\text{N}}$ back exchange at fast MAS rates (>60 kHz) and 1 GHz magnetic field [56]. Pintacuda and coworkers recently established for two proteins (GB1, 56 residues and AP205CP, 130 residues) that with 100 kHz MAS and ultrahigh fields, protein structure determination from ^1H -detected data sets of fully protonated proteins is feasible [57].

In recent years, heteronuclear detection has become increasingly utilized in solution NMR [58, 59]. Heteronuclear chemical shifts are sensitive to different dynamics timescales than protons, allowing researchers to gain access to information that may be absent in ^1H -detected experiments. Heteronuclear detection eliminates issues associated with solvent exchange, an important consideration at or near physiological conditions [60]. Furthermore, heteronuclear chemical shifts often have greater dispersion than ^1H shifts, an effect that becomes more pronounced at higher magnetic fields and is particularly advantageous for disordered systems [61]. Under certain conditions, heteronuclear detection may not result in loss of sensitivity compared to ^1H detection [62].

2.1.2 Fast Magic Angle Spinning

With significant advancements in probe and coil design, it is currently possible to spin samples at rates up to ~ 110 kHz. Fast MAS is a powerful tool for proton detection in bimolecular systems, which in conjunction with ultrahigh fields leads to further improvement in spectral resolution. Fast magic angle spinning is able to more efficiently average and at upward of 100 kHz effectively eliminate

the strong ^1H - ^1H homonuclear dipolar coupling. This leads to substantially narrowed ^1H linewidths, making ^1H -detected experiments feasible in the solid state and with recent advances, eliminating requirements for deuteration. Critically, sensitivity losses due to the decreased sample amounts necessitated by the small rotor sizes are compensated by the increased sensitivity of ^1H detection and high-magnetic fields [63]. Reinstra and coworkers first demonstrated proton-detected spectra of fully protonated GB1 at 40 kHz MAS and a 750 MHz magnetic field strength [64]. Subsequently, Marchetti et al. reported the resolution enhancement achieved by increasing the magnetic field strength to 1 GHz with 60 kHz MAS, illustrated by dipolar-based ^1H - ^{15}N CP-HSQC spectra of fully protonated single-stranded-DNA binding (SSB) protein from *E. coli*. They also determined the unambiguous assignments of this medium-sized protein (18 kDa per monomer) with a set of 2D and 3D proton-detected ^1H - ^{15}N , ^1H - ^{13}C correlation experiments [65]. Weingarth and coworkers demonstrated a labeling scheme dubbed inverse fractional deuteration (iFD) to obtain high-resolution ^1H -detected spectra for non-exchangeable protonation sites, such as membrane proteins [66]. Recently, the combination of fast MAS (100 kHz) and proton detection at ultrahigh field was applied to a range of structurally diverse, fully protonated proteins including GB1, the bacteriophage coat protein AP205CP, and amyloid-forming HET-s [57, 67]. The introduction of fast MAS permits the exploitation of fully protonated proteins for structural and dynamics investigations of biological macromolecules with ^1H detection. Working with protonated proteins is preferred, not only for cost and time savings, but also for the ability to access ^1H - ^1H distances, which can provide additional long-range information among side-chains to improve accuracy of protein structure determination.

2.1.3 Nonuniform Sampling

One approach for resolution or sensitivity enhancement is nonuniform sampling (NUS). Traditionally, in uniformly sampled data sets, points acquired in indirect dimensions are evenly spaced with respect to time. In NUS applications, during acquisition, acquired points are distributed nonlinearly according to a chosen sampling scheme. Rather than discrete Fourier transform, NUS data sets are processed using reconstruction algorithms [68]. Nonuniform sampling using random schedules weighted by a decaying function results in bona fide time domain sensitivity enhancements by reducing the number of points acquired, thereby allowing for more transients to be acquired in a given time frame [69–72]. NUS can also be used to obtain resolution enhancement by acquiring out to a longer acquisition time without increasing the number of points. This is a significant advantage given the enhanced resolution (and thus longer FIDs) obtained at higher magnetic fields [73].

With respect to high-magnetic field applications in solution NMR, NUS is particularly valuable for multi-dimensional heteronuclear-detected experiments and challenging applications such as metabolomics [74], disordered proteins [75–77], and in-cell NMR [78], with higher dimensionality experiments offering improved resolution. Maximum entropy reconstruction algorithms have been demonstrated to be robust for these applications [79]. Several important applications of NUS in the solid state relevant for ultrahigh field experiments have also been demonstrated including quadrupolar nuclei (^{17}O MQMAS) [80], protein assemblies [70, 81], and high-dimensionality (e.g., 4D) proton-detected experiments [81–83] to obtain inter-residue ^1H - ^1H distances (methyl-methyl or amide-amide) for structure determination.

2.2 Resonance Assignment Strategies

To determine structures and examine dynamics of biological macromolecules by NMR, the first, essential step is unequivocal resonance assignments of the individual backbone and side-chain atoms in the target protein. Resonance assignments are determined by a set of 2D and 3D (and sometimes higher dimensionality) experiments that provide intra- and inter-residue backbone and side-chain correlations. To complete assignments and overcome challenges of spectral resolution associated with large proteins and complex biological systems, several strategies have been developed. In conjunction with ultrahigh fields, choice of appropriate labeling schemes for sample preparation and novel pulse sequences further contribute to obtaining high-resolution spectra for accurate and rapid resonance assignments.

2.2.1 Labeling Schemes

Many isotope labeling schemes exist to achieve spectral simplification as well as reduced linewidths due to decreased spin diffusion/transverse relaxation. In MAS NMR, precise structure calculations typically require ^{13}C - ^{13}C distance restraints and torsion angles. There are a number of sparse/extensive (as opposed to uniform) ^{13}C labeling schemes that utilize metabolic precursors such as [1,3- ^{13}C]-glycerol, [2- ^{13}C]-glycerol, [1,6- ^{13}C]-glucose, and [2- ^{13}C]-glucose [52, 84–86] to obtain distance restraints of 2–7 Å [52, 87]. In addition, partial deuteration labeling schemes such as RAP, ILV, and SAIL are suitable to reduce homonuclear dipolar couplings and transverse relaxation, and resolve aliphatic ^1H - ^{13}C correlations with high resolution and sensitivity in both solution and solid-state NMR [53, 55]. To obtain increased resolution enhancement of aliphatic protons for tertiary structure determination, Reif and coworkers introduced Reduced Adjoining Protonation (RAP, [88]), in which side-chain protonation is diluted by expressing the protein with ^2H -glucose and a low $\text{H}_2\text{O}/\text{D}_2\text{O}$ ratio (5–15% and 85–95% respectively) to obtain aliphatic ^1H linewidths ~25 Hz. In this approach, re-protonation at exchangeable sites is not needed. ILV labeling schemes, introduced

by Kay and coworkers for selective, stereospecific protonation of methyl groups in Ile, Leu, and Val [89–91], have been extended to Ala, Thr, and Met (for review *see* refs. 92, 93). This labeling scheme allows for the measurement of more ^1H - ^1H NOEs in a deuterated protein and for the study of larger proteins by solution NMR. A cell-free protocol that has been used is stereo-array isotope labeling (SAIL) of methyl, methylene, and aromatic groups with a combination of $^1\text{H}/^2\text{H}/^{12}\text{C}/^{13}\text{C}$ patterns [94]. The cell-free expression protocol prevents isotope scrambling and like ILV-labeling yields narrower lines than uniform ^2H , ^{13}C , ^{15}N -labeling. All of these labeling schemes are fully applicable for experiments at high-magnetic fields.

2.2.2 Pulse Sequence Design

Just a few decades ago, solution NMR studies of biological systems were limited by the size of the target protein (~30 kDa). Since the introduction of Transverse Relaxation Optimized Spectroscopy (TROSY), solution NMR can be applied to analyze targets with molecular weights of up to 100 kDa in solution [95, 96]. At high fields, the transverse relaxation (T_2) is dominated by the chemical shift anisotropy (CSA) and dipole-dipole coupling (DD). The TROSY pulse sequence suppresses T_2 by employing constructive interference between the CSA and dipolar coupling [96]. Thus, the sensitivity enhancement yielded by the TROSY sequence is more pronounced at higher magnetic fields, with the optimal field strength for ^1H -detected TROSY being 1 GHz [97, 98].

While ^1H -detection has been the method of choice in solution NMR, ^{15}N - and ^{13}C -detected pulse sequences have been developed and make use of increased magnetic fields to achieve higher resolution and site-specific information. Takeuchi et al. reported a modified TROSY sequence selective for amide protons (TROSY $^{15}\text{N}_\text{H}$), which benefits from the slow T_2 and overcomes the low sensitivity of ^{15}N [62, 99]. Theoretical simulations indicate maximum sensitivity of TROSY $^{15}\text{N}_\text{H}$ at a strength field of 1.2 GHz, and the narrowest linewidth at 900 MHz (Fig. 2a, [99]), allowing solution NMR spectroscopists to take advantage of higher field strengths where ^1H -detected TROSY suffers from reduced resolution. With this approach, deuterated samples are not necessary and high-resolution spectra can be obtained using a ^{15}N -detected TROSY-HSQC experiment (Fig. 2b, c). Recently, Yoshimura et al. demonstrated a ^{13}C -detected NMR experiment for probing arginine side-chain $^{15}\text{N}^{\eta/\epsilon}$ - $^{13}\text{C}^\zeta$ correlations at high-magnetic field strengths, which facilitates determination of the arginine ionization states indicating conformational changes in the target protein [100]. This method also makes use of J-based cross polarization, a method that has shown particular promise for disordered proteins as well [101, 102].

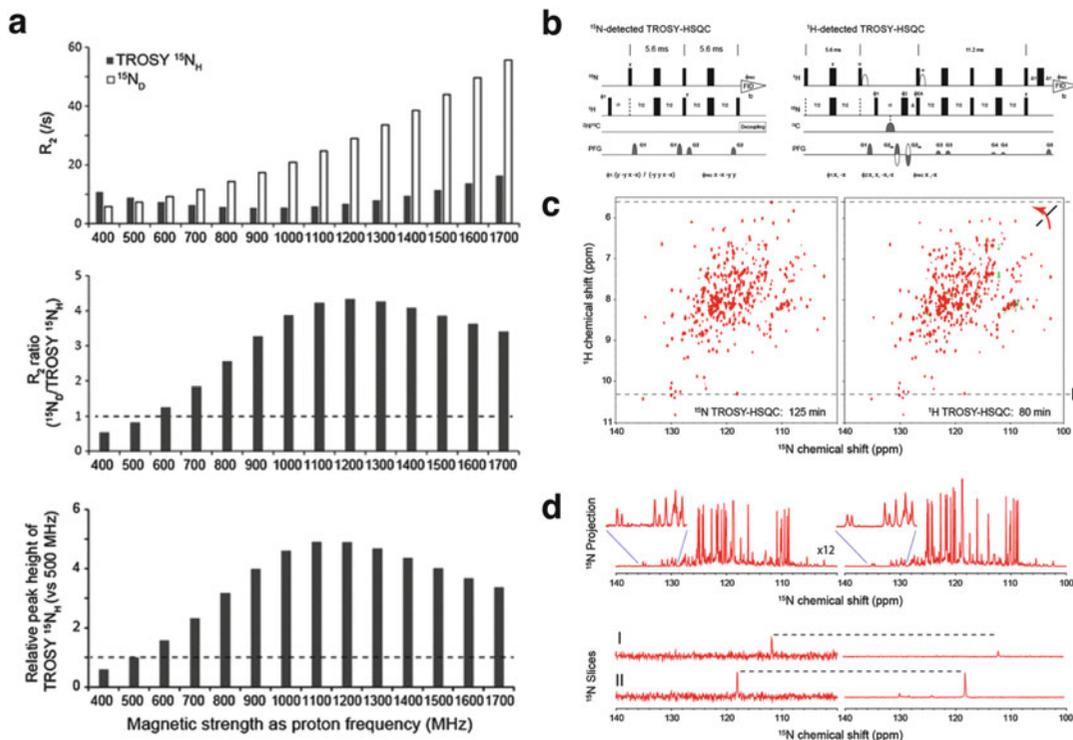


Fig. 2 (a) Simulated plots of TROSY $^{15}\text{N}_\text{H}$ relaxation rates and relative peak heights as a function of magnetic field strength. (b) Pulse sequences for the ^1H - (right) and ^{15}N - (left) detected 2D TROSY-HSQC experiments. (c) Left: ^{15}N -detected TROSY-HSQC, right: ^1H -detected TROSY-HSQC. (d) Top: ^{15}N projection of the 2D ^{15}N -detected TROSY-HSQC (left) and ^1H -detected TROSY-HSQC (right). Bottom: Selected ^{15}N cross sections of the 2D ^{15}N -detected TROSY-HSQC (left) and ^1H -detected TROSY-HSQC (right). Reprinted with permission from Takeuchi, K et al., *J. Biomol. NMR.*, **2015**, *63* (4), 323–331. Copyright 2015 Springer

In the solid state, the development of proton detection in conjunction with fast MAS and ultrahigh fields enables the use of solution-state-derived pulse sequences including 3D ^1H -detected pulse sequences for resonance assignments. ^1H -detected CP-HSQC experiments [103] have been utilized to build 3D ^1H - ^{13}C - ^{15}N correlation experiments, for example the inter-residue (H)CONH and intra-residue (H)CANH experiments [103], with additional ^{13}C - ^{13}C correlation dimensions added for increased information content, e.g., DREAM (dipolar recoupling enhanced by amplitude modulation [104, 105]). At spinning speeds above 40 kHz, dipolar magnetization transfers become less efficient. By employing ^{13}C - ^{13}C J-based coherence transfer blocks, multiple ^{13}C - ^{13}C transfers are achieved in additional experiments for complete backbone resonance assignments: (H)CO(CA)NH, (H)(CO)CA(CO)NH, (H)(CA)CB(CA)NH, and (H)(CA)CB(CACO)NH [63, 106]). For RAP labeled samples, a pair of 3D (H)CCH and HCC(H) experiments were developed by Asami and Reif (Fig. 3, [107]) for assigning $^1\text{H}\alpha$ resonances based on correlations with

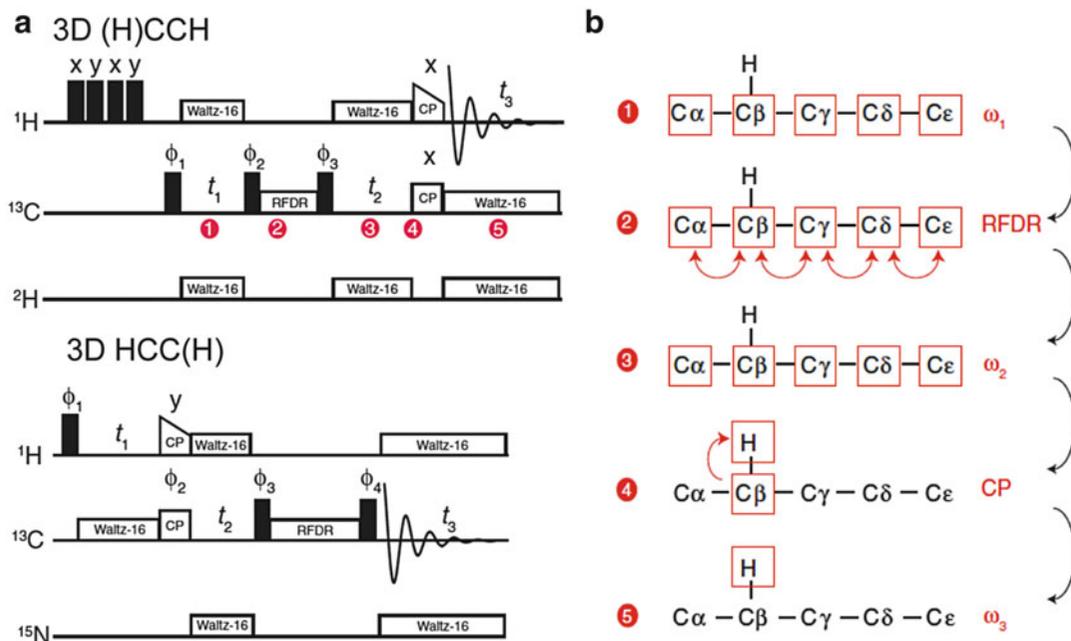


Fig. 3 (a) Pulse sequences of 3D (H)CCH and HCC(H) experiments for RAP-labeled proteins with (b) schematic of magnetization transfers at each step of the pulse sequence. Adapted with permission from Asami, S et al., *J. Biomol. NMR.*, **2012**, 52 (1), 31–39. Copyright 2012 Springer

$^{13}\text{C}\alpha$ and $^{13}\text{CO}/\text{C}\beta$. High resolution was achieved with ^{13}CO and $^{13}\text{C}\beta$ homonuclear scalar decoupling during the $^{13}\text{C}\alpha$ evolution period [108]. An extension of these sequences, (HxCx) $\text{C}\gamma\text{C}\alpha\text{Hx}$ experiments, can also be applied for assignments of ^{13}C -ILV-methyl labeled protonated samples [109].

2.3 Methods for Structural and Dynamics Studies of Biomolecules

Many methods for the structural and dynamics characterization of biomolecules benefit from applications at high and ultrahigh magnetic fields. Certain methods, such as measurement of ^1H CSAs, benefit directly from the higher fields as a result of changes in field-dependent nuclear properties. Other methods become easier to execute at higher fields due to inherent improvements in sensitivity and resolution, such as applications of relaxation and paramagnetism in the solid state. Here, we highlight several methods that have been successfully applied at high and ultrahigh magnetic fields. Given the detailed structural and dynamics information these methods can provide, they will be essential components of ongoing and future work at ultrahigh magnetic fields.

2.3.1 ^1H - ^1H Distance Restraints

Distance restraints are a principal input for structure determination by NMR. In solution NMR structure determination, NOE-derived ^1H - ^1H distances are a key restraint for structure determination, with an upper limit of ~ 5 Å. Until recently, distance restraints in

the solid state were largely limited to heteroatoms such as ^{13}C - ^{13}C and ^{15}N - ^{13}C distances, which can typically probe distances up to ~ 7 Å. High-magnetic fields in conjunction with fast MAS and ^1H detection present the opportunity to obtain accurate, high-resolution ^1H - ^1H distance restraints. ^1H - ^1H restraints can access longer distances (up to ~ 13 Å) and correlations not readily detected otherwise, such as inter-residue side-chain methyl-methyl distances with high-quality multi-dimensional ^1H - ^{15}N and ^1H - ^{13}C spectra and have proven valuable for tertiary structure determination [81]. Early work on small molecules and model proteins utilized moderate MAS rates with heteronuclear detection [110, 111]. Increased MAS rates and the use of deuteration schemes enabled the application of proton-detected ^1H - ^1H distance restraints for structure determination [88, 103, 112]. While previously extensive deuteration and/or selective ^{13}C labeling (e.g., RAP or ILV) had been required, it was recently demonstrated that ^1H - ^1H restraints of up to ~ 5.5 Å can be obtained for uniformly ^1H , ^{13}C , ^{15}N -labeled samples at 1 GHz field with ~ 100 kHz MAS [57]. Numerous schemes for acquiring proton-detected ^1H - ^1H distance restraints at ultrahigh fields with fast MAS have been proposed. 3D (H)XHH with H-H RFDR mixing [113] has been used in structure determination of SOD [114], GB1 [57], and AP205 [57], with higher dimensionality experiments providing enhanced resolution and unambiguous assignments [112]. Meier and coworkers applied DREAM (Dipolar Recoupling Enhanced by Amplitude Modulation) ^1H - ^1H recoupling to obtain methyl-methyl contacts [82], and this was applied for structure determination of M2 [115] and the cytoskeletal assembly BacA [81]. Several groups have applied nonuniform sampling for 4D ^1H - ^1H correlation experiments [81–83]. Based on the success of these solid-state NMR methods for obtaining ^1H - ^1H distance restraints applied to small- and medium-sized proteins, measurements of ^1H - ^1H distance restraints are very promising for applications to complex biological systems at ultrahigh magnetic fields.

2.3.2 Measurement of ^1H Chemical Shift Anisotropy

Hydrogen bonding in proteins is a key determinant of secondary and tertiary structures, and thus is a very valuable parameter for structure determination. The proton chemical shift tensor, and especially its anisotropic component, is highly sensitive to the hydrogen bonding environment [116, 117]. Further, the ^1H CSA is sensitive to dynamics on the submillisecond to microsecond timescales. In solution NMR, knowledge of tensor magnitudes is also required for quantitative relaxation studies in the characterization of protein motions and to optimize cross-correlated relaxation to achieve narrow linewidths. Thus, it is of great interest to establish robust protocols for the measurements of site-specific ^1H chemical shift anisotropies (CSAs) with high resolution. Measurements of

^1H CSAs have presented a challenge for a number of reasons including the strong ^1H - ^1H homonuclear dipolar coupling and the relatively small magnitude of the ^1H CSA tensor. Two approaches for measuring ^1H CSAs under MAS are R-symmetry recoupling sequences [118] and rotary resonance spin-locking [119]. With the use of an appropriate R-symmetry element one can select for the ^1H CSA and ^1H -X heteronuclear dipole while eliminating contributions from ^1H - ^1H homonuclear couplings for measurements in fully protonated proteins. Site-specific CSAs can be determined by incorporating the R-symmetry recoupling as one dimension in a 3D experiment with heteronuclear detection [120, 121]. Two methods for the measurement of ^1H CSAs in the solution are CSA-dipolar cross-correlated relaxation, and residual CSA (RCSA) measurements in aligned media [122–124]. Cross-correlated relaxation measures the interconversion between the CSA and dipolar operators (e.g., H_y and $2H_yN_z$ in the case of amide proton CSAs) from which the chemical shift tensor components (σ_{xx} , σ_{yy} , and σ_{zz}) can be extracted using known formalism [125]. A downside of this approach is that several cross- and autocorrelation rate measurements are needed for the accurate determination of the tensor. The relationship between magnetic field strengths and cross-correlated relaxation rates is discussed below. With RCSA measurements in weakly aligned media, the magnitude and orientation of the CSA tensor can be determined based on changes in the chemical shift relative to an isotropic (i.e., tumbling) sample. Given the relatively small magnitude of the ^1H CSA tensor and the B_0 field dependence of the CSA interaction, executing these measurements at higher magnetic fields allows for increased resolution and accuracy of measurements [24]. With the development of ultrahigh magnetic fields, use of ^1H CSA tensors may become a routine parameter for use in protein structure determination by NMR.

2.3.3 Relaxation and Paramagnetic Effects

In the context of dynamics studies in biological systems by NMR, nuclear spin relaxation parameters provide a wealth of information related to domain flexibility, structural plasticity, and biological functions [126]. Relaxation studies at ultrahigh magnetic fields in the solid state can take advantage of the ability to quickly acquire well-resolved $^{15}\text{N}/^{13}\text{C}$ - ^1H HSQC spectra at several relaxation intervals (e.g., pulse delay or spin-lock pulse length) [127–129]. Furthermore, fast MAS generally applied during these ultrahigh field measurements facilitates the accurate measurement of relaxation rates by reducing coherent contributions to the observed relaxation [130, 131]. MAS NMR R_1 and $R_{1\rho}$ measurements at high fields have been used to characterize site-specific protein backbone dynamics of several microcrystalline proteins [130, 132] and were recently applied for the study of protein-nucleic acid interactions [133].

The application of relaxation measurements at high fields in solution NMR requires careful consideration of how properties of the system, such as the chemical shift timescale, CSA, CSA-dipolar cross correlation, and longitudinal relaxation, are affected by the magnetic field strength. For example, for large systems, ^{15}N longitudinal relaxation rates (R_1) typically decrease at higher magnetic fields [134]. Ishima demonstrated that for ^{15}N R_1 measurements at 900 MHz, the ^1H - ^{15}N dipole/ ^{15}N CSA cross correlation was not sufficiently suppressed in the initial decay affecting accuracy of the R_1 measurements. The aforementioned cross-correlated CSA-dipolar relaxation interference is a critical component of many solution NMR methods including the TROSY and nuclear Overhauser (NOE) techniques and CSA measurements. TROSY line narrowing is dependent on the CSA/DD cross-correlated relaxation, and hence the CSA magnitude and B_0 field [96]. ^{15}N -detected TROSY takes advantage of the cross-correlated relaxation rate at high-magnetic fields (900 MHz–1.2 GHz) to achieve comparable resolution and sensitivity to ^1H -detected TROSY [99], allowing solution NMR spectroscopists to exploit higher magnetic fields without losing sensitivity or resolution due to unfavorable relaxation effects.

Paramagnetic effects provide much long-range structural information for proteins that contain paramagnetic centers, which may be endogenous or synthetically incorporated. The presence of a paramagnetic center may induce four distinct effects: paramagnetic relaxation enhancement (PRE), hyperfine shift (which includes the contact and pseudocontact shifts (PCS)), residual dipolar coupling (RDC), and cross-correlated relaxation (CCR) effects [135]. Paramagnetic centers used in NMR applications include nitroxide radicals, Mn^{2+} , Gd^{3+} , Cu^{2+} , Co^{2+} , and lanthanides [136]. Paramagnetic centers provide angular and distance information to facilitate the determination and further refinement of protein structures, with the distance and orientation dependence of PREs, PCSs, and RDCs used as restraints in NMR structure calculations [137]. Paramagnetic centers have been used to examine many biological systems, such as metalloproteins, cysteine-containing targets attached with a metal tag, and biomolecules modified with a nitroxide spin-label [138–142]. Measuring site-specific PREs requires examination of relaxation decays dependent on the proximity of the paramagnetic center while PCSs are determined by the difference in chemical shifts between a diamagnetic and a paramagnetic species of the target protein [143].

With remarkable progress in proton detection and fast MAS at high-magnetic fields, directly detected ^1H spectra offer high sensitivity and resolution for rapid, quantitative measurements of PREs and pseudocontact shifts in MAS NMR. A further advantage of solid-state NMR studies of paramagnetic systems is the absence of

Curie relaxation, which leads to line broadening in solution, and very fast MAS allows for the observation of signals close to the paramagnetic center [142]. ^1H -detected, fast MAS PRE and PCS measurements of the metalloenzyme superoxide dismutase (SOD) doped with paramagnetic Cu^{2+} or Co^{2+} demonstrate the potential power of ultrahigh fields for the study of paramagnetic systems [128, 143, 144]. The different properties of these two paramagnetic centers made one ideal for PRE measurements (Cu^{2+}) and the other better suited for PCS measurements (Co^{2+} , Fig. 4b–d). With the high-resolution ^{15}N - ^1H CP-HSQC spectra, PREs were obtained from ^{15}N and ^{13}CO R_1 relaxation rates for distances of up to ~ 20 Å (Fig. 4a). High sensitivity enabled the use of 3D experiments for unambiguous assignment of the PCSs. These paramagnetic restraints were combined with ^1H - ^1H distances to subsequently determine a well-defined backbone geometry and metal binding sites in human SOD (Fig. 4e–h). It has also been demonstrated that ^1H T_2 relaxation rates can be used to obtain distance restraints of up to 32 Å [145]. With access to spectrometers at and beyond 1 GHz, site-specific PRE and PCS measurements will be a significant contribution to structural and dynamics investigations of complex biological environments and high molecular weight macromolecules.

2.3.4 *Quadrupolar Nuclei*

Quadrupolar nuclei (spin $I > 1/2$) account for a significant number of NMR active nuclei and these elements, including Li, B, O, Na, Al, Cl, and Ca, are essential components of biology, materials science, and pharmaceuticals. NMR studies of quadrupolar nuclei can yield information such as the chemical and hydrogen bonding environment and molecular motions. The study of quadrupolar nuclei with NMR has been a challenge due to line broadening induced by the strong quadrupolar couplings and the very broad (MHz) spectra limiting sensitivity and making it difficult to obtain uniform excitation. Higher magnetic fields are beneficial in the study of quadrupolar nuclei because the second order quadrupolar coupling, which typically cannot be eliminated by MAS alone, becomes smaller; hence, observed central transition lines are narrower. Improved sensitivity is also beneficial, particularly for low- γ nuclei.

Early methods for the study of quadrupolar nuclei in the solid state included double frequency sweeps (DFS), rotor-assisted population transfer (RAPT), hyperbolic secant (HS) pulses, and double rotation (DOR) (for review *see* ref. 146 and references therein). A significant advancement in the study of quadrupolar nuclei in the solid state was the development of multiple quantum magic angle spinning (MQMAS [147]), which removes the quadrupolar broadening, allows for the resolution of different species by retaining the isotropic shift, and enables quadrupole to spin-1/2 correlations

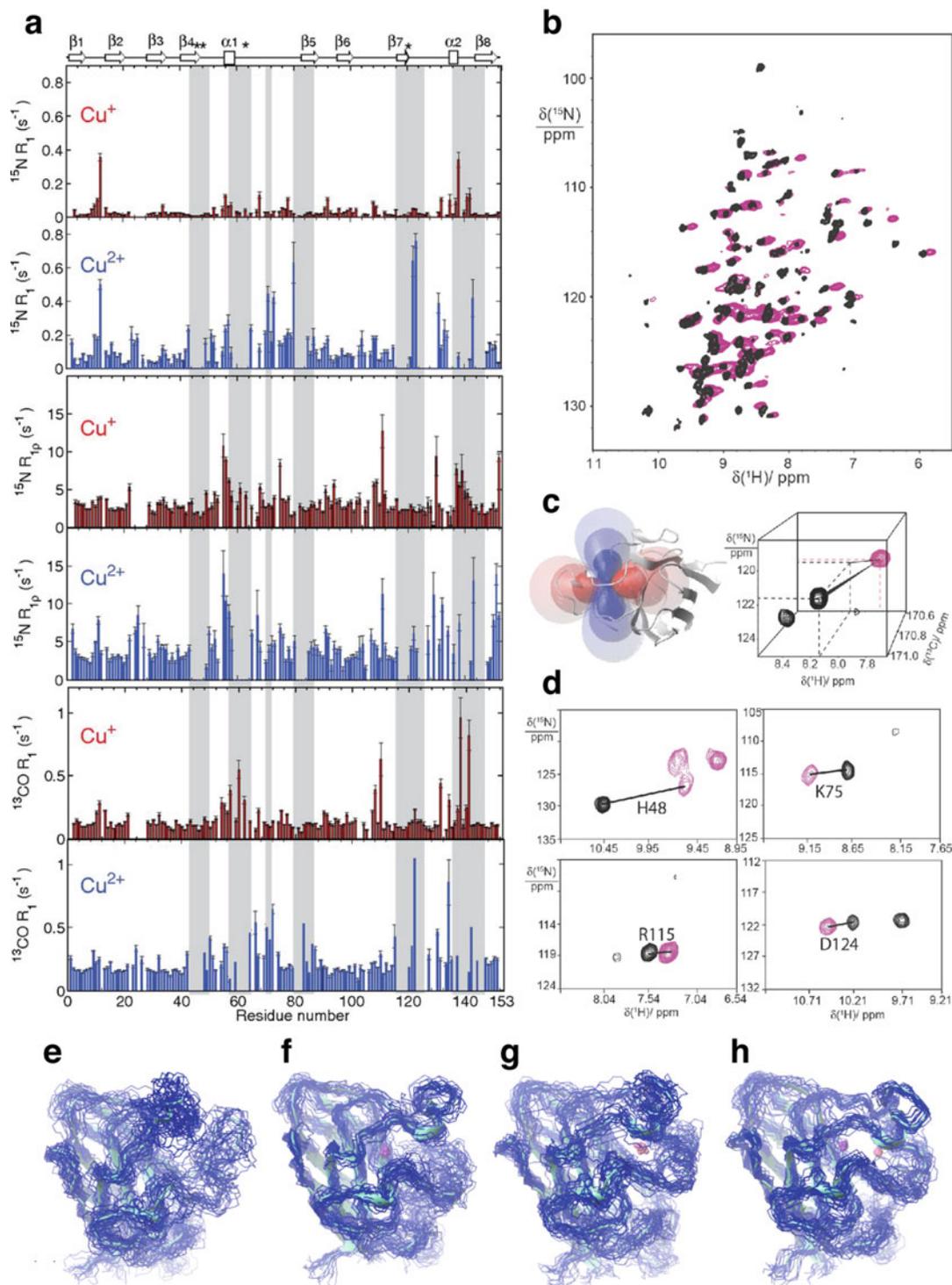


Fig. 4 Paramagnetic effects in $\text{Cu}^{2+}/\text{Co}^{2+}$ loaded SOD by MAS NMR. **(a)** ^{15}N and ^{13}CO R_1 and R_{1p} relaxation rates for Cu^+ - (red, diamagnetic) and Cu^{2+} - (blue, paramagnetic) bound SOD [144]. **(b)** CP-HSQC spectra of Zn^{2+} - (black, diamagnetic) and Co^{2+} - (purple, paramagnetic) bound SOD [144]. **(c)** Left: anisotropic

[148]. Though sensitivity can be limited, the method is more robust and easier to implement than others [149, 150]. When very broad quadrupolar lines are present, static methods can be a better choice than MAS. Static methods for ultra-wideline measurement of quadrupolar nuclei include quadrupolar echo and QCPMG (quadrupolar Carr-Purcell-Meiboom-Gill [151, 152]) which uses refocusing pulses during acquisition to record the spin echo and recapitulates the NMR lineshape. Typically, several frequency steps (i.e., several spectra at different transmitter frequencies) are required to record the full spectrum. Alternatively, Schurko and coworkers have demonstrated the use of adiabatic pulses to excite a broader region of the spectrum, dubbed adiabatic WURST (wide-band uniform-rate smooth truncation [153]) in combination with QCPMG to obtain broadband excitation and high sensitivity [154, 155].

Quadrupolar relaxation can be exploited for solution NMR structure and dynamics studies of quadrupolar nuclei [156, 157]. One approach that exploits quadrupolar relaxation and has been successful for the study of quadrupolar nuclei in the context of biological macromolecules (including ^{17}O , ^{27}Al , and ^{51}V) is quadrupole central transition (QCT) NMR [48, 158–160], which utilizes relaxation properties of the central transition at high-magnetic fields (i.e., the slow tumbling regime) to obtain narrow lines. The combination of high-magnetic fields and narrow lines allows for the observation of fine structural details and applications to larger biological systems [158], as demonstrated in ^{17}O QCT experiments to study the catalytic mechanism of the enzyme Tryptophan synthase [48].

3 Applications of Ultrahigh Fields to the Study of Biological Systems

Ultrahigh field NMR spectroscopy is a powerful tool for investigating the structural and dynamic properties of macromolecules and their physiological relevance. In the past decade, numerous studies have convincingly demonstrated that increased magnetic fields bring improved resolution and sensitivity, enabling applications in both solution and solid states, including protein-ligand binding

Fig. 4 (continued) susceptibility tensor for Co^{2+} with respect to SOD structure. *Right*: Identification and assignment of PCS from 3D (H)CONH experiment, color scheme as in (b). (d) Expansions of select PCSs from 3D (H)CONH experiment, color scheme as in (b). (e–h) Structure refinement of SOD [143]. (e) No paramagnetic restraints, (f) with PRE restraints, (g) with PCS restraints, (h) with PRE and PCS restraints. (a) Reprinted with permission from Knight, MJ et al., *Proc. Nat. Acad. Sci.*, **2012**, 46 (9), 2108–2116. (b–d) Reprinted with permission from Knight, MJ et al., *J. Am. Chem. Soc.*, **2012**, 134 (36), 14,730–14,733. Copyright 2012 American Chemical Society. (e–h) Reprinted with permission from Knight, MJ et al., *Acc. Chem. Res.*, **2013**, 46 (9), 2108–2116. Copyright 2013 American Chemical Society

interactions [161, 162], and structure determination of soluble macromolecules [163, 164], cellular and viral assemblies [165], intact virus particles [166], membrane proteins [167], and amyloids [168, 169]. Ultrahigh field NMR has demonstrated promising high-quality results for future atomic-resolution studies of these highly challenging macromolecule targets and others, such as intrinsically disordered proteins (IDPs, [61, 170]) and whole cells [171]. Here, we highlight several remarkable studies benefited by ultrahigh field NMR techniques to solve interesting biological problems, as well as the current state-of-the-art in emerging fields that will benefit greatly from these technological developments.

The development of proton detection techniques under fast MAS frequencies at ultrahigh fields significantly expands the application of MAS NMR to biological complexes and large proteins. In recent years, the Pintacuda group has developed a rapid proton detection protocol (discussed in Subheading 2.2) and applied it to numerous proteins in a range of assembly states, including the metalloenzyme SOD, microcrystalline SH3, $\beta 2m$, sedimented nucleocapsids of AP205, the membrane protein M2 channel, and OmpG [63].

Structure determination using proton detection in fully protonated, uniform ^{13}C , ^{15}N -labeled AP205CP was achieved with MAS frequencies of 100–111 kHz at 1 GHz magnetic field [57]. AP205CP is a dimeric protein comprising the 2.5-MDa viral capsid of a single-stranded RNA bacteriophage. As shown in Fig. 5, a series of remarkably high-resolution ^1H -detected spectra including ^{13}C - ^1H , ^{15}N - ^1H CP-HSQC, and (H)CCH experiments were used to obtain unambiguous assignments of about 78% of backbone protons and 65% of all proton resonances. In 3D (H)CCH experiments, ^{13}C resonances of the aliphatic side chains were correlated to their protons, providing proton side-chain information. 3D (H)CHH_{RFD} spectra were acquired to record the inter-molecular ^1H - ^1H contacts for the final structure calculation (Fig. 5c). Importantly, 104 inter-monomer contacts were detected, which served to define the dimer interface. The AP205CP dimer structure was determined by this approach with 0.5 mg of protein, less than 2 weeks of instrument time, and rapid analysis of the high-quality data. Further, together with this MAS NMR structure of AP205CP, an atomic-resolution model of the AP205 virus-like particle (VLP) was proposed based on the crystal structure of the protein dimer and the cryo-EM (cryo-electron microscopy) structure of the assembled particle [172]. Significantly, this somewhat heterogeneous sample exhibited broader ^1H linewidths than GB1 (0.15–0.2 ppm vs. 0.1 ppm for $^1\text{H}^{\text{N}}$), but a high-resolution structure could still be determined, suggesting the applicability of the approach to other challenging systems.

Andreas et al. reported the MAS NMR structure of a drug-resistant mutation of Influenza A M2_{18–60} in lipid bilayers (S31N

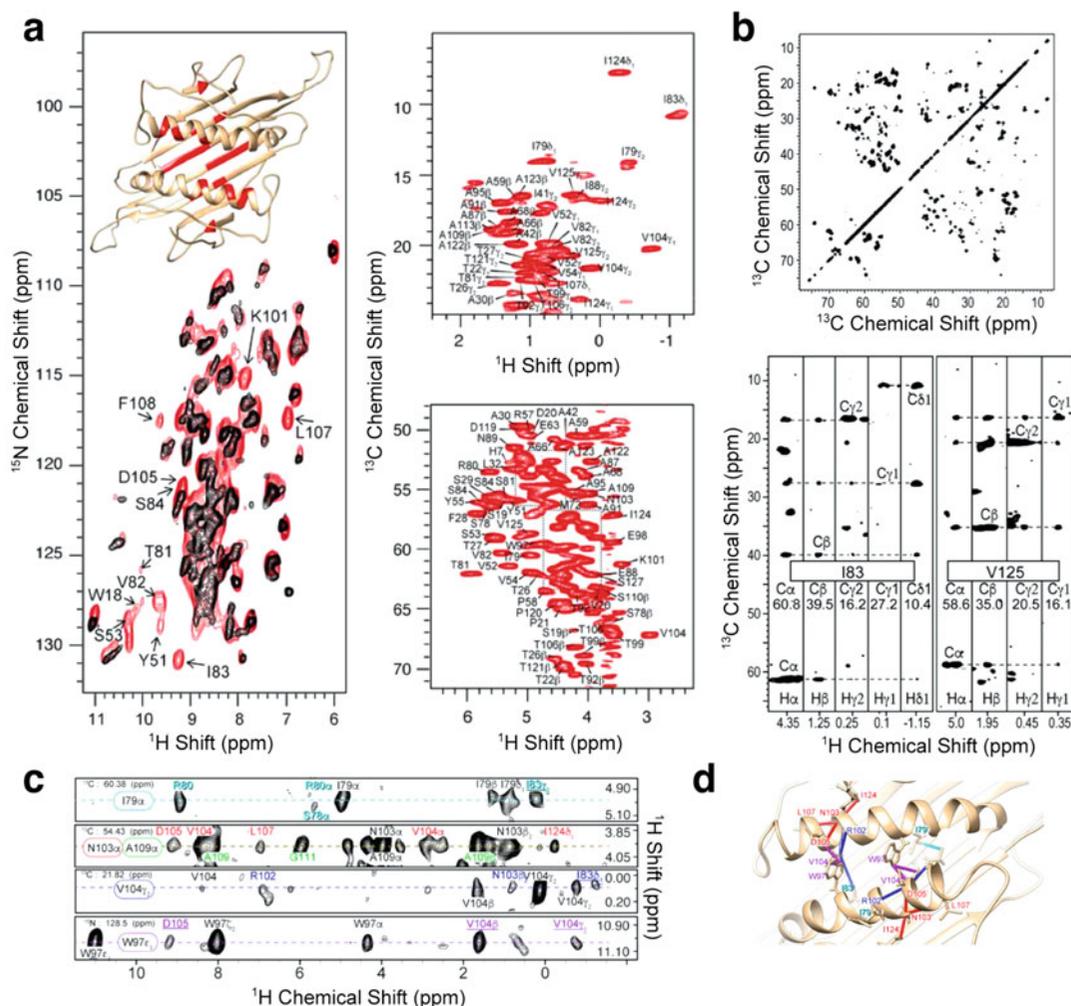


Fig. 5 Structure determination of the viral protein AP205CP assemblies by MAS NMR at ultrahigh field. **(a)** (*left*) A superposition of ^{15}N - ^1H CP-HSQC spectra of (*red*) a fully protonated sample at 100 kHz and (*black*) a perdeuterated, 100% N^{H} sample at 60 kHz MAS frequency. (*right*) Expansions of ^{13}C - ^1H CP-HSQC spectra of AP205CP methyl (*top*) and $\text{H}\alpha$ - $\text{C}\alpha$ (*bottom*) regions at 100 kHz MAS and 1 GHz field. **(b)** ^{13}C - ^{13}C correlation spectrum (*top*) and select strips of a 3D (H)CCH spectrum of AP205CP (*bottom*). **(c)** Selected strips of a 3D (H)CHH spectrum for determination of distance restraints. The intermolecular peaks are *underlined*. **(d)** Expansion of the AP205CP dimer structure showing the restraints extracted from **(c)**. Adapted with permission from Andreas, LB et al., *Proc. Nat. Acad. Sci.*, **2016**, 113 (33), 9187–9192

mutant) [115]. M2 is a tetrameric transmembrane protein that triggers membrane fusion by transporting protons at low pH during viral replication [173, 174]. Understanding the structure and H^+ transport mechanism of this drug-resistant mutant is essential for the development of novel influenza inhibitors. To obtain long-range distance restraints, several samples were prepared with different isotopic labeling schemes for various measurements, including uniform $^{13}\text{C}/^{15}\text{N}$ labeling, extensive 1,6- ^{13}C glucose labeling,

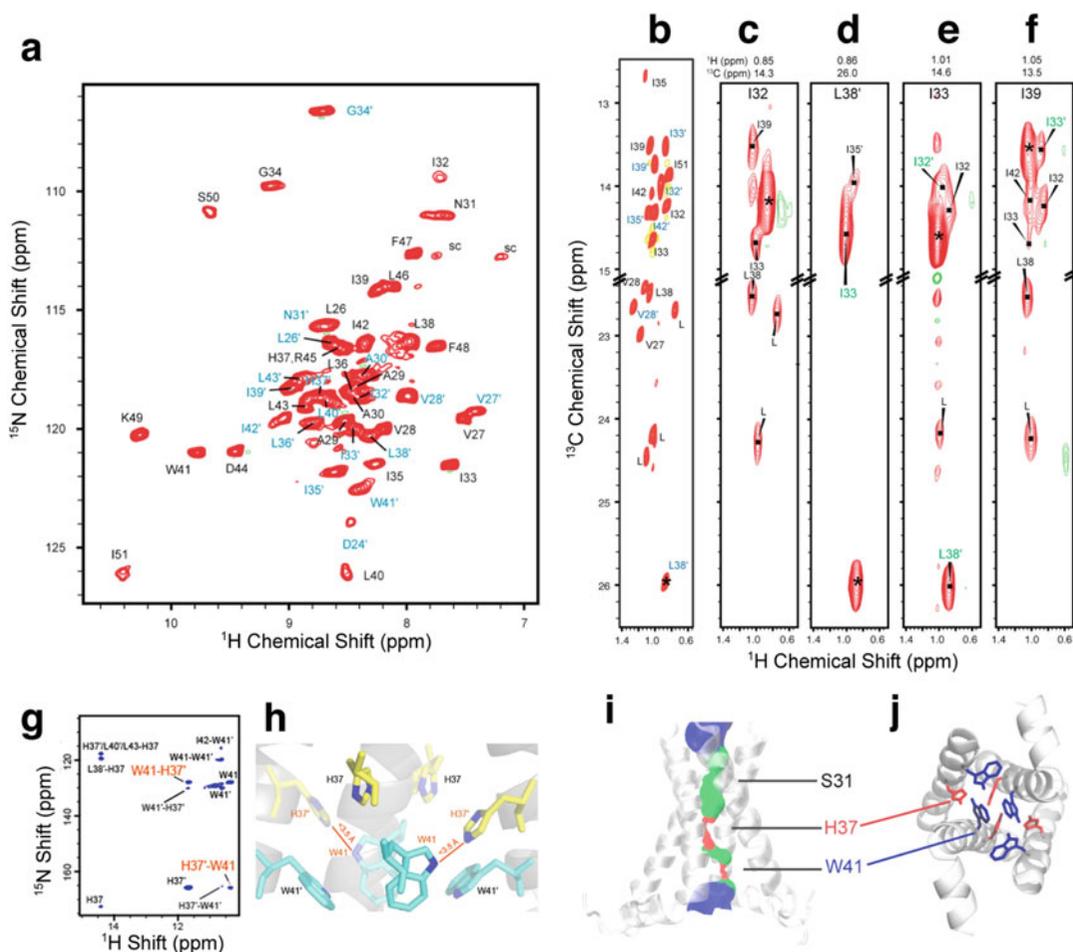


Fig. 6 Structural and mechanistic studies of the membrane protein Influenza A M2 by MAS NMR at ultrahigh field. **(a)** Assigned ^{15}N - ^1H CP-HSQC spectrum recorded at 60 kHz MAS and 1 GHz field. Peak doubling labeled in *black* and *blue* indicates the different conformations for each subunit of the dimer. **(b)** Methyl region of a 2D ^{13}C - ^1H J-based spectrum of ^{13}C , $^2\text{H}_2$, ^1H -ILV labeled M2 and **(c-f)** selected strips of a 4D HCHCHCH spectrum. **(g)** H37'-W41 inter-residue cross peaks from an (H)NHH_{RFDR} experiment. **(h)** Positions of H37 and W41 in the M2 dimer of dimers. **(i)** M2 pore surface indicating water accessibility/pore width: *red* < 1 H₂O, *green* = 1 H₂O, and *blue* > 1 H₂O. **(j)** C-terminal H37 and W41 adopts two different conformations in the dimer of dimers. Adapted with permission from Andreas, LB et al., *J. Am. Chem. Soc.*, **2015**, *137*(47), 14877–14886. Copyright 2015 American Chemical Society

labeling by residue type, deuteration, and ILV-methyl labeling. Using MAS NMR at high and ultrahigh fields (750 MHz–1 GHz), well-resolved 2D and 3D ^{15}N - ^{13}C , ^{13}C - ^{13}C , and a series of ^1H -detected spectra [63] were utilized for resonance assignments of M2. Many residues in the spectra exhibited peak doubling, which indicated M2_{18–60} forms a twofold symmetric dimer of dimers (Fig. 6a). In addition to ^{13}C - ^{13}C and ^{15}N - ^{13}C long-range correlations, ^1H -detected 3D and 4D experiments were performed to measure inter-helical distances necessary for the tetramer structure

determination. 3D and 4D variations of (H)CHH (Fig. 6b–f) and (H)NHH experiments provided intermolecular contacts including methyl-methyl contacts and H37-W41 imidazole-indole contacts (Fig. 6g). The M2 structure was determined with a total of 283 structural constraints. Different side-chain conformations were observed for the gating and pH-sensing residues W41 and H37, indicating that two of the four helices exhibited an “indole in” conformation in one dimer and an “indole out” conformation in the other (Fig. 6h–j). This novel structure suggests that W41 and H37 are likely important in the H⁺ transport process. Furthermore, the conformation at the S31N mutation lends insight into the mechanism of drug resistance. Through this work, a protocol was established for structure determination of membrane proteins and other challenging protein targets at ultrahigh fields.

Barbet-Massin et al. demonstrated the potential to investigate large and complex assemblies with fast MAS at ultrahigh fields [175]. They characterized the structure and dynamics of a native-like Measles virus (MeV) nucleocapsid. The MeV nucleocapsid consists of a macromolecular assembly of a 525-residue protein, which has two domains: an ordered N_{CORE} domain (residues 1–400) and a disordered N_{TAIL} domain (residues 401–525). CP and J-based ¹⁵N-¹H HSQC experiments were used to observe the rigid N_{CORE} and dynamic N_{TAIL} domains respectively. Bulk ¹⁵N *T*_{1ρ} measurements indicated differences in local dynamics between N_{CORE} of intact nucleocapsids and cleaved N_{CORE} domains (where N_{TAIL} is removed by trypsin digestion). An additional advantage of ¹H-detection at ultrahigh fields is the ability to characterize hydration, solvent accessibility, and inter-molecular packing, which in this work detailed the increased disorder of the intact nucleocapsids relative to the cleaved N_{CORE} domain. These studies further demonstrated that sample preparation by sedimentation protocols yields narrow proton linewidths under the experimental conditions utilized (i.e., 1 GHz field and 60 kHz MAS), paving the way for future studies of large, heterogeneous systems such as intact viral particles.

The development of ultrahigh field NMR spectrometers opens further opportunities for the study of emerging phenomena such as whole cells, intact viral particles, and complex biomaterials. These samples are often very insensitive (low natural abundance, low γ nuclei, or isotopically labeled component is a small percentage of the total sample mass), and, given the complexity of *in vivo* systems, spectra are often crowded and suffer from line broadening. Furthermore, many nuclei of interest such as ⁴³Ca are quadrupolar. These challenges can be overcome by the use of ultrahigh field spectrometers, and take advantage of the numerous methodological advances described above.

3.1 Whole Cell NMR

Whole cell NMR enables the study of biological systems under physiological conditions, such as in vivo cell density and effects of binding interactions and other elements of the cellular environment [176]. Challenges to whole cell NMR include sample heterogeneity and cell integrity over the relatively long acquisition times of NMR experiments. To date, only a few examples of in vivo protein structure determination have been presented. The first in-cell structure determination by NMR was of a 66-residue heavy metal binding protein overexpressed in *E. coli* using nonuniform sampling to efficiently acquire data before sample degradation [78]. The same protocol was subsequently applied to the structure determination of GB1 at 250 μM concentration, closer to physiological cellular concentrations [177]. Another recent study presented structure determination of GB1 with a paramagnetic tag for PCS structural restraints in eukaryotic cells [178]. Enhanced sensitivity from ultrahigh magnetic fields will likely increase the efficiency of in vivo structure determination of low concentration proteins. While remarkable results have been accomplished with respect to in vivo studies of disordered proteins [179], ultrahigh fields in combination with ^{13}C detection methods [60, 102] will further benefit this area of research. NMR of IDPs suffers from severe sensitivity and resolution challenges due to dynamics, solvent exchange, and the lack of an ordered secondary structure. Lippens and coworkers presented a method for combined J- and CP-based correlation with ^{13}C detection at 900 MHz for characterization of disordered α -synuclein in *E. coli* cells [102]. Whole cell solid-state NMR studies have covered a wide range of systems including bacteria and biofilms to address many significant questions including perturbations in bacterial cell walls induced by drug binding [180, 181], cellular composition [182], and metabolomics [183], as well as monitoring of protein expression and cellular processes [184–186]. These studies are essential to address pressing issues such as antibiotic resistance [187–189]. In cell MAS NMR studies typically rely on ^{13}C detection and due to the complex, heterogeneous environments resolution is a hindrance to obtaining atomic level information. The expanded use of ultrahigh magnetic fields, associated gains in sensitivity and resolution, and continued methods developments will significantly advance whole cell structural and dynamics studies by NMR.

Biomaterials is a broad discipline with both technological and clinical applications. NMR is a potentially powerful method for characterizing the structure and dynamics of biomaterials. Recent advances in NMR including ultrahigh field spectrometers will further expand the role of NMR in the study of biomaterials. There is great interest in furthering our understanding of the effects of disease and aging on human bone. MAS NMR at ultrahigh fields is uniquely suited to characterize the high-resolution structural features of bone for both high-resolution studies of ^1H nuclei, as

well as improved resolution for studies of quadrupolar nuclei that form essential components of bone such as ^{23}Na and ^{43}Ca . Ramamoorthy and coworkers demonstrated the first ^1H - ^1H correlation spectra of bone at fast MAS (up to 110 kHz [190]) and extracted relaxation parameters to characterize the water distribution in bone. ^{43}Ca (0.14% natural abundance) studies of hydroxyapatite, the primary component of bone, detected two different calcium coordination environments not observed at lower B_0 fields, and determined the NMR parameters (δ_{iso} , C_Q , η_Q) of the two local environments [191] which were subsequently observed in bovine cortical bone at 833 MHz [192]. In addition to ^{43}Ca NMR, Laurencin et al. used ^{23}Na of equine bone and bovine tooth to identify two sodium sites in those materials [193]. For these studies, high-magnetic fields (≥ 800 MHz) were valuable to overcome challenges associated with quadrupolar nuclei. Biosilica materials, such as enzymes encapsulated in or immobilized on an inorganic matrix of silica, are of great interest for their applications in biotechnology and chemical catalysis [194, 195]. Catalytic activity of biosilica-associated enzymes requires conservation of active site structure upon association with the silica network. Stability of the catalytic complex is also essential. It has recently been demonstrated that MAS NMR at ultrahigh fields is a potentially powerful means to characterizing protein structure in these biosilica materials [196]. Luchinat and coworkers showed sufficient resolution can be achieved at fast MAS and high fields (850 MHz) for ^1H -detected studies of biosilica-encapsulated proteins [190]. Another application of biosilica materials, diatoms are unicellular algae with a cell wall composed of silica. Structural features of the diatom cell wall are of interest for a wide range of applications including biotechnology, catalysis, and nanotechnology [197]. ^{29}Si NMR is a common approach for studies of both whole cell and isolated cell wall components of diatoms (*see* ref. 198 for review). While to our knowledge, ultrahigh field studies have not been reported to date, this is another active field of biosilica research that may benefit from greater magnetic field strengths and associated advances.

4 Conclusions

Recent advancements in superconductor technology have made magnetic field strengths over 1 GHz accessible for NMR studies of biomolecular systems. Significant challenges such as field inhomogeneity have needed to be overcome for these instruments to provide atomic level information. Ultrahigh field spectrometers will have a significant impact in a wide range of disciplines including biology and materials science. Numerous methods have been developed that will further enhance the resolution and sensitivity gains that are realized as increased magnetic field strengths. The

pioneering applications highlighted in this work indicate the promise of experiments at ultrahigh fields of 1 GHz and beyond for a variety of complex biomolecules and biomaterials.

References

1. Hashi K, Ohki S, Matsumoto S, Nishijima G, Goto A, Deguchi K, Yamada K, Noguchi T, Sakai S, Takahashi M, Yanagisawa Y, Iguchi S, Yamazaki T, Maeda H, Tanaka R, Nemoto T, Suematsu H, Miki T, Saito K, Shimizu T (2015) Achievement of 1020 MHz NMR. *J Magn Reson* 256:30–33. doi:[10.1016/j.jmr.2015.04.009](https://doi.org/10.1016/j.jmr.2015.04.009)
2. Hashi K, Shimizu T, Goto A, Kiyoshi T, Matsumoto S (2002) Achievement of a 920-MHz high resolution NMR. *J Magn Reson* 156(2):318–321. doi:[10.1006/jmre.2002.2559](https://doi.org/10.1006/jmre.2002.2559)
3. Yanagisawa Y, Nakagome H, Tennmei K, Hamada M, Yoshikawa M, Otsuka A, Hosono M, Kiyoshi T, Takahashi M, Yamazaki T, Maeda H (2010) Operation of a 500 MHz high temperature superconducting NMR: towards an NMR spectrometer operating beyond 1 GHz. *J Magn Reson* 203(2):274–282. doi:[10.1016/j.jmr.2010.01.007](https://doi.org/10.1016/j.jmr.2010.01.007)
4. Bird MD, Bai HY, Bole S, Chen JP, Dixon IR, Ehmler H, Gavrilin AV, Painter TA, Smeibidl P, Toth J, Weijers H, Xu T, Zhai YH (2009) The NHMFL hybrid magnet projects. *IEEE Trans Appl Supercon* 19(3):1612–1616. doi:[10.1109/TASC.2009.2018269](https://doi.org/10.1109/TASC.2009.2018269)
5. Polenova T, Gupta R, Goldbourt A (2015) Magic angle spinning NMR spectroscopy: a versatile technique for structural and dynamic analysis of solid-phase systems. *Anal Chem* 87(11):5458–5469. doi:[10.1021/ac504288u](https://doi.org/10.1021/ac504288u)
6. Ardenkjaer-Larsen JH, Boebinger GS, Comment A, Duckett S, Edison AS, Engelke F, Griesinger C, Griffin RG, Hilty C, Maeda H, Parigi G, Prisner T, Ravera E, van Bentum J, Vega S, Webb A, Luchinat C, Schwalbe H, Frydman L (2015) Facing and overcoming sensitivity challenges in biomolecular NMR spectroscopy. *Angew Chem Int Ed* 54(32):9162–9185. doi:[10.1002/anie.201410653](https://doi.org/10.1002/anie.201410653)
7. Man PP (2000) Quadrupole couplings in nuclear magnetic resonance, general. In: Meyers RA (ed) *Encyclopedia of analytical chemistry*. Wiley, Chichester, pp 12224–12265
8. Budinger TF, Bird MD, Frydman L, Long JR, Mareci TH, Rooney WD, Rosen B, Schenck JF, Schepkin VD, Sherry AD, Sodickson DK, Springer CS, Thulborn KR, Ugurbil K, Wald LL (2016) Toward 20 T magnetic resonance for human brain studies: opportunities for discovery and neuroscience rationale. *Magn Reson Mater Phys* 29(3):617–639. doi:[10.1007/s10334-016-0561-4](https://doi.org/10.1007/s10334-016-0561-4)
9. Niendorf T, Barth M, Kober F, Trattning S (2016) From ultrahigh to extreme field magnetic resonance: where physics, biology and medicine meet. *Magn Reson Mater Phys* 29(3):309–311. doi:[10.1007/s10334-016-0564-1](https://doi.org/10.1007/s10334-016-0564-1)
10. Bascunan J, Lee HG, Bobrov ES, Iwasa Y (2003) A low- and high-temperature superconducting NMR magnet: design and performance results. *IEEE Trans Appl Supercon* 13(2):1550–1553. doi:[10.1109/TASC.2003.812771](https://doi.org/10.1109/TASC.2003.812771)
11. Bascunan J, Hahn S, Park DK, Iwasa Y (2011) A 1.3-GHz LTS/HTS NMR magnet - a progress report. *IEEE Trans Appl Supercon* 21(3):2092–2095. doi:[10.1109/TASC.2010.2086995](https://doi.org/10.1109/TASC.2010.2086995)
12. Bascunan J, Hahn S, Lecrevisse T, Song JB, Miyagi D, Iwasa Y (2016) An 800-MHz all-REBCO insert for the 1.3-GHz LTS/HTS NMR magnet program - a progress report. *IEEE Trans Appl Supercon* 26(4). doi:[10.1109/Tasc.2015.2512045](https://doi.org/10.1109/Tasc.2015.2512045)
13. Yanagisawa Y, Piao R, Iguchi S, Nakagome H, Takao T, Kominato K, Hamada M, Matsumoto S, Suematsu H, Jin X, Takahashi M, Yamazaki T, Maeda H (2014) Operation of a 400 MHz NMR magnet using a (RE:Rare Earth)Ba₂Cu₃O_{7-x} high-temperature superconducting coil: towards an ultra-compact super-high field NMR spectrometer operated beyond 1 GHz. *J Magn Reson* 249:38–48. doi:[10.1016/j.jmr.2014.10.006](https://doi.org/10.1016/j.jmr.2014.10.006)
14. Scanlan RM, Malozemoff AP, Larbalestier DC (2004) Superconducting materials for large scale applications. *Proc IEEE* 92(10):1639–1654. doi:[10.1109/JPROC.2004.833673](https://doi.org/10.1109/JPROC.2004.833673)
15. Matsumoto S, Choi S, Kiyoshi T, Otsuka A, Hamada M, Maeda H, Yanagisawa Y, Nakagome H, Suematsu H (2012) REBCO layer-wound coil tests under electromagnetic forces in an external magnetic field of up to 17.2 T. *IEEE Trans Appl Supercon* 22(3). doi:[10.1109/Tasc.2012.2185024](https://doi.org/10.1109/Tasc.2012.2185024)

16. Hazelton DW, Selvamanickam V, Duval JM, Larbalestier DC, Markiewicz WD, Weijers HW, Holtz RL (2009) Recent developments in 2G HTS coil technology. *IEEE Trans Appl Supercon* 19(3):2218–2222. doi:[10.1109/TASC.2009.2018791](https://doi.org/10.1109/TASC.2009.2018791)
17. Cheng MC, Yan BP, Lee KH, Ma QY, Yang ES (2005) A high temperature superconductor tape RF receiver coil for a low field magnetic resonance-imaging system. *Supercond Sci Technol* 18(8):1100–1105. doi:[10.1088/0953-2048/18/8/013](https://doi.org/10.1088/0953-2048/18/8/013)
18. Parizh M, Lvovsky Y, Sumption M (2017) Conductors for commercial MRI magnets beyond NbTi: requirements and challenges. *Supercond Sci Technol* 30(1). doi:[10.1088/0953-2048/30/1/014007](https://doi.org/10.1088/0953-2048/30/1/014007)
19. Kajita K, Iguchi K, Xu Y, Nawa M, Hamada M, Takao T, Nakagome H, Matsumoto S, Nishijima G, Suematsu H, Takahashi M, Yanagisawa Y (2016) Degradation of a REBCO coil due to cleavage and peeling originating from an electromagnetic force. *IEEE Trans Appl Supercon* 26(4). doi:[10.1109/Tasc.2016.2515564](https://doi.org/10.1109/Tasc.2016.2515564)
20. Brittles GD, Mousavi T, Grovenor CRM, Aksoy C, Speller SC (2015) Persistent current joints between technological superconductors. *Supercond Sci Technol* 28(9). doi:[10.1088/0953-2048/28/9/093001](https://doi.org/10.1088/0953-2048/28/9/093001)
21. Roth G (2015) Ultra-high field magnets at Bruker. In: Polenova T, Budinger TF (eds) *Ultrahigh field NMR and MRI: science at a crossroads*, Bethesda, MD
22. Nishijima G, Matsumoto S, Hashi K, Ohki S, Goto A, Noguchi T, Iguchi S, Yanagisawa Y, Takahashi M, Maeda H, Miki T, Saito K, Tanaka R, Shimizu T (2016) Successful upgrading of 920-MHz NMR superconducting magnet to 1020 MHz using Bi-2223 innermost coil. *IEEE Trans Appl Supercon* 26(3). doi:[10.1109/Tasc.2016.2524466](https://doi.org/10.1109/Tasc.2016.2524466)
23. Hashi K, Deguchi K, Yamazaki T, Ohki S, Matsumoto S, Nishijima G, Goto A, Yamada K, Noguchi T, Sakai S, Takahashi M, Yanagisawa Y, Iguchi S, Maeda H, Tanaka R, Nemoto T, Suematsu H, To J, Torres J, Perushin K, Shimizu T (2016) Efficiency of high magnetic fields in solid-state NMR. *Chem Lett* 45(2):209–210. doi:[10.1246/cl.151063](https://doi.org/10.1246/cl.151063)
24. Pandey MK, Zhang RC, Hashi K, Ohki S, Nishijima G, Matsumoto S, Noguchi T, Deguchi K, Goto A, Shimizu T, Maeda H, Takahashi M, Yanagisawa Y, Yamazaki T, Iguchi S, Tanaka R, Nemoto T, Miyamoto T, Suematsu H, Saito K, Miki T, Ramamoorthy A, Nishiyama Y (2015) 1020 MHz single-channel proton fast magic angle spinning solid-state NMR spectroscopy. *J Magn Reson* 261:1–5. doi:[10.1016/j.jmr.2015.10.003](https://doi.org/10.1016/j.jmr.2015.10.003)
25. Iguchi S, Piao R, Hamada M, Matsumoto S, Suematsu H, Takao T, Saito AT, Li J, Nakagome H, Jin X, Takahashi M, Maeda H, Yanagisawa Y (2016) Advanced field shimming technology to reduce the influence of a screening current in a REBCO coil for a high-resolution NMR magnet. *Supercond Sci Technol* 29(4). doi:[10.1088/0953-2048/29/4/045013](https://doi.org/10.1088/0953-2048/29/4/045013)
26. Iguchi S, Yanagisawa Y, Takahashi M, Takao T, Hashi K, Ohki S, Nishijima G, Matsumoto S, Noguchi T, Tanaka R, Suematsu H, Saito K, Shimizu T (2016) Shimming for the 1020 MHz LTS/Bi-2223 NMR magnet. *IEEE Trans Appl Supercon* 26(4). doi:[10.1109/Tasc.2016.2519516](https://doi.org/10.1109/Tasc.2016.2519516)
27. Otsuka A, Kiyoshi T, Matsumoto S, Kominato K, Takeda M (2008) Field stability of a 600 MHz NMR magnet in the driven-mode operation. *IEEE Trans Appl Supercon* 18(2):852–855. doi:[10.1109/TASC.2008.920558](https://doi.org/10.1109/TASC.2008.920558)
28. Takahashi M, Ebisawa Y, Tennmei K, Yanagisawa Y, Hosono M, Takasugi K, Hase T, Miyazaki T, Fujito T, Nakagome H, Kiyoshi T, Yamazaki T, Maeda H (2012) Towards a beyond 1 GHz solid-state nuclear magnetic resonance: external lock operation in an external current mode for a 500 MHz nuclear magnetic resonance. *Rev Sci Instrum* 83(10). doi:[10.1063/1.4757576](https://doi.org/10.1063/1.4757576)
29. Matsumoto S, Kiyoshi T, Nishijima G, Hashi K, Takahashi M, Noguchi T, Ohki S, Maeda H, Shimizu T (2016) Equipment for power outage in operation of driven-mode NMR magnet. *IEEE Trans Appl Supercon* 26(4). doi:[10.1109/Tasc.2016.2519515](https://doi.org/10.1109/Tasc.2016.2519515)
30. Gavrilin AV, Weijers HW, et al (2016) Comprehensive modelling study of quench behaviour of the NHMFL 32 T all-superconducting magnet system. Input data and methodology aspects. In: *The 5th international workshop on numerical modelling of high temperature superconductors*, Bologna, Italy, 2016
31. Bai HY, Bird MD, Cantrell KR, Dixon IR, Gavrilin AV (2009) Design of cryogenic system for SCH magnets. *IEEE Trans Appl Supercon* 19(3):1596–1599. doi:[10.1109/TASC.2009.2018218](https://doi.org/10.1109/TASC.2009.2018218)
32. Choi YS, Kim DL, Painter TA, Markiewicz WD, Lee BS, Yang HS, Yoo JS (2008) Closed-loop cryogenic cooling for a 21 T FT-ICR magnet system. *IEEE Trans Appl Supercon* 18(2):1471–1474. doi:[10.1109/TASC.2008.922538](https://doi.org/10.1109/TASC.2008.922538)

33. National High Field Magnet Laboratory (2015) MagLab claims record with novel superconducting magnet. <https://nationalmaglab.org/news-events/news/maglab-claims-record-with-novel-superconducting-magnet>
34. Bird MD, Dixon IR, Toth J (2015) Large, high-field magnet projects at the NHMFL. *IEEE Trans Appl Supercon* 25(3):1–6
35. Toth J (2016) Resistive insert magnet designs for the series-connected hybrids developed at the NHMFL. *IEEE Trans Appl Supercon* 26(4). doi:10.1109/Tasc.2015.2510227
36. Hu JZ, Kwak JH, Yang Z, Wan X, Shaw LL (2009) Direct observation of ion exchange in mechanically activated LiH + MgB₂ system using ultrahigh field nuclear magnetic resonance spectroscopy. *Appl Phys Lett* 94(14):1–3
37. Vijayakumar M, Kerisit S, Yang ZG, Graff GL, Liu J, Sears JA, Burton SD, Rosso KM, Hu JZ (2009) Combined ⁶Li, ⁷Li NMR and molecular dynamics study of Li diffusion in Li₂TiO₃. *J Phys Chem C* 113(46):20108–20116
38. Kroeker S, Neuhoff PS, Stebbins JF (2001) Enhanced resolution and quantitation from ‘ultrahigh’ field NMR spectroscopy of glasses. *J Non-Cryst Solids* 293:440–445
39. Wu G (2016) Solid-state ¹⁷O NMR studies of organic and biological molecules: recent advances and future directions. *Solid State Nucl Magn Reson* 73:1–14. doi:10.1016/j.ssnmr.2015.11.001
40. Crewdson P, Bryce DL, Rominger F, Hofmann P (2008) Application of ultrahigh-field ⁵⁹Co solid-state NMR spectroscopy in the investigation of the 1,2-polybutadiene catalyst [Co(C₈H₁₃)(C₄H₆)]. *Angew Chem Int Ed* 47(18):3454–3457. doi:10.1002/anie.200705204
41. Ooms KJ, Terskikh VV, Wasylshen RE (2007) Ultrahigh-field solid-state ⁵⁹Co NMR studies of Co(C₂B₉H₁₁)²⁻ and Co(C₅H₅)²⁺ salts. *J Am Chem Soc* 129(21):6704–6705. doi:10.1021/ja0710095
42. Greer BJ, Michaelis VK, Terskikh VV, Kroeker S (2011) Reconnaissance of diverse structural and electronic environments in germanium halides by solid-state Ge NMR and quantum chemical calculations. *Can J Chem* 89(9):1118–1129
43. Widdifield CM, Bryce DL (2010) Solid-state (127)I NMR and GIPAW DFT study of metal iodides and their hydrates: structure, symmetry, and higher-order quadrupole-induced effects. *J Phys Chem A* 114(40):10810–10823. doi:10.1021/jp108237x
44. Widdifield CM, Chapman RP, Bryce DL (2009) Chlorine, bromine, and iodine solid-state NMR spectroscopy. *Annu Rep NMR Spectrosc* 66:195–326
45. Pandey A, Hashi K, Ohki S, Nishijima G, Matsumoto S, Noguchi T, Deguchi K, Goto A, Shimizu T, Maeda H, Takahashi M, Yanagisawa Y, Yamazaki T, Iguchi S, Tanaka R, Nemoto T, Miyamoto T, Suematsu H, Saito K, Miki T, Nishiyama Y (2016) 24 T high-resolution and -sensitivity solid-state NMR measurements of low-gamma half-integer quadrupolar nuclei ³⁵Cl and ³⁷Cl. *Anal Sci* 32:1339–1345
46. Wong A, Poli F (2014) Solid-state ¹⁷O NMR studies of biomolecules. *Annu Rep NMR Spectrosc* 83:145–220
47. Gerothanassis IP (2010) Oxygen-17 NMR spectroscopy: basic principles and applications. Part II. *Prog Nucl Magn Reson Spectrosc* 57(1):1–110. doi:10.1016/j.pnmrs.2009.12.001
48. Young RP, Caulkins BG, Borchardt D, Bulloch DN, Larive CK, Dunn MF, Mueller LJ (2016) Solution-state ¹⁷O quadrupole central-transition NMR spectroscopy in the active site of tryptophan synthase. *Angew Chem Int Ed* 55(4):1350–1354. doi:10.1002/anie.201508898
49. Sun SJ, Han Y, Paramasivam S, Yan S, Siglin AE, Williams JC, Byeon IJL, Ahn J, Gronenborn AM, Polenova T (2012) Solid-state NMR spectroscopy of protein complexes. *Methods Mol Biol* 831:303–331. doi:10.1007/978-1-61779-480-3_17
50. Igumenova TI, McDermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ (2004) Assignments of carbon NMR resonances for microcrystalline ubiquitin. *J Am Chem Soc* 126(21):6720–6727. doi:10.1021/ja030547o
51. Igumenova TI, Wand AJ, McDermott AE (2004) Assignment of the backbone resonances for microcrystalline ubiquitin. *J Am Chem Soc* 126(16):5323–5331. doi:10.1021/ja030546w
52. 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(6911):98–102. doi:10.1038/nature01070
53. Chevelkov V, van Rossum BJ, Castellani F, Rehbein K, Diehl A, Hohwy M, Steuernagel S, Engelke F, Oschkinat H, Reif B (2003) ¹H detection in MAS solid-state NMR spectroscopy of biomacromolecules employing pulsed

- field gradients for residual solvent suppression. *J Am Chem Soc* 125(26):7788–7789. doi:[10.1021/ja029354b](https://doi.org/10.1021/ja029354b)
54. Paulson EK, Morcombe CR, Gaponenko V, Dancheck B, Byrd RA, Zilm KW (2003) Sensitive high resolution inverse detection NMR spectroscopy of proteins in the solid state. *J Am Chem Soc* 125(51):15831–15836. doi:[10.1021/ja037315+](https://doi.org/10.1021/ja037315+)
 55. Chevelkov V, Rehbein K, Diehl A, Reif B (2006) Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew Chem Int Ed* 45(23):3878–3881. doi:[10.1002/anie.200600328](https://doi.org/10.1002/anie.200600328)
 56. Lewandowski JR, Dumez JN, Akbey U, Lange S, Emsley L, Oschkinat H (2011) Enhanced resolution and coherence lifetimes in the solid-state NMR spectroscopy of perdeuterated proteins under ultrafast magic-angle spinning. *J Phys Chem Lett* 2(17):2205–2211. doi:[10.1021/jz200844n](https://doi.org/10.1021/jz200844n)
 57. Andreas LB, Jaudzems K, Stanek J, Lalli D, Bertarello A, Le Marchand T, Cala-De Paepe D, Kotelovica S, Akopjana I, Knott B, Wegner S, Engelke F, Lesage A, Emsley L, Tars K, Herrmann T, Pintacuda G (2016) Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc Natl Acad Sci USA* 113(33):9187–9192. doi:[10.1073/pnas.1602248113](https://doi.org/10.1073/pnas.1602248113)
 58. Felli IC, Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. *ChemPhysChem* 10(9–10):1356–1368. doi:[10.1002/cphc.200900133](https://doi.org/10.1002/cphc.200900133)
 59. Takeuchi K, Gal M, Shimada I, Wagner G (2012) Low-gamma nuclei detection experiments for biomolecular NMR, Chapter 2. In: Clore M, Potts J (eds) *Recent developments in biomolecular NMR*, vol 25. Royal Society of Chemistry, Cambridge, pp 25–52. doi:[10.1039/9781849735391-00025](https://doi.org/10.1039/9781849735391-00025)
 60. Felli IC, Gonnelli L, Pierattelli R (2014) In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat Protoc* 9(9):2005–2016. doi:[10.1038/nprot.2014.124](https://doi.org/10.1038/nprot.2014.124)
 61. Felli IC, Pierattelli R (2014) Novel methods based on ^{13}C detection to study intrinsically disordered proteins. *J Magn Reson* 241:115–125. doi:[10.1016/j.jmr.2013.10.020](https://doi.org/10.1016/j.jmr.2013.10.020)
 62. Takeuchi K, Arthanari H, Imai MK, Wagner G, Shimada I (2016) Nitrogen-detected TROSY yields comparable sensitivity to proton-detected TROSY for non-deuterated, large proteins under physiological salt conditions. *J Biomol NMR* 64(2):143–151. doi:[10.1007/s10858-016-0015-3](https://doi.org/10.1007/s10858-016-0015-3)
 63. Barbet-Massin E, Pell AJ, Retel JS, Andreas LB, Jaudzems K, Franks WT, Nieuwkoop AJ, Hiller M, Higman V, Guerry P, Bertarello A, Knight MJ, Felletti M, Le Marchand T, Kotelovica S, Akopjana I, Tars K, Stoppini M, Bellotti V, Bolognesi M, Ricagno S, Chou JJ, Griffin RG, Oschkinat H, Lesage A, Emsley L, Herrmann T, Pintacuda G (2014) Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J Am Chem Soc* 136(35):12489–12497. doi:[10.1021/ja507382j](https://doi.org/10.1021/ja507382j)
 64. Sperling LJ, Nieuwkoop AJ, Lipton AS, Berthold DA, Rienstra CM (2010) High resolution NMR spectroscopy of nanocrystalline proteins at ultra-high magnetic field. *J Biomol NMR* 46(2):149–155. doi:[10.1007/s10858-009-9389-9](https://doi.org/10.1007/s10858-009-9389-9)
 65. Marchetti A, Jehle S, Felletti M, Knight MJ, Wang Y, Xu ZQ, Park AY, Otting G, Lesage A, Emsley L, Dixon NE, Pintacuda G (2012) Backbone assignment of fully protonated solid proteins by ^1H detection and ultrafast magic-angle-spinning NMR spectroscopy. *Angew Chem Int Ed* 51(43):10756–10759. doi:[10.1002/anie.201203124](https://doi.org/10.1002/anie.201203124)
 66. Medeiros-Silva J, Mance D, Daniëls M, Jekhmane S, Houben K, Baldus M, Weingarth M (2016) ^1H -detected solid-state NMR studies of water-inaccessible proteins in vitro and in situ. *Angew Chem Int Ed* 55(43):13606–13610. doi:[10.1002/anie.201606594](https://doi.org/10.1002/anie.201606594)
 67. Stanek J, Andreas LB, Jaudzems K, Cala D, Lalli D, Bertarello A, Schubeis T, Akopjana I, Kotelovica S, Tars K, Pica A, Leone S, Picone D, Xu ZQ, Dixon NE, Martinez D, Berbon M, El Mammeri N, Noubhani A, Saupe S, Habenstein B, Loquet A, Pintacuda G (2016) NMR spectroscopic assignment of backbone and side-chain protons in fully protonated proteins: microcrystals, sedimented assemblies, and amyloid fibrils. *Angew Chem Int Ed* 55(50):15503–15509. doi:[10.1002/anie.201607084](https://doi.org/10.1002/anie.201607084)
 68. Hyberts SG, Arthanari H, Wagner G (2012) Applications of non-uniform sampling and processing. *Top Curr Chem* 316:125–148. doi:[10.1007/128_2011_187](https://doi.org/10.1007/128_2011_187)
 69. Rovnyak D, Sarcone M, Jiang Z (2011) Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. *Magn Reson Chem* 49(8):483–491. doi:[10.1002/mrc.2775](https://doi.org/10.1002/mrc.2775)
 70. Paramasivam S, Suiter CL, Hou GJ, Sun SJ, Palmer M, Hoch JC, Rovnyak D, Polenova T (2012) Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *J Phys Chem B* 116(25):7416–7427. doi:[10.1021/jp3032786](https://doi.org/10.1021/jp3032786)

71. Suiter CL, Paramasivam S, Hou G, Sun S, Rice D, Hoch JC, Rovnyak D, Polenova T (2014) Sensitivity gains, linearity, and spectral reproducibility in nonuniformly sampled multidimensional MAS NMR spectra of high dynamic range. *J Biomol NMR* 59(2):57–73. doi:[10.1007/s10858-014-9824-4](https://doi.org/10.1007/s10858-014-9824-4)
72. Palmer MR, Suiter CL, Henry GE, Rovnyak J, Hoch JC, Polenova T, Rovnyak D (2015) Sensitivity of nonuniform sampling NMR. *J Phys Chem B* 119(22):6502–6515. doi:[10.1021/jp5126415](https://doi.org/10.1021/jp5126415)
73. Rovnyak D, Hoch J, Stern A, Wagner G (2004) Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. *J Biomol NMR* 30(1):1–10. doi:[10.1023/B:JNMR.0000042946.04002.19](https://doi.org/10.1023/B:JNMR.0000042946.04002.19)
74. Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luthardt H, Fejzo J, Chorev M, Aktas H, Colson K, Falchuk KH, Halperin JA, Wagner G (2007) Ultrahigh-resolution ^1H - ^{13}C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J Am Chem Soc* 129(16):5108–5116. doi:[10.1021/ja068541x](https://doi.org/10.1021/ja068541x)
75. Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Kozminski W (2013) High-dimensional NMR spectra for structural studies of biomolecules. *ChemPhysChem* 14(13):3015–3025. doi:[10.1002/cphc.201300277](https://doi.org/10.1002/cphc.201300277)
76. Roche J, Ying J, Maltsev AS, Bax A (2013) Impact of hydrostatic pressure on an intrinsically disordered protein: a high-pressure NMR study of alpha-synuclein. *ChemBioChem* 14(14):1754–1761. doi:[10.1002/cbic.201300244](https://doi.org/10.1002/cbic.201300244)
77. Marion D (2010) Combining methods for speeding up multi-dimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins. *J Magn Reson* 206(1):81–87. doi:[10.1016/j.jmr.2010.06.007](https://doi.org/10.1016/j.jmr.2010.06.007)
78. Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 458(7234):102–105. doi:[10.1038/nature07814](https://doi.org/10.1038/nature07814)
79. Hoch JC, Maciejewski MW, Mobli M, Schuyler AD, Stern AS (2014) Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR. *Acc Chem Res* 47(2):708–717. doi:[10.1021/ar400244v](https://doi.org/10.1021/ar400244v)
80. Rovnyak D, Filip C, Itin B, Stern AS, Wagner G, Griffin RG, Hoch JC (2003) Multiple-quantum magic-angle spinning spectroscopy using nonlinear sampling. *J Magn Reson* 161(1):43–55
81. Shi C, Fricke P, Lin L, Chevelkov V, Wegstroth M, Giller K, Becker S, Thanbichler M, Lange A (2015) Atomic-resolution structure of cytoskeletal bacteriophage by solid-state NMR. *Sci Adv* 1(11):e1501087. doi:[10.1126/sciadv.1501087](https://doi.org/10.1126/sciadv.1501087)
82. Huber M, Hiller S, Schanda P, Ernst M, Bockmann A, Verel R, Meier BH (2011) A proton-detected 4D solid-state NMR experiment for protein structure determination. *ChemPhysChem* 12(5):915–918. doi:[10.1002/cphc.201100062](https://doi.org/10.1002/cphc.201100062)
83. Linsler R, Bardiaux B, Andreas LB, Hyberts SG, Morris VK, Pintacuda G, Sunde M, Kwan AH, Wagner G (2014) Solid-state NMR structure determination from diagonal-compensated, sparsely nonuniform-sampled 4D proton-proton restraints. *J Am Chem Soc* 136(31):11002–11010. doi:[10.1021/ja504603g](https://doi.org/10.1021/ja504603g)
84. LeMaster DM, Kushlan DM (1996) Dynamical mapping of E-coli thioredoxin via ^{13}C NMR relaxation analysis. *J Am Chem Soc* 118(39):9255–9264. doi:[10.1021/Ja960877r](https://doi.org/10.1021/Ja960877r)
85. 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 ^{13}C -labelled proteins. *J Biomol NMR* 44(4):245–260. doi:[10.1007/s10858-009-9338-7](https://doi.org/10.1007/s10858-009-9338-7)
86. Hong M (1999) Determination of multiple phi-torsion angles in proteins by selective and extensive ^{13}C labeling and two-dimensional solid-state NMR. *J Magn Reson* 139(2):389–401. doi:[10.1006/jmre.1999.1805](https://doi.org/10.1006/jmre.1999.1805)
87. Lian LY, Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog Nucl Magn Reson Spectrosc* 39(3):171–190. doi:[10.1016/S0079-6565\(01\)00034-6](https://doi.org/10.1016/S0079-6565(01)00034-6)
88. Asami S, Schmieder P, Reif B (2010) High resolution ^1H -detected solid-state NMR spectroscopy of protein aliphatic resonances: access to tertiary structure information. *J Am Chem Soc* 132(43):15133–15135. doi:[10.1021/ja106170h](https://doi.org/10.1021/ja106170h)
89. Gardner KH, Rosen MK, Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* 36(6):1389–1401. doi:[10.1021/Bi9624806](https://doi.org/10.1021/Bi9624806)
90. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* 13(4):369–374. doi:[10.1023/A:1008393201236](https://doi.org/10.1023/A:1008393201236)

91. Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed* 49(11):1958–1962. doi:[10.1002/anie.200905660](https://doi.org/10.1002/anie.200905660)
92. Zhang H, van Ingen H (2016) Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Curr Opin Struct Biol* 38:75–82. doi:[10.1016/j.sbi.2016.05.008](https://doi.org/10.1016/j.sbi.2016.05.008)
93. Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1(2):749–754. doi:[10.1038/nprot.2006.101](https://doi.org/10.1038/nprot.2006.101)
94. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440(7080):52–57. doi:[10.1038/nature04525](https://doi.org/10.1038/nature04525)
95. Salzman M, Pervushin K, Wider G, Senn H, Wuthrich K (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *J Am Chem Soc* 122(31):7543–7548. doi:[10.1021/Ja0003268](https://doi.org/10.1021/Ja0003268)
96. Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94(23):12366–12371. doi:[10.1073/Pnas.94.23.12366](https://doi.org/10.1073/Pnas.94.23.12366)
97. Skinner AL, Laurence JS (2008) High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. *J Pharm Sci* 97(11):4670–4695. doi:[10.1002/jps.21378](https://doi.org/10.1002/jps.21378)
98. Wuthrich K (1998) The second decade - into the third millennium. *Nat Struct Biol* 5:492–495. doi:[10.1038/728](https://doi.org/10.1038/728)
99. Takeuchi K, Arthanari H, Shimada I, Wagner G (2015) Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR. *J Biomol NMR* 63(4):323–331. doi:[10.1007/s10858-015-9991-y](https://doi.org/10.1007/s10858-015-9991-y)
100. Yoshimura Y, Oktaviani NA, Yonezawa K, Kamikubo H, Mulder FA (2017) Unambiguous determination of protein arginine ionization states in solution by NMR spectroscopy. *Angew Chem Int Ed* 56(1):239–242. doi:[10.1002/anie.201609605](https://doi.org/10.1002/anie.201609605)
101. Yuwen T, Skrynnikov NR (2014) CP-HISQC: a better version of HSQC experiment for intrinsically disordered proteins under physiological conditions. *J Biomol NMR* 58(3):175–192. doi:[10.1007/s10858-014-9815-5](https://doi.org/10.1007/s10858-014-9815-5)
102. Lopez J, Schneider R, Cantrelle FX, Huvent I, Lippens G (2016) Studying intrinsically disordered proteins under true in vivo conditions by combined cross-polarization and carbonyl-detection NMR spectroscopy. *Angew Chem Int Ed* 55(26):7418–7422. doi:[10.1002/anie.201601850](https://doi.org/10.1002/anie.201601850)
103. Zhou DH, Shea JJ, Nieuwkoop AJ, Franks WT, Wylie BJ, Mullen C, Sandoz D, Rienstra CM (2007) Solid-state protein-structure determination with proton-detected triple-resonance 3D magic-angle-spinning NMR spectroscopy. *Angew Chem Int Ed* 46(44):8380–8383. doi:[10.1002/anie.200702905](https://doi.org/10.1002/anie.200702905)
104. Zhou DH, Nieuwkoop AJ, Berthold DA, Comellas G, Sperling LJ, Tang M, Shah GJ, Brea EJ, Lemkau LR, Rienstra CM (2012) Solid-state NMR analysis of membrane proteins and protein aggregates by proton detected spectroscopy. *J Biomol NMR* 54(3):291–305. doi:[10.1007/s10858-012-9672-z](https://doi.org/10.1007/s10858-012-9672-z)
105. Verel R, Ernst M, Meier BH (2001) Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme. *J Magn Reson* 150(1):81–99. doi:[10.1006/jmre.2001.2310](https://doi.org/10.1006/jmre.2001.2310)
106. Barbet-Massin E, Pell AJ, Jaudzems K, Franks WT, Retel JS, Kotelovica S, Akopjana I, Tars K, Emsley L, Oschkinat H, Lesage A, Pintacuda G (2013) Out-and-back ^{13}C - ^{13}C scalar transfers in protein resonance assignment by proton-detected solid-state NMR under ultra-fast MAS. *J Biomol NMR* 56(4):379–386. doi:[10.1007/s10858-013-9757-3](https://doi.org/10.1007/s10858-013-9757-3)
107. Asami S, Reif B (2012) Assignment strategies for aliphatic protons in the solid-state in randomly protonated proteins. *J Biomol NMR* 52(1):31–39. doi:[10.1007/s10858-011-9591-4](https://doi.org/10.1007/s10858-011-9591-4)
108. Hennig M, Bermel W, Schwalbe H, Griesinger C (2000) Determination of psi torsion angle restraints from $(3\text{J})(\text{C}\alpha, \text{C}\alpha)$ and $3\text{J}(\text{C}\alpha, \text{HN})$ coupling constants in proteins. *J Am Chem Soc* 122(26):6268–6277. doi:[10.1021/JA9928834](https://doi.org/10.1021/JA9928834)
109. Andreas LB, Le Marchand T, Jaudzems K, Pintacuda G (2015) High-resolution proton-detected NMR of proteins at very fast MAS. *J Magn Reson* 253:36–49. doi:[10.1016/j.jmr.2015.01.003](https://doi.org/10.1016/j.jmr.2015.01.003)
110. Lange A, Luca S, Baldus M (2002) Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *J Am Chem Soc* 124(33):9704–9705. doi:[10.1021/ja026691b](https://doi.org/10.1021/ja026691b)

111. Reif B, van Rossum BJ, Castellani F, Rehbein K, Diehl A, Oschkinat H (2003) Characterization of ^1H – ^1H distances in a uniformly ^2H , ^{15}N -labeled SH3 domain by MAS solid-state NMR spectroscopy. *J Am Chem Soc* 125(6):1488–1489. doi:10.1021/ja0283697
112. Linser R, Bardiaux B, Higman V, Fink U, Reif B (2011) Structure calculation from unambiguous long-range amide and methyl ^1H – ^1H distance restraints for a microcrystalline protein with MAS solid-state NMR spectroscopy. *J Am Chem Soc* 133(15):5905–5912. doi:10.1021/ja110222h
113. Paulson EK, Morcombe CR, Gaponenko V, Dancheck B, Byrd RA, Zilm KW (2003) High-sensitivity observation of dipolar exchange and NOEs between exchangeable protons in proteins by 3D solid-state NMR spectroscopy. *J Am Chem Soc* 125(47):14222–14223. doi:10.1021/ja037559u
114. Knight MJ, Webber AL, Pell AJ, Guerry P, Barbet-Massin E, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Emsley L, Lesage A, Herrmann T, Pintacuda G (2011) Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state MAS NMR spectroscopy. *Angew Chem Int Ed* 50(49):11697–11701. doi:10.1002/anie.201106340
115. Andreas LB, Reese M, Eddy MT, Gelev V, Ni QZ, Miller EA, Emsley L, Pintacuda G, Chou JJ, Griffin RG (2015) Structure and mechanism of the influenza A M2(18–60) dimer of dimers. *J Am Chem Soc* 137(47):14877–14886. doi:10.1021/jacs.5604802
116. Sharma Y, Kwon OY, Brooks B, Tjandra N (2002) An ab initio study of amide proton shift tensor dependence on local protein structure. *J Am Chem Soc* 124(2):327–335. doi:10.1021/ja016859d
117. Wu G, Freure CJ, Verdurand E (1998) Proton chemical shift tensors and hydrogen bond geometry: a ^1H – ^2H dipolar NMR study of the water molecule in crystalline hydrates. *J Am Chem Soc* 120(50):13187–13193. doi:10.1021/ja983126t
118. Brouwer DH, Ripmeester JA (2007) Symmetry-based recoupling of proton chemical shift anisotropies in ultrahigh-field solid-state NMR. *J Magn Reson* 185(1):173–178. doi:10.1016/j.jmr.2006.12.003
119. Duma L, Abergel D, Tekely P, Bodenhausen G (2008) Proton chemical shift anisotropy measurements of hydrogen-bonded functional groups by fast magic-angle spinning solid-state NMR spectroscopy. *Chem Commun* 20:2361–2363. doi:10.1039/b801154k
120. Hou G, Paramasivam S, Yan S, Polenova T, Vega AJ (2013) Multidimensional magic angle spinning NMR spectroscopy for site-resolved measurement of proton chemical shift anisotropy in biological solids. *J Am Chem Soc* 135(4):1358–1368. doi:10.1021/ja3084972
121. Hou G, Gupta R, Polenova T, Vega AJ (2014) A magic-angle spinning NMR method for the site-specific measurement of proton chemical-shift anisotropy in biological and organic solids. *Isr J Chem* 54(1–2):171–183. doi:10.1002/ijch.201300099
122. Loth K, Pelupessy P, Bodenhausen G (2005) Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. *J Am Chem Soc* 127(16):6062–6068. doi:10.1021/ja042863o
123. Yao L, Grishaev A, Cornilescu G, Bax A (2010) The impact of hydrogen bonding on amide ^1H chemical shift anisotropy studied by cross-correlated relaxation and liquid crystal NMR spectroscopy. *J Am Chem Soc* 132(31):10866–10875. doi:10.1021/ja103629e
124. Cornilescu G, Bax A (2000) Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. *J Am Chem Soc* 122(41):10143–10154. doi:10.1021/ja0016194
125. Goldman M (1984) Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. *J Magn Reson* 60(3):437–452. doi:10.1016/0022-2364(84)90055-6
126. Lewandowski JR (2013) Advances in solid-state relaxation methodology for probing site-specific protein dynamics. *Acc Chem Res* 46(9):2018–2027. doi:10.1021/ar300334g
127. Tollinger M, Sivertsen AC, Meier BH, Ernst M, Schanda P (2012) Site-resolved measurement of microsecond-to-millisecond conformational-exchange processes in proteins by solid-state NMR spectroscopy. *J Am Chem Soc* 134(36):14800–14807. doi:10.1021/ja303591y
128. Knight MJ, Pell AJ, Bertini I, Felli IC, Gonnelli L, Pierattelli R, Herrmann T, Emsley L, Pintacuda G (2012) Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proc Natl Acad Sci USA* 109(28):11095–11100. doi:10.1073/pnas.1204515109

129. Ma PX, Haller JD, Zajacka J, Macek P, Sivertsen AC, Willbold D, Boisbouvier J, Schanda P (2014) Probing transient conformational states of proteins by solid-state $R_{1\rho}$ relaxation-dispersion NMR spectroscopy. *Angew Chem Int Ed* 53(17):4312–4317. doi:10.1002/anie.201311275
130. Lewandowski JR, Sass HJ, Grzesiek S, Blackledge M, Emsley L (2011) Site-specific measurement of slow motions in proteins. *J Am Chem Soc* 133(42):16762–16765. doi:10.1021/ja206815h
131. Lewandowski JR, Sein J, Sass HJ, Grzesiek S, Blackledge M, Emsley L (2010) Measurement of site-specific ^{13}C spin-lattice relaxation in a crystalline protein. *J Am Chem Soc* 132(24):8252–8254. doi:10.1021/ja102744b
132. Schanda P, Meier BH, Ernst M (2010) Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. *J Am Chem Soc* 132(45):15957–15967. doi:10.1021/ja100726a
133. Dannatt HRW, Felletti M, Jehle S, Wang Y, Emsley L, Dixon NE, Lesage A, Pintacuda G (2016) Weak and transient protein interactions determined by solid-state NMR. *Angew Chem Int Ed* 55(23):6637–6640. doi:10.1002/anie.201511609
134. Ishima R (2014) A probe to monitor performance of N-15 longitudinal relaxation experiments for proteins in solution. *J Biomol NMR* 58(2):113–122. doi:10.1007/s10858-013-9809-8
135. Su XC, Otting G (2010) Paramagnetic labeling of proteins and oligonucleotides for NMR. *J Biomol NMR* 46(1):101–112. doi:10.1007/s10858-009-9331-1
136. Otting G (2010) Protein NMR using paramagnetic ions. *Annu Rev Biophys* 39:387–405. doi:10.1146/annurev.biophys.093008.131321
137. Balayssac S, Bertini I, Lelli M, Luchinat C, Maletta M (2007) Paramagnetic ions provide structural restraints in solid-state NMR of proteins. *J Am Chem Soc* 129(8):2218–2219. doi:10.1021/ja068105a
138. Jahnke W (2002) Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chembiochem* 3(2–3):167–173. doi:10.1002/1439-7633(20020301)3:2/3<167::Aid-Cbic167>3.0.Co;2-S
139. Banci L, Bertini I, Luchinat C, Mori M (2010) NMR in structural proteomics and beyond. *Prog Nucl Magn Reson Spectrosc* 56(3):247–266. doi:10.1016/j.pnmrs.2009.12.003
140. Bertini I, Bhaumik A, De Paeppe G, Griffin RG, Lelli M, Lewandowski JR, Luchinat C (2010) High-resolution solid-state NMR structure of a 17.6 kDa protein. *J Am Chem Soc* 132(3):1032–1040. doi:10.1021/ja906426p
141. Sengupta I, Nadaud PS, Helmus JJ, Schwitters CD, Jaroniec CP (2012) Protein fold determined by paramagnetic magic-angle spinning solid-state NMR spectroscopy. *Nat Chem* 4(5):410–417. doi:10.1038/NCHEM.1299
142. Bertini I, Emsley L, Lelli M, Luchinat C, Mao J, Pintacuda G (2010) Ultrafast MAS solid-state NMR permits extensive ^{13}C and ^1H detection in paramagnetic metalloproteins. *J Am Chem Soc* 132(16):5558–5559. doi:10.1021/ja100398q
143. Knight MJ, Felli IC, Pierattelli R, Emsley L, Pintacuda G (2013) Magic angle spinning NMR of paramagnetic proteins. *Acc Chem Res* 46(9):2108–2116. doi:10.1021/ar300349y
144. Knight MJ, Felli IC, Pierattelli R, Bertini I, Emsley L, Herrmann T, Pintacuda G (2012) Rapid measurement of pseudocontact shifts in metalloproteins by proton-detected solid-state NMR spectroscopy. *J Am Chem Soc* 134(36):14730–14733. doi:10.1021/ja306813j
145. Rovo P, Grohe K, Giller K, Becker S, Linser R (2015) Proton transverse relaxation as a sensitive probe for structure determination in solid proteins. *ChemPhysChem* 16(18):3791–3796. doi:10.1002/cphc.201500799
146. Ashbrook SE (2009) Recent advances in solid-state NMR spectroscopy of quadrupolar nuclei. *Phys Chem Chem Phys* 11(32):6892–6905. doi:10.1039/b907183k
147. Frydman L, Harwood JS (1995) Isotropic spectra of half-integer quadrupolar spins from bidimensional magic-angle spinning NMR. *J Am Chem Soc* 117(19):5367–5368. doi:10.1021/ja00124a023
148. Wang SH, De Paul SM, Bull LM (1997) High-resolution heteronuclear correlation between quadrupolar and spin-1/2 nuclei using multiple-quantum magic-angle spinning. *J Magn Reson* 125(2):364–368. doi:10.1006/jmre.1997.1137
149. Gan Z (2000) Isotropic NMR spectra of half-integer quadrupolar nuclei using satellite transitions and magic-angle spinning. *J Am Chem Soc* 122(13):3242–3243. doi:10.1021/ja9939791
150. Thrippleton MJ, Ball TJ, Steuernagel S, Ashbrook SE, Wimperis S (2006) STARTMAS: a MAS-based method for acquiring isotropic NMR spectra of spin $I = 3/2$ nuclei in real

- time. *Chem Phys Lett* 431(4–6):390–396.
doi:[10.1016/j.cplett.2006.09.075](https://doi.org/10.1016/j.cplett.2006.09.075)
151. Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29(8):688–691.
doi:[10.1063/1.1716296](https://doi.org/10.1063/1.1716296)
152. Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* 94(3):630–638. doi:[10.1103/PhysRev.94.630](https://doi.org/10.1103/PhysRev.94.630)
153. Kupce E, Freeman R (1995) Adiabatic pulses for wide-band inversion and broad-band decoupling. *J Magn Reson A* 115(2):273–276. doi:[10.1006/Jmra.1995.1179](https://doi.org/10.1006/Jmra.1995.1179)
154. O'Dell LA, Rossini AJ, Schurko RW (2009) Acquisition of ultra-wideline NMR spectra from quadrupolar nuclei by frequency stepped WURST-QCPMG. *Chem Phys Lett* 468(4–6):330–335.
doi:[10.1016/j.cplett.2008.12.044](https://doi.org/10.1016/j.cplett.2008.12.044)
155. O'Dell LA, Schurko RW (2008) QCPMG using adiabatic pulses for faster acquisition of ultra-wideline NMR spectra. *Chem Phys Lett* 464(1–3):97–102.
doi:[10.1016/j.cplett.2008.08.095](https://doi.org/10.1016/j.cplett.2008.08.095)
156. Palmer AG, Williams J, McDermott A (1996) Nuclear magnetic resonance studies of biopolymer dynamics. *J Phys Chem A* 100(31):13293–13310. doi:[10.1021/jp9606117](https://doi.org/10.1021/jp9606117)
157. Wu G (2016) An approximate analytical expression for the nuclear quadrupole transverse relaxation rate of half-integer spins in liquids. *J Magn Reson* 269:176–178.
doi:[10.1016/j.jmr.2016.06.012](https://doi.org/10.1016/j.jmr.2016.06.012)
158. Zhu J, Wu G (2011) Quadrupole central transition ^{17}O NMR spectroscopy of biological macromolecules in aqueous solution. *J Am Chem Soc* 133(4):920–932.
doi:[10.1021/ja1079207](https://doi.org/10.1021/ja1079207)
159. Butler A, Eckert H (1989) Vanadium-51 NMR as a probe of vanadium(V) coordination to human apotransferrin. *J Am Chem Soc* 111(8):2802–2809. doi:[10.1021/ja00190a010](https://doi.org/10.1021/ja00190a010)
160. Germann MW, Aramini JM, Vogel HJ (1994) Quadrupolar metal ion NMR study of ovotransferrin at 17.6 T. *J Am Chem Soc* 116(15):6971–6972. doi:[10.1021/ja00094a076](https://doi.org/10.1021/ja00094a076)
161. Sasakawa H, Sakata E, Yamaguchi Y, Masuda M, Mori T, Kurimoto E, Iguchi T, Hisanaga SI, Iwatsubo T, Hasegawa M, Kato K (2007) Ultra-high field NMR studies of antibody binding and site-specific phosphorylation of alpha-synuclein. *Biochem Biophys Res Commun* 363(3):795–799.
doi:[10.1016/j.bbrc.2007.09.048](https://doi.org/10.1016/j.bbrc.2007.09.048)
162. Yamaguchi Y, Hanashima S, Yagi H, Takahashi Y, Sasakawa H, Kurimoto E, Iguchi T, Kon S, Uede T, Kato K (2010) NMR characterization of intramolecular interaction of osteopontin, an intrinsically disordered protein with cryptic integrin-binding motifs. *Biochem Biophys Res Commun* 393(3):487–491. doi:[10.1016/j.bbrc.2010.02.030](https://doi.org/10.1016/j.bbrc.2010.02.030)
163. Blundell CD, Reed MAC, Overduin M, Almond A (2006) NMR spectra of oligosaccharides at ultra-high field (900 MHz) have better resolution than expected due to favourable molecular tumbling. *Carbohydr Res* 341(12):1985–1991.
doi:[10.1016/j.carres.2006.05.017](https://doi.org/10.1016/j.carres.2006.05.017)
164. Park TJ, Kim JS, Ahn HC, Kim Y (2011) Solution and solid-state NMR structural studies of antimicrobial peptides LPcin-I and LPcin-II. *Biophys J* 101(5):1193–1201.
doi:[10.1016/j.bpj.2011.06.067](https://doi.org/10.1016/j.bpj.2011.06.067)
165. Quinn CM, Polenova T (2017) Structural biology of supramolecular assemblies by magic-angle spinning NMR spectroscopy. *Q Rev Biophys* 50.
doi:[10.1017/S0033583516000159](https://doi.org/10.1017/S0033583516000159)
166. Morag O, Sgourakis NG, Baker D, Goldbourn A (2015) The NMR-Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc Natl Acad Sci USA* 112(4):971–976.
doi:[10.1073/pnas.1415393112](https://doi.org/10.1073/pnas.1415393112)
167. Kaplan M, Pinto C, Houben K, Baldus M (2016) Nuclear magnetic resonance (NMR) applied to membrane-protein complexes. *Q Rev Biophys* 49.
doi:[10.1017/S003358351600010X](https://doi.org/10.1017/S003358351600010X)
168. Karamanos TK, Kalverda AP, Thompson GS, Radford SE (2015) Mechanisms of amyloid formation revealed by solution NMR. *Prog Nucl Magn Reson Spectrosc* 88–89:86–104.
doi:[10.1016/j.pnmrs.2015.05.002](https://doi.org/10.1016/j.pnmrs.2015.05.002)
169. Tycko R (2014) Physical and structural basis for polymorphism in amyloid fibrils. *Protein Sci* 23(11):1528–1539.
doi:[10.1002/pro.2544](https://doi.org/10.1002/pro.2544)
170. Meyer NH, Zangger K (2014) Enhancing the resolution of multi-dimensional heteronuclear NMR spectra of intrinsically disordered proteins by homonuclear broadband decoupling. *Chem Commun* 50(12):1488–1490.
doi:[10.1039/c3cc48135b](https://doi.org/10.1039/c3cc48135b)
171. Banci L, Barbieri L, Bertini I, Luchinat E, Secci E, Zhao YG, Aricescu AR (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol* 9(5):297–299.
doi:[10.1038/NCHEMBIO.1202](https://doi.org/10.1038/NCHEMBIO.1202)
172. Shishovs M, Rumnieks J, Diebold C, Jaudzems K, Andreas LB, Stanek J, Kazaks A,

- Kotelovica S, Akopjana I, Pintacuda G, Koning RI, Tars K (2016) Structure of AP205 coat protein reveals circular permutation in ssRNA bacteriophages. *J Mol Biol* 428 (21):4267–4279. doi:[10.1016/j.jmb.2016.08.025](https://doi.org/10.1016/j.jmb.2016.08.025)
173. Holsinger LJ, Lamb RA (1991) Influenza virus-M2 integral membrane-protein is a homotetramer stabilized by formation of disulfide bonds. *Virology* 183(1):32–43. doi:[10.1016/0042-6822\(91\)90115-R](https://doi.org/10.1016/0042-6822(91)90115-R)
174. Pinto LH, Holsinger LJ, Lamb RA (1992) Influenza-virus M2 protein has ion channel activity. *Cell* 69(3):517–528. doi:[10.1016/0092-8674\(92\)90452-I](https://doi.org/10.1016/0092-8674(92)90452-I)
175. Barbet-Massin E, Felletti M, Schneider R, Jehle S, Communie G, Martinez N, Jensen MR, Ruigrok RW, Emsley L, Lesage A, Blackledge M, Pintacuda G (2014) Insights into the structure and dynamics of measles virus nucleocapsids by ^1H -detected solid-state NMR. *Biophys J* 107(4):941–946. doi:[10.1016/j.bpj.2014.05.048](https://doi.org/10.1016/j.bpj.2014.05.048)
176. Luchinat E, Banci L (2016) A unique tool for cellular structural biology: in-cell NMR. *J Biol Chem* 291(8):3776–3784. doi:[10.1074/jbc.R115.643247](https://doi.org/10.1074/jbc.R115.643247)
177. Ikeya T, Hanashima T, Hosoya S, Shimazaki M, Ikeda S, Mishima M, Guntert P, Ito Y (2016) Improved in-cell structure determination of proteins at near-physiological concentration. *Sci Rep* 6:38312. doi:[10.1038/srep38312](https://doi.org/10.1038/srep38312)
178. Pan BB, Yang F, Ye Y, Wu Q, Li C, Huber T, Su XC (2016) 3D structure determination of a protein in living cells using paramagnetic NMR spectroscopy. *Chem Commun* 52 (67):10237–10240. doi:[10.1039/c6cc05490k](https://doi.org/10.1039/c6cc05490k)
179. Theillet FX, Binolfi A, Bekei B, Martorana A, Rose HM, Stuiver M, Verzini S, Lorenz D, van Rossum M, Goldfarb D, Selenko P (2016) Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* 530(7588):45–50. doi:[10.1038/nature16531](https://doi.org/10.1038/nature16531)
180. Bhunia A, Bhattacharjya S, Chatterjee S (2012) Applications of saturation transfer difference NMR in biological systems. *Drug Discov Today* 17(9–10):505–513. doi:[10.1016/j.drudis.2011.12.016](https://doi.org/10.1016/j.drudis.2011.12.016)
181. Romaniuk JAH, Cegelski L (2015) Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR. *Philos Trans R Soc B* 370(1679). doi:[10.1098/Rstb.2015.0024](https://doi.org/10.1098/Rstb.2015.0024)
182. Arnold AA, Genard B, Zito F, Tremblay R, Warschawski DE, Marcotte I (2015) Identification of lipid and saccharide constituents of whole microalgal cells by ^{13}C solid-state NMR. *Biochim Biophys Acta* 1848(1 Pt B):369–377. doi:[10.1016/j.bbamem.2014.07.017](https://doi.org/10.1016/j.bbamem.2014.07.017)
183. Coen M, Holmes E, Lindon JC, Nicholson JK (2008) NMR-based metabolic profiling and metabolomic approaches to problems in molecular toxicology. *Chem Res Toxicol* 21 (1):9–27. doi:[10.1021/tx700335d](https://doi.org/10.1021/tx700335d)
184. Vogel EP, Curtis-Fisk J, Young KM, Weliky DP (2011) Solid-state nuclear magnetic resonance (NMR) spectroscopy of human immunodeficiency virus gp41 protein that includes the fusion peptide: NMR detection of recombinant Fgp41 in inclusion bodies in whole bacterial cells and structural characterization of purified and membrane-associated Fgp41. *Biochemistry* 50(46):10013–10026. doi:[10.1021/bi201292e](https://doi.org/10.1021/bi201292e)
185. Vogel EP, Weliky DP (2013) Quantitation of recombinant protein in whole cells and cell extracts via solid-state NMR spectroscopy. *Biochemistry* 52(25):4285–4287. doi:[10.1021/bi4007034](https://doi.org/10.1021/bi4007034)
186. Curtis-Fisk J, Spencer RM, Weliky DP (2008) Native conformation at specific residues in recombinant inclusion body protein in whole cells determined with solid-state NMR spectroscopy. *J Am Chem Soc* 130(38):12568–12569. doi:[10.1021/ja8039426](https://doi.org/10.1021/ja8039426)
187. Sharif S, Kim SJ, Labischinski H, Schaefer J (2009) Characterization of peptidoglycan in fem-deletion mutants of methicillin-resistant *Staphylococcus aureus* by solid-state NMR. *Biochemistry* 48(14):3100–3108. doi:[10.1021/bi801750u](https://doi.org/10.1021/bi801750u)
188. Cegelski L, Steuber D, Mehta AK, Kulp DW, Axelsen PH, Schaefer J (2006) Conformational and quantitative characterization of oritavancin-peptidoglycan complexes in whole cells of *Staphylococcus aureus* by *in Vivo* ^{13}C and ^{15}N labeling. *J Mol Biol* 357(4):1253–1262. doi:[10.1016/j.jmb.2006.01.040](https://doi.org/10.1016/j.jmb.2006.01.040)
189. Kim SJ, Matsuoka S, Patti GJ, Schaefer J (2008) Vancomycin derivative with damaged D-Ala-D-Ala binding cleft binds to cross-linked peptidoglycan in the cell wall of *Staphylococcus aureus*. *Biochemistry* 47 (12):3822–3831. doi:[10.1021/bi702232a](https://doi.org/10.1021/bi702232a)
190. Mroue KH, Nishiyama Y, Pandey MK, Gong B, McNerny E, Kohn DH, Morris MD, Ramamoorthy A (2015) Proton-detected solid-state NMR spectroscopy of bone with ultrafast magic angle spinning. *Sci Rep* 5. doi:[10.1038/Srep11991](https://doi.org/10.1038/Srep11991)
191. Laurencin D, Wong A, Dupree R, Smith ME (2008) Natural abundance ^{43}Ca solid-state

- NMR characterisation of hydroxyapatite: identification of the two calcium sites. *Magn Reson Chem* 46(4):347–350. doi:[10.1002/mrc.2117](https://doi.org/10.1002/mrc.2117)
192. Xu J, Zhu P, Gan Z, Sahar N, Tecklenburg M, Morris MD, Kohn DH, Ramamoorthy A (2010) Natural-abundance ^{43}Ca solid-state NMR spectroscopy of bone. *J Am Chem Soc* 132(33):11504–11509. doi:[10.1021/ja101961x](https://doi.org/10.1021/ja101961x)
193. Laurencin D, Wong A, Chrzanowski W, Knowles JC, Qiu D, Pickup DM, Newport RJ, Gan Z, Duer MJ, Smith ME (2010) Probing the calcium and sodium local environment in bones and teeth using multinuclear solid state NMR and X-ray absorption spectroscopy. *Phys Chem Chem Phys* 12(5):1081–1091. doi:[10.1039/b915708e](https://doi.org/10.1039/b915708e)
194. Ravera E, Cerofolini L, Martelli T, Louka A, Fragai M, Luchinat C (2016) ^1H -detected solid-state NMR of proteins entrapped in bioinspired silica: a new tool for biomaterials characterization. *Sci Rep* 6:27851. doi:[10.1038/srep27851](https://doi.org/10.1038/srep27851)
195. Ravera E, Martelli T, Geiger Y, Fragai M, Goobes G, Luchinat C (2016) Biosilica and bioinspired silica studied by solid-state NMR. *Coord Chem Rev* 327–328:110–122. doi:[10.1016/j.ccr.2016.06.003](https://doi.org/10.1016/j.ccr.2016.06.003)
196. Fragai M, Luchinat C, Martelli T, Ravera E, Sagi I, Solomonov I, Udi Y (2014) SSNMR of biosilica-entrapped enzymes permits an easy assessment of preservation of native conformation in atomic detail. *Chem Commun* 50(4):421–423. doi:[10.1039/C3CC46896H](https://doi.org/10.1039/C3CC46896H)
197. Bozarth A, Maier UG, Zauner S (2009) Diatoms in biotechnology: modern tools and applications. *Appl Microbiol Biotechnol* 82(2):195–201. doi:[10.1007/s00253-008-1804-8](https://doi.org/10.1007/s00253-008-1804-8)
198. Brunner E, Groger C, Lutz K, Richthammer P, Spinde K, Sumper M (2009) Analytical studies of silica biomineralization: towards an understanding of silica processing by diatoms. *Appl Microbiol Biotechnol* 84(4):607–616. doi:[10.1007/s00253-009-2140-3](https://doi.org/10.1007/s00253-009-2140-3)

Experimental Aspects of Polarization Optimized Experiments (POE) for Magic Angle Spinning Solid-State NMR of Microcrystalline and Membrane-Bound Proteins

T. Gopinath and Gianluigi Veglia

Abstract

Conventional NMR pulse sequences record one spectrum per experiment, while spending most of the time waiting for the spin system to return to the equilibrium. As a result, a full set of multidimensional NMR experiments for biological macromolecules may take up to several months to complete. Here, we present a practical guide for setting up a new class of MAS solid-state NMR experiments (POE or polarization optimized experiments) that enable the simultaneous acquisition of multiple spectra of proteins, accelerating data acquisition. POE exploit the long-lived ^{15}N polarization of isotopically labeled proteins and enable one to obtain up to eight spectra, by concatenating classical NMR pulse sequences. This new strategy propels data throughput of solid-state NMR spectroscopy of fibers, microcrystalline preparations, as well as membrane proteins.

Key words Solid-state NMR, Multiple acquisitions, POE, DUMAS, MEIOSIS, MAeSTOSO, SIM-CP

1 Introduction

Magic angle spinning solid-state NMR (MAS ssNMR) spectroscopy plays a central role in the characterization of structure, motions, and interactions of biological macromolecules [1–6]. Unlike X-ray crystallography, MAS ssNMR does not require well-diffracting crystals and is able to analyze protein fibrils, microcrystalline preparations, as well as membrane proteins in their natural lipid environments [7–11]. However, the low-throughput of ssNMR data is a significant concern. As a result, the number of protein structures or dynamic studies utilizing ssNMR is rather sparse.

A significant hurdle is the intrinsic low sensitivity of NMR spectroscopy, which makes data acquisition time-consuming. NMR researchers must invest a significant portion of time and budgetary resources to carry out several multidimensional NMR

experiments for resonance assignments and structure determination. Recent developments in both software and hardware have alleviated these problems to a large extent. Technical developments include higher static magnetic fields (B_0), low-E or E-free probes that increase the sensitivity of the RF coil, avoiding heating caused by high-power decoupling pulses [12–14], dynamic nuclear polarization (DNP) [15], paramagnetic relaxation enhancement (PRE) [16], as well as the most recent ^1H detection methods using fast and ultra-fast MAS probes [17–19].

We recently introduced a new class of experiments, called POE for polarization optimized experiments, which are based on long-lived ^{15}N polarization generated for biological solids that enable the acquisition of multiple discrete NMR spectra simultaneously [20–24]. POE are synergistic to the above methods and promise to speed up the acquisition of solid-state NMR data. Specifically, POE are based on the detection of multiple FIDs from orphan polarization that is discarded in classical ssNMR experiments. The development of POE was inspired by the work of Alex Pines, who noted that multiple cross-polarization experiments enable one to record several FIDs sequentially [25]. Along these lines, the “afterglow” effects of polarization have been previously introduced for both solution and solid-state NMR experiments for acquiring multiple spectra simultaneously [26–28]. Unlike solution NMR experiments, where the T_1 relaxation is quite rapid, solid-state NMR experiments can be designed in a way that ^{15}N polarization originating from a simultaneous cross-polarization scheme (SIM-CP) can be stored along the z-direction and exploited for multiple experiments using a single pulse sequence.

In this paper, we provide a step-by-step guide to the design and concatenation of several multidimensional MAS ssNMR experiments to optimize the time on the NMR spectrometers. We will describe our recent pulse schemes, DUMAS, MEIOSIS, and MAE-STOSO with a detailed description of experimental setup, data acquisition, and processing.

2 POE: A New Class of Experiments for Simultaneous Acquisition of ssNMR Spectra

Solid-state NMR experiments utilize about 5–10% of the total experimental time for pulse execution and acquisition of the signal (FID), while for 90–95% of the time the spectrometer is idle (2–3 s) waiting for the spin system to return to the equilibrium (i.e., through longitudinal relaxation, T_1) and to conform to probe duty cycle guidelines. Classical ssNMR methods are designed for a single acquisition period, where each 2D or 3D experiment is acquired separately from the evolution, coherence transfer, and detection of a specific polarization pathway. The basic ideas behind POE include (1) the creation of multiple spin polarization

pathways, and (2) the acquisition of multiple 2D experiments using two to four acquisition periods per each scan. Using these two concepts, we are able to concatenate several different experiments and obtain multiple spectra using a single pulse sequence. Specifically, we have developed four different strategies for POE (Fig. 1). Our most recent implementation, MAeSTOSO-8, makes it possible to acquire up to eight different experiments simultaneously. In principle, there are multiple combinations possible for each of these experiments. A few examples are reported in Fig. 2.

The central element in each of these methods is the simultaneous cross-polarization (SIM-CP) that creates two polarization pathways for ^{13}C and ^{15}N . For instance, in the DUMAS experiment, ^{13}C and ^{15}N polarization originating from SIM-CP is used to acquire the CXCX and NCA experiments in the first and second acquisition periods, respectively (Fig. 2a). A similar scheme is used in MAeSTOSO-4, where CXCX and NCACB are acquired in the first and second acquisitions, and two more experiments (NCACX and NCO) can be also acquired utilizing the residual ^{15}N polarization resulting from the NC transfer periods. MEIOSIS, on the other hand, was designed to concatenate C(N)C with CC and NC correlation experiments using two acquisitions per scan. In the MEIOSIS experiment (Fig. 2c), CXCX and NCACX are acquired in first acquisition, whereas CA(N)CO and NCO are recorded in the second acquisition period. To deconvolute the two polarization pathways, two data sets with alternating ϕ phases (x and $-x$) of ^{15}N spin-lock are recorded. Using similar strategies, the simultaneous acquisition of up to eight FIDs (MAeSTOSO-8) can be achieved with four acquisition periods (Fig. 2d).

3 Experimental Setup of POE and Data Acquisition

Modern MAS ssNMR experiments on proteins are generally carried out using either fast spinning speeds with 1.3 mm diameter rotors or moderate spinning speeds with 3.2 mm rotors. The typical sample volume for a 1.3 mm rotor is about 3 μL , reaching high spinning speeds (~40–60 kHz) that truncate the ^1H – ^1H dipolar coupling network and enabling ^1H detection for high sensitivity [17, 29, 30]. Due to the small volume, the 1.3 mm probes and rotors are mainly used for highly concentrated protein samples such as microcrystalline preparations with low hydration levels. In contrast, the MAS probes with 3.2 mm rotors can accommodate up to 30 μL sample with maximum spinning speed of approximately 10–15 kHz. Due to the larger sample volume, these rotors can accommodate fully hydrated samples such as membrane proteins, whose concentrations are quite dilute due to the presence of lipids necessary to maintain their native state. Under these moderate spinning speed conditions, ^1H resonances are very broad, and

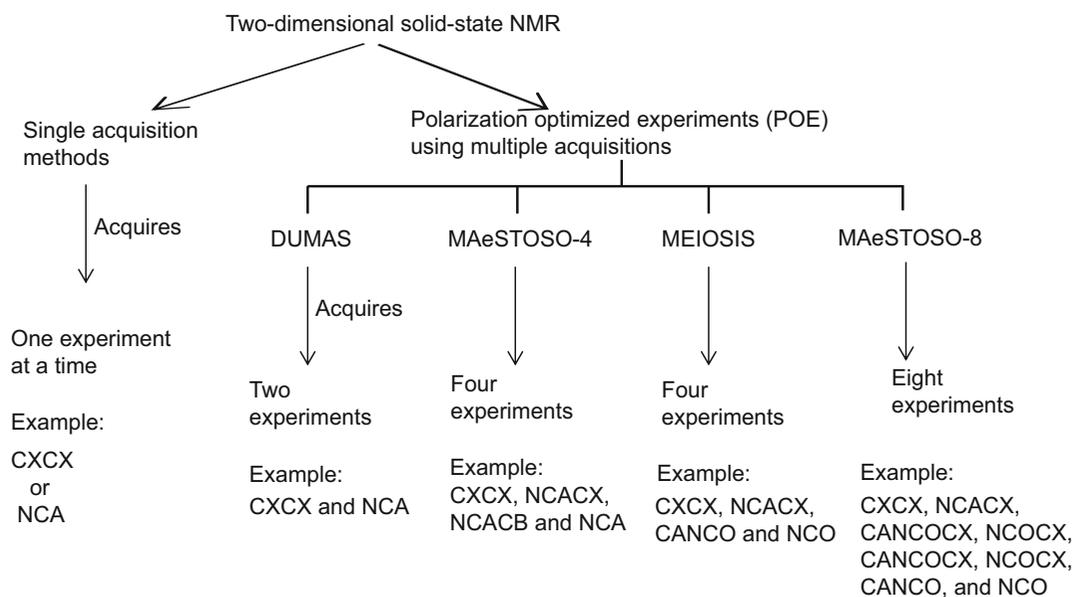


Fig. 1 Schematic of two-dimensional ssNMR methodologies. Conventional single-acquisition methods record one experiment at a time. POE class of experiments enable the acquisition of multiple 2D spectra using DUMAS, MAeSTOSO-4, MEIOSIS, and MAeSTOSO-8 strategies

spectroscopic studies are usually carried out using ^{13}C -detected multidimensional experiments. Indeed, POE are most effective for membrane protein samples under moderate spinning speed conditions and with 3.2 mm or 4 mm MAS probes.

To demonstrate the performance of the POE methods, we utilized a microcrystalline preparation with $\text{U-}^{13}\text{C},^{15}\text{N}$ labeled ubiquitin and two $\text{U-}^{13}\text{C},^{15}\text{N}$ labeled membrane proteins (phospholamban, PLN, and sarcolipin, SLN) as benchmarks. The microcrystalline $\text{U-}^{13}\text{C},^{15}\text{N}$ ubiquitin sample consisted of ~ 8 mg of protein, whereas PLN and SLN were reconstituted in 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipids to an approximate 1:100 protein-to-lipid ratio using our standard protocols [3, 24, 31]. Note that our membrane protein preparations contain at least 100 lipids per protein to maintain native-like conditions and avoid aggregations. Typically, the reconstituted samples contain 2–3 mg of protein and 15 mg of DMPC lipids. After organic solvent- or detergent-mediated reconstitution, the lipoproteins are centrifuged and the pellets with approximately 30% (w/v) hydration level are packed into 3.2 mm rotors for spectral analysis.

POE have been implemented on 600 MHz Agilent spectrometers using 3.2 mm scroll coil MAS probe. All of the relevant experimental parameters are summarized in Table 1. For hydrated samples such as membrane proteins, the sample temperature plays a major role for optimal sensitivity and resolution. For proteoliposomes, the sample temperature can affect the lipid phase and

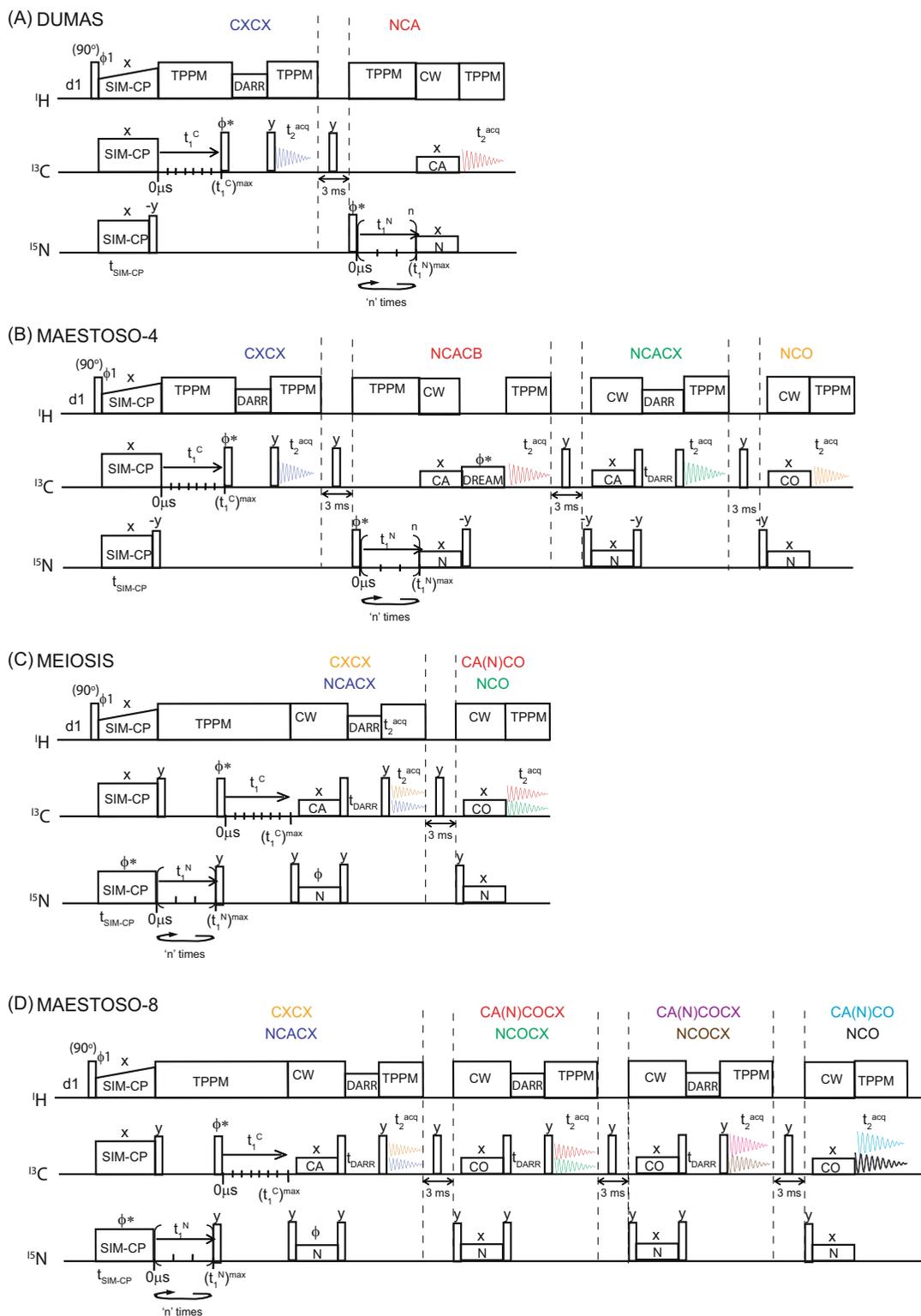


Fig. 2 Concatenation of 2D ssNMR experiments using DUMAS, MAESTOSO-4, MEIOSIS, and MAESTOSO-8

Table 1
Experimental parameters used for the POE experiments reported in Fig. 2

Parameter	Value	Parameter	Value	Parameter	Value
d1 (Recycle delay)	2 s	t_{2acq}	15–20 ms	Dwell time of t_1^c	30 μ s
MAS rate	12 kHz	Dwell time of t2	10 μ s	Number of t_1^c increments	256
A_{SIM-CP}^H	59 kHz	A_{TPPM}^H	100 kHz	$(t_1^c)^{max}$	7680 μ s
A_{SIM-CP}^C	35 kHz	1H offset	5 ppm	Dwell time of t_1^N	300 μ s
A_{SIM-CP}^N	35 kHz	^{15}N offset	120 ppm	Number of t_1^N increments	32
Δ_{SIM-CP}^H	2 kHz	^{15}C offset	100 ppm	$(t_1^N)^{max}$	9600 μ s
t_{SIM-CP}	300–1000 μ s			Number of t_1^N loops	8
90° pulses					
A_{90}^H	100 kHz	A_{90}^C	41.6 kHz	A_{90}^N	41.6 kHz
t_{90}^H	2.5 μ s	t_{90}^C	6 μ s	t_{90}^N	6 μ s
CC mixing			NC mixing		
CXCX		CACB		NCA/NCO	
^{13}C offset	100 ppm	^{13}C offset	52 ppm	^{13}C offset (NCA)	55 ppm
t_{DARR}	20–500 ms	t_{DREAM}	2.53 ms	A^C (NCA)	18.5 kHz
A_{DARR}^H	12 kHz	A_{DREAM}^H	6.69 kHz	^{13}C offset (NCO)	170 ppm
		Δ_{DREAM}^C	3.12 kHz	A^C (NCO)	39.8 kHz
		β_{DREAM}^C	1.56 kHz	^{15}N offset (NCA/NCO)	120 ppm
				A^N (NCA/NCO)	31 kHz
				Δ^C (NCA/NCO)	2.8 kHz
				β^C (NCA/NCO)	0.9 kHz
				$t_{NCA/NCO}$	3 ms

X_p^N : RF amplitude of “X” applied on nucleus “N” during the pulse sequence block “P”; t_p : Time period “t” of the pulse sequence block “p”

protein’s rotational diffusion, influencing both sensitivity and resolution. At 25 °C, DMPC lipids are in the liquid crystalline phase, which promotes the protein’s rotational diffusion, which scales dipolar couplings (DC) and lowers the performance of the cross-polarization. On the other hand, protein spectra recorded at 0–4 °C with the lipids in gel phase are more sensitive. In addition to temperature, sensitivity changes are also dictated by the protein size, lipid composition, and protein-lipid interactions. Although it is always preferable to carry out the NMR experiments under conditions close to the liquid crystalline phase, the lower sensitivity may represent a serious hurdle, extending the experimental acquisition time significantly.

For all POE reported in Figs. 1 and 2, and Table 1, the recycle delay was set to 2 s, and MAS rate (ω_r) to 12 kHz. As per probe requirements, the maximum RF amplitude on the ^1H channel was set to 100 kHz, which corresponds to a $2.5 \mu\text{s}$ 90° pulse; whereas for ^{13}C and ^{15}N a $6 \mu\text{s}$ 90° pulses were used, corresponding to 41.6 kHz RF amplitude. During the t_1 evolution and t_2 acquisition periods, a TPPM decoupling sequence was used on the ^1H channel with a 100 kHz RF amplitude [32]. During SIM-CP, the RF amplitudes for ^{13}C and ^{15}N were set to 35 kHz, and ^1H RF amplitude was ramped, with the center of ramp set to 59 kHz satisfying the Hartman-Hahn matching condition at 12 kHz MAS rate [33]. Typical contact times for SIM-CP range from 300 to 1000 μs . Usually, the backbone $^{15}\text{N}^{\text{H}}$ and ^{13}CH groups are cross-polarized at contact times ranging from 300 to 500 μs , whereas ^{13}CO , $^{13}\text{CH}_2$, and NH_2 groups require longer contact times up to 1000–1500 μs . In our experiments, we generally set SIM-CP contact time at 500 μs to obtain maximum intensity for $^{13}\text{C}\alpha$, and $^{15}\text{N}^{\text{H}}$ nuclei of the protein backbone. Note that optimal contact times can vary significantly among different samples depending on the hydration levels.

Specific CP that transfers polarization between ^{13}C and ^{15}N nuclei is another important building block for classical triple-resonance MAS ssNMR experiments as well as for POE. Unlike for ^1H - ^{13}C and ^1H - ^{15}N polarization transfer via CP or SIM-CP, the NC transfer is mediated by relatively weak DC between ^{13}C and ^{15}N , and thus requires careful optimization of specific-CP RF parameters [34, 35]. In other words, the rotating frame matching condition for NC transfer is relatively narrow compared to HC or HN CP transfer. In our case NC transfer was obtained by using tangent-shaped ramp pulse on ^{13}C and constant amplitude pulse was applied on ^{15}N , while applying a 100 kHz RF for ^1H decoupling. Theoretical RF amplitudes for the NCA transfer are $1.5\omega_r$ and $2.5\omega_r$ for ^{13}C , ^{15}N , respectively, where ω_r is the spinning rate. For the NCO transfer, the theoretical values are $3.5\omega_r$ and $2.5\omega_r$ for ^{13}C and ^{15}N , respectively. Note that the optimal experimental RF values for NC transfer slightly deviate from the theoretical ones (Table 1). Typically, one needs to optimize these values, starting from the theoretical values and arraying the acquisition of multiple 1D experiments to select the best parameters for the highest sensitivity. We typically carry out this optimization acquiring the 1st increment of the 2D DUMAS experiments, where the second acquisition used to optimize the RF parameters for the NC transfer including the tangential ^{13}C ramp parameters Δ and β . Similarly, we optimize the DREAM mixing period for the CACB transfer using the second acquisition of MAeSTOSO-4 [36]. For the CXCX experiments, the DARR mixing times are based on the desired intra- or inter-residue correlations [37]. Typical DARR mixing times are in the range of 20–500 ms. Note that longer mixing

times cause a reduction in the overall sensitivity of the experiment and require more scans.

The t_2 acquisition times ($t_{2\text{acq}}$) for DUMAS and MEIOSIS are usually set to 20 ms, whereas for MAeSTOSO experiments $t_{2\text{acq}}$ are set to 15 ms to avoid RF heating. While all $t_{2\text{acq}}$ on ^{13}C use identical parameters, the indirect t_1 evolution times for ^{13}C and ^{15}N ($t_{1\text{C}}$ and $t_{1\text{N}}$) must be synchronized. Usually, the ^{15}N evolution requires fewer t_1 increments than ^{13}C due to the longer dwell times. For example, a t_1 dwell time of 30 μs and 256 increments covers a ^{13}C spectral width of 33.33 kHz with a total t_1 evolution time of 7.68 ms. In contrast, the ^{15}N evolution requires a 300 μs dwell time with only 32 increments for a total t_1 evolution period of 9.6 ms. In order to synchronize the 256 increments of ^{13}C evolution with the ^{15}N evolution periods, each of the 32 increments for the ^{15}N evolution is repeated eight times (number of loops in Table 1) and added in the spectral processing. As a result, the effective number of scans for the processed NC spectra will be eight times higher compared to the CC correlation spectra. All the data sets are externally referenced to the CH_2 resonance of adamantane at 40.48 ppm and processed using NMRPipe as described below [38].

4 Processing POE Experiments

POE require special data processing to deconvolute the individual FID data sets and then Fourier transform them into multiple 2D spectra. Figure 3 shows the layout of the processing procedure using NMRpipe scripts [38]. Briefly, the raw data are first converted into multiple 2D FID data sets by executing *fid.com* script that contains the experimental acquisition parameters. The *fid.com* scripts for DUMAS and MAeSTOSO-4 are reported in Table 2, while the *fid.com* for MEIOSIS and MAeSTOSO-8 are shown in Table 3. For all four experiments (DUMAS, MAeSTOSO, MEIOSIS, and MAeSTOSO-8) the numbers of t_1 complex points for ^{13}C and ^{15}N evolution are set to 256 and 32, respectively, with eight loops for ^{15}N t_1 evolution. Note that in the *fid.com* script for conventional 2D experiments (one acquisition per scan) the “yN” and “yT” variables correspond to real and complex points, respectively. In the POE data sets, however, the definitions of yN and yT are given in Table 4. Depending on the experimental scheme, the outputs from the *fid.com* input file generate one to four ^{13}C - and ^{15}N -edited time domains in the 2D data sets. For instance, in the DUMAS experiment, the *fid.com* input file deconvolutes the FID into two data sets, CC.fid and NC.fid, which correspond to ^{13}C - and ^{15}N -edited experiments acquired in the first and second acquisition periods. On the other hand, in the MAeSTOSO-8-experiment, the ^{13}C -edited 2D data sets (CC.fid, CNC1.fid,

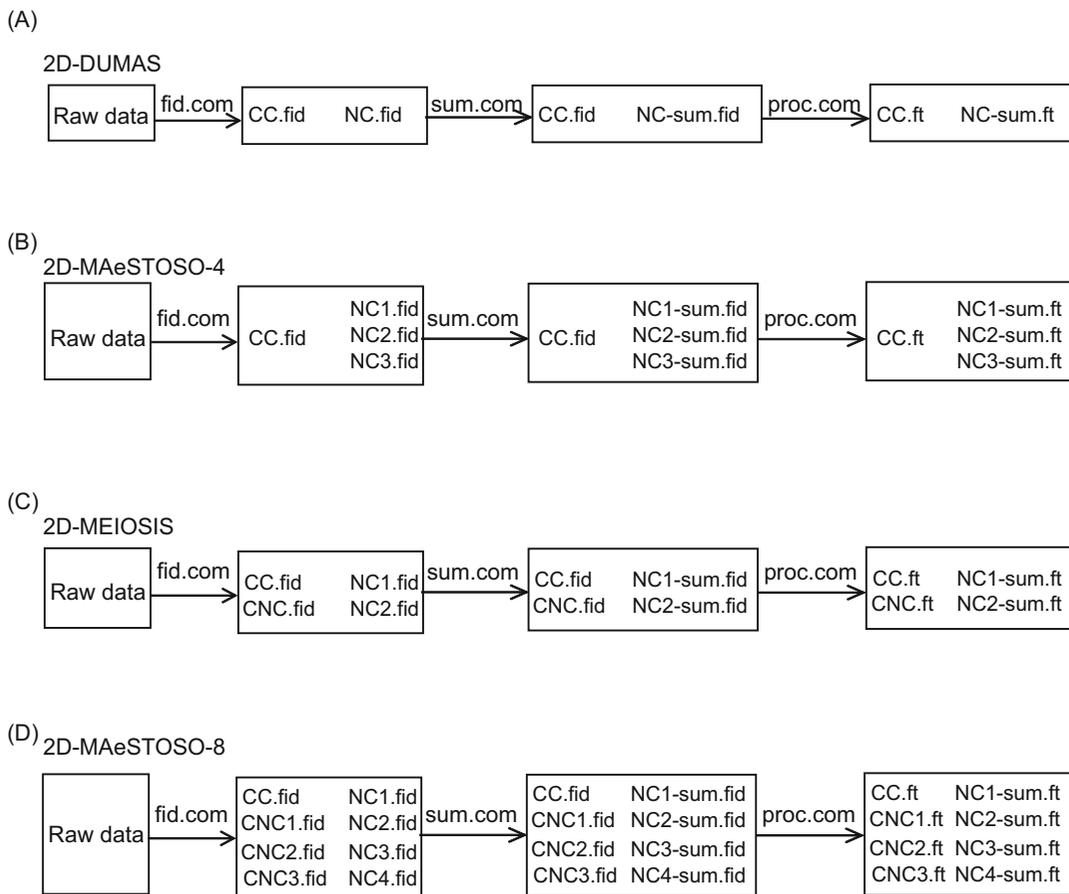


Fig. 3 Data processing procedure of 2D ssNMR data acquired from DUMAS, MAeSTOSO-4, MEIOSIS, and MAeSTOSO-8 experiments. NMRpipe scripts for [fid.com](#), [sum.com](#), and [proc.com](#) are given in Tables 2–6

CNC2.fid, and CNC3.fid) are deconvoluted from the FID (test.fid) using the COADD command line in the [fid.com](#) script. Similarly, the ^{15}N -edited 2D data sets (NC.fid, NC1.fid, NC2.fid, and NC3.fid) are deconvoluted from the test1.fid file. Following this step, the ^{13}C -edited data sets remain the same, whereas each of the ^{15}N -edited 2D data sets is further separated into eight t_1 data sets ($n = 8$ loops) and then corresponding t_1 increments in each of the eight data sets is added using the [sum.com](#) script as shown in Table 5. In the final step, multiple 2D spectra are generated by 2D Fourier transformation executing the [proc.com](#) script (Table 6) on each of the ^{13}C - and ^{15}N - edited 2D data sets. Figure 4 shows the 2D DUMAS spectra of sarcolipin membrane protein, and

Table 2
NMRpipe script for DUMAS and MAeSTOSO-4 for converting the raw data into nmrpipe format

<pre>#2D DUMAS 'fid.com' script #Number of 13C t1 complex points=256 #Number of 15N t1 complex points=32 #Number of 15N t1 loops=8 varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 1024 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 10000.000 -ySW 3333.33 \ -xOBS 150.787 -yOBS 150.787 \ -xCAR 100.0 -yCAR 100.0 \ -xLAB C13 -yLAB C13 \ -ndim 2 -aq2D Real \ -out ./test.fid -verb -ov nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 -time \ -verb -ov -out CC.fid varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 1024 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 10000.000 -ySW 3333.33\ -xOBS 150.787 -yOBS 60.766 \ -xCAR 96.87 -yCAR 118.74 \ -xLAB C13 -yLAB N15 \ -ndim 2 -aq2D Real \ -out ./test1.fid -verb -ov nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 -time \ -verb -ov -out NC.fid</pre>	<pre>#2D MAeSTOSO-4 'fid.com' script #Number of 13C t1 complex points=256 #Number of 15N t1 complex points=32 #Number of 15N t1 loops=8 varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 2048 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 33333.33 \ -xOBS 150.787 -yOBS 150.787 \ -xCAR 100.0 -yCAR 100.0 \ -xLAB C13 -yLAB C13 \ -ndim 2 -aq2D Real \ -out ./test.fid -verb -ov nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 0 0-time \ -verb -ov -out CC.fid varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 1024 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 3333.33\ -xOBS 150.787 -yOBS 60.766 \ -xCAR 96.87 -yCAR 118.74 \ -xLAB C13 -yLAB N15 \ -ndim 2 -aq2D Real \ -out ./test1.fid -verb -ov nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 0 0 -time \ -verb -ov -out NC1.fid nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 1 0 -time \ -verb -ov -out NC2.fid nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 1 0 -time \ -verb -ov -out NC3.fid nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 0 1 -time \ -verb -ov -out NC4.fid</pre>
---	--

MAeSTOSO-4 spectra of ubiquitin microcrystalline preparation using the acquisition parameters shown in Table 1. The spectra were processed using the *fid.com* (Table 2), *sum.com* (Table 5), *proc.com* (Table 6) scripts in NMRpipe.

Table 3

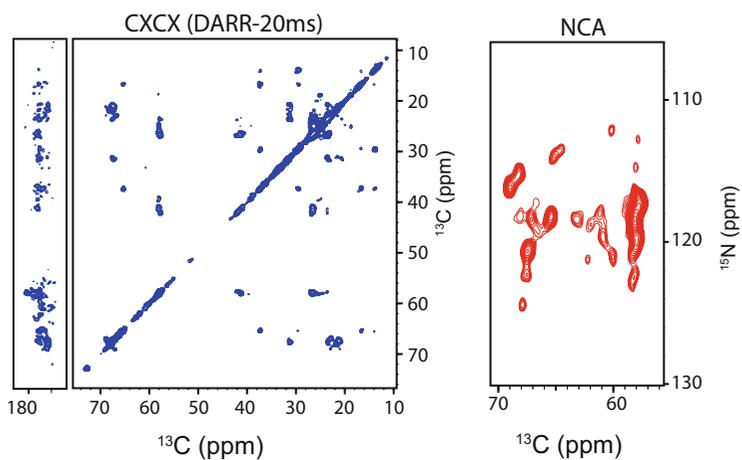
NMRpipe script for MEIOSIS and MeESTOSO-8 for converting the raw data into nmrpipe format

<pre>#2D MEIOSIS 'fid.com' script #Number of 13C t1 complex points=256 #Number of 15N t1 complex points=32 #Number of 15N t1 loops=8 varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 2048 \ -xT 2000 -yT 512 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 33333.33 \ -xOBS 150.787 -yOBS 150.787 \ -xCAR 100.0 -yCAR 100.0 \ -xLAB C13 -yLAB C13 \ -ndim 2 -aq2D Real \ -out ./test.fid -verb -ov nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 1 0 -time \ -verb -ov -out CC.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 0 1 -time \ -verb -ov -out CNC.fid varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 2048 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 3333.33\ -xOBS 150.787 -yOBS 60.766 \ -xCAR 96.87 -yCAR 118.74 \ -xLAB C13 -yLAB N15 \ -ndim 2 -aq2D Real \ -out ./test1.fid -verb -ov nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 -1 0 -time \ -verb -ov -out NC1.fid nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 0 -1 -time \ -verb -ov -out NC2.fid</pre>	<pre>#2D MAeSTOSO-8 'fid.com' script #Number of 13C t1 complex points=256 #Number of 15N t1 complex points=32 #Number of 15N t1 loops=8 varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 4096 \ -xT 2000 -yT 1024 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 33333.33 \ -xOBS 150.787 -yOBS 150.787 \ -xCAR 100.0 -yCAR 100.0 \ -xLAB C13 -yLAB C13 \ -ndim 2 -aq2D Real \ -out ./test.fid -verb -ov nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 0 0 1 0 0 0 -time \ -verb -ov -out CC.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 0 0 0 1 0 0 -time \ -verb -ov -out CNC1.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 1 0 0 0 1 0 -time \ -verb -ov -out CNC2.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 0 1 0 0 0 1 -time \ -verb -ov -out CNC3.fid varAdjust -in fid -out fid.adj -ov var2pipe -in fid.adj -noaswap -pw 0 \ -xN 4000 -yN 4096 \ -xT 2000 -yT 256 \ -xMODE Complex -yMODE complex \ -xSW 100000.000 -ySW 3333.33\ -xOBS 150.787 -yOBS 60.766 \ -xCAR 96.87 -yCAR 118.74 \ -xLAB C13 -yLAB N15 \ -ndim 2 -aq2D Real \ -out ./test1.fid -verb -ov nmrPipe -in test1.fid \ nmrPipe -fn COADD -axis Y -cList 1 0 0 0 -1 0 0 0 -time \ -verb -ov -out NC1.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 1 0 0 0 -1 0 0 -time \ -verb -ov -out NC2.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 1 0 0 0 -1 0 -time \ -verb -ov -out NC3.fid nmrPipe -in test.fid \ nmrPipe -fn COADD -axis Y -cList 0 0 0 1 0 0 0 -1 -time \ -verb -ov -out NC3.fid</pre>
--	--

Table 4
Definition of “yN” and “yT” used in [fid.com](#) script of Table 2

POE scheme	No. of FIDs per scan (nf)	No. of polarization pathways per FID (np)	Total no. of 2D experiments combined (nf*np)	No. of complex points (nc)	No. of t1 loops (n)	yN = 2nf*np*nc*n	yT
DUMAS	2	1	2	256 (for ¹³ C) 32 (for ¹⁵ N)	1 (for ¹³ C) 8 (for ¹⁵ N)	1024	256
MAcSTOSO-4	4	1	4	256 (for ¹³ C) 32 (for ¹⁵ N)	1 (for ¹³ C) 8 (for ¹⁵ N)	2048	256
MEIOSIS	2	2	4	256 (for ¹³ C) 32 (for ¹⁵ N)	1 (for ¹³ C) 8 (for ¹⁵ N)	2048	512
MAcSTOSO-8	4	2	8	256 (for ¹³ C) 32 (for ¹⁵ N)	1 (for ¹³ C) 8 (for ¹⁵ N)	4096	1024

(A) DUMAS:



(B) MAeSTOSO-4:

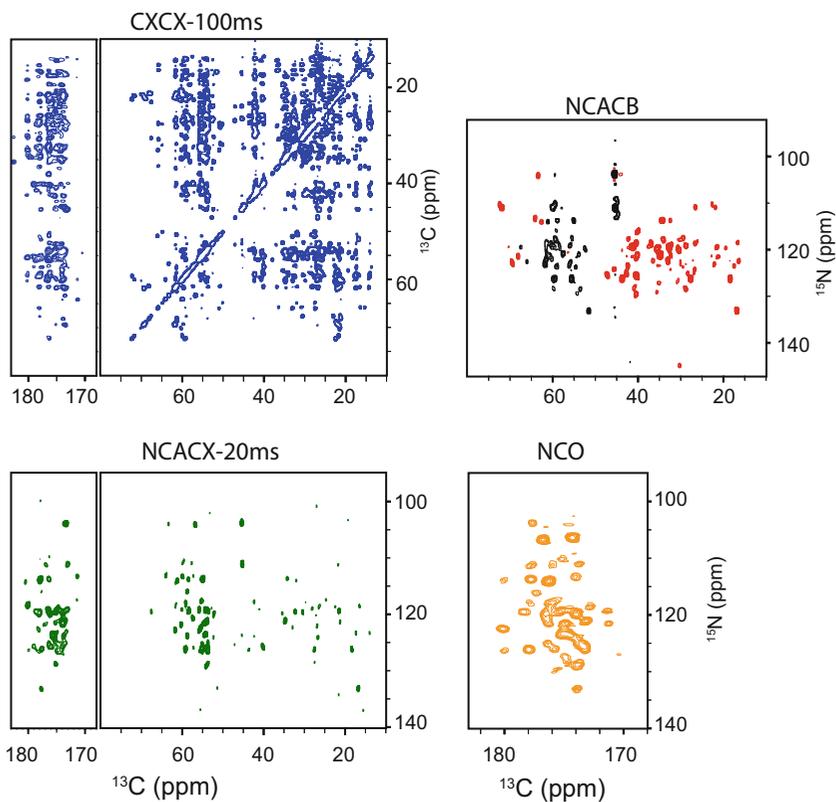


Fig. 4 (a) CXCX and NCA spectra of sarcolipin membrane protein acquired using 2D DUMAS pulse sequence of Fig. 2a. (b) Application of MAeSTOSO-4 (Fig. 2b) for Simultaneous acquisition of CXCX, NCACB, NCACX, and NCO spectra of ubiquitin microcrystalline protein

Table 5
NMRpipe script for adding the sub 2D data sets of NC.fid

#sum.com script	
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 64 -y1 1 -verb \ -verb -ov -out 1.fid	addNMR -in1 1.fid -in2 2.fid -out 12.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 128 -y1 65 -verb \ -verb -ov -out 2.fid	addNMR -in1 3.fid -in2 4.fid -out 34.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 192 -y1 129 -verb \ -verb -ov -out 3.fid	addNMR -in1 5.fid -in2 6.fid -out 56.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 256 -y1 193 -verb \ -verb -ov -out 4.fid	addNMR -in1 7.fid -in2 8.fid -out 78.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 320 -y1 257 -verb \ -verb -ov -out 5.fid	addNMR -in1 12.fid -in2 34.fid -out 1234.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 384 -y1 321 -verb \ -verb -ov -out 6.fid	addNMR -in1 56.fid -in2 78.fid -out 5678.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 448 -y1 385 -verb \ -verb -ov -out 7.fid	addNMR -in1 1234.fid -in2 5678.fid -out NC.fid -add
nmrPipe -in NC.fid \ nmrPipe -fn EXT -yn 512 -y1 449 -verb \ -verb -ov -out 8.fid	rm -r 1.fid 2.fid 3.fid 4.fid 5.fid 6.fid 7.fid 8.fid rm -r 1234.fid 5678.fid

Table 6
NMRpipe script for 2D Fourier transformation

#proc.com script #!/bin/csh
nmrPipe -in CC.fid \ nmrPipe -fn GM -g1 80 -g2 120 \ nmrPipe -fn ZF -size 16000 \ nmrPipe -fn FT \ nmrPipe -fn PS -p0 0.0 -p1.0 -di \ nmrPipe -fn EXT -x1 0ppm -xn 200ppm -sw \ nmrPipe -fn TP \ nmrPipe -fn ZF -size 1024 \ nmrPipe -fn FT \ nmrPipe -fn PS -p0 0.0 -p1 0.0 -di \ nmrPipe -fn EXT -x1 0ppm -xn 80ppm -sw \ nmrPipe -fn TP \ nmrPipe -fn POLY -auto \ -verb -ov -out cc.ft

5 Conclusions

POE presented in this chapter integrate various NMR pulse sequences using state-of-art NMR probe technology. This new class of experiments is not an alternative to the existing approaches to speed up data acquisition and increase sensitivity; rather it provides a clever strategy to concatenate multiple experiments into a single pulse sequence to optimize the timing of the spectrometers and speed up data acquisition through the acquisition of multiple 2D and 3D experiments. Utilizing POE-based strategies, for our proteins, we observed a reduction of approximately 50% of the experimental time compared to acquiring the corresponding experiments individually in a serial fashion. POE represent a general strategy for multiple acquisitions of almost all types of double- and triple-resonance ^{13}C detected experiments. When applied in concert with other fast acquisition and sensitivity enhancement techniques (e.g., Dynamic Nuclear Polarization, Paramagnetic Relaxation Enhancements, etc.), this approach can further push the boundaries of ssNMR applications to structural biology.

Acknowledgments

This research is supported by the National Institute of Health (GM 64742 and GM 72701 to G. V.). All of the NMR Experiments were carried out at the Minnesota NMR Center.

References

1. Ader C, Schneider R, Seidel K, Etzkorn M, Becker S, Baldus M (2009) Structural rearrangements of membrane proteins probed by water-edited solid-state NMR spectroscopy. *J Am Chem Soc* 131(1):170–176. doi:[10.1021/ja806306e](https://doi.org/10.1021/ja806306e)
2. 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(6911):98–102. doi:[10.1038/nature01070](https://doi.org/10.1038/nature01070)
3. Gustavsson M, Verardi R, Mullen DG, Mote KR, Traaseth NJ, Gopinath T, Veglia G (2013) Allosteric regulation of SERCA by phosphorylation-mediated conformational shift of phospholamban. *Proc Natl Acad Sci USA* 110(43):17338–17343. doi:[10.1073/pnas.1303006110](https://doi.org/10.1073/pnas.1303006110)
4. Hong M, Zhang Y, Hu F (2012) Membrane protein structure and dynamics from NMR spectroscopy. *Annu Rev Phys Chem* 63:1–24. doi:[10.1146/annurev-physchem-032511-143731](https://doi.org/10.1146/annurev-physchem-032511-143731)
5. Hu F, Luo W, Hong M (2010) Mechanisms of proton conduction and gating in influenza M2 proton channels from solid-state NMR. *Science* 330(6003):505–508. doi:[10.1126/science.1191714](https://doi.org/10.1126/science.1191714)
6. Wang S, Ladizhansky V (2014) Recent advances in magic angle spinning solid state NMR of membrane proteins. *Prog Nucl Magn Reson Spectrosc* 82:1–26. doi:[10.1016/j.pnmrs.2014.07.001](https://doi.org/10.1016/j.pnmrs.2014.07.001)
7. Zhou HX, Cross TA (2013) Influences of membrane mimetic environments on membrane protein structures. *Annu Rev Biophys* 42:361–392. doi:[10.1146/annurev-biophys-083012-130326](https://doi.org/10.1146/annurev-biophys-083012-130326)
8. Zachariae U, Schneider R, Briones R, Gattin Z, Demers JP, Giller K, Maier E, Zweckstetter M, Griesinger C, Becker S, Benz R, de Groot BL, Lange A (2012) β -barrel mobility underlies closure of the voltage-dependent anion channel. *Structure* 20(9):1540–1549. doi:[10.1016/j.str.2012.06.015](https://doi.org/10.1016/j.str.2012.06.015)
9. Vostrikov VV, Gustavsson M, Gopinath T, Mullen D, Dicke AA, Truong V, Veglia G (2016) Ca^{2+} ATPase conformational transitions in lipid bilayers mapped by site-directed Ethylation and

- solid-state NMR. *ACS Chem Biol* 11(2): 329–334. doi:10.1021/acschembio.5b00953
10. McDermott A (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu Rev Biophys* 38:385–403. doi:10.1146/annurev.biophys.050708.133719
 11. Mainz A, Peschek J, Stavropoulou M, Back KC, Bardiaux B, Asami S, Prade E, Peters C, Weinkauff S, Buchner J, Reif B (2015) The chaperone alphaB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat Struct Mol Biol* 22(11):898–905. doi:10.1038/nsemb.3108
 12. Gor'kov PL, Chekmenev EY, Li C, Cotten M, Buffy JJ, Traaseth NJ, Veglia G, Brey WW (2007) Using low-E resonators to reduce RF heating in biological samples for static solid-state NMR up to 900 MHz. *J Magn Reson* 185(1):77–93. doi:10.1016/j.jmr.2006.11.008
 13. McNeill SA, Gor'kov PL, Shetty K, Brey WW, Long JR (2009) A low-E magic angle spinning probe for biological solid state NMR at 750 MHz. *J Magn Reson* 197(2):135–144. doi:10.1016/j.jmr.2008.12.008
 14. Stringer JA, Bronnimann CE, Mullen CG, Zhou DH, Stellfox SA, Li Y, Williams EH, Rienstra CM (2005) Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes. *J Magn Reson* 173(1): 40–48. doi:10.1016/j.jmr.2004.11.015
 15. Maly T, Debelouchina GT, Bajaj VS, Hu KN, Joo CG, Mak-Jurkuskas ML, Sirigiri JR, van der Wel PC, Herzfeld J, Temkin RJ, Griffin RG (2008) Dynamic nuclear polarization at high magnetic fields. *J Chem Phys* 128(5):052211. doi:10.1063/1.2833582
 16. Wickramasinghe NP, Parthasarathy S, Jones CR, Bhardwaj C, Long F, Kotecha M, Mehboob S, Fung LW, Past J, Samoson A, Ishii Y (2009) Nanomole-scale protein solid-state NMR by breaking intrinsic IHT1 boundaries. *Nat Methods* 6(3):215–218. doi:10.1038/nmeth.1300
 17. Chevelkov V, Rehbein K, Diehl A, Reif B (2006) Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew Chem Int Ed Engl* 45(23):3878–3881. doi:10.1002/anie.200600328
 18. Chevelkov V, Xiang S, Giller K, Becker S, Lange A, Reif B (2015) Perspectives for sensitivity enhancement in proton-detected solid-state NMR of highly deuterated proteins by preserving water magnetization. *J Biomol NMR* 61(2): 151–160. doi:10.1007/s10858-015-9902-2
 19. Fu R, Brey WW, Shetty K, Gor'kov P, Saha S, Long JR, Grant SC, Chekmenev EY, Hu J, Gan Z, Sharma M, Zhang F, Logan TM, Bruschweiler R, Edison A, Blue A, Dixon IR, Markiewicz WD, Cross TA (2005) Ultra-wide bore 900 MHz high-resolution NMR at the national high magnetic field laboratory. *J Magn Reson* 177(1):1–8. doi:10.1016/j.jmr.2005.07.013
 20. Gopinath T, Veglia G (2012) Dual acquisition magic-angle spinning solid-state NMR-spectroscopy: simultaneous acquisition of multi-dimensional spectra of biomacromolecules. *Angew Chem Int Ed Engl* 51(11):2731–2735. doi:10.1002/anie.201108132
 21. Gopinath T, Veglia G (2013) Orphan spin operators enable the acquisition of multiple 2D and 3D magic angle spinning solid-state NMR spectra. *J Chem Phys* 138(18):184201. doi:10.1063/1.4803126
 22. Gopinath T, Veglia G (2016) Multiple acquisitions via sequential transfer of orphan spin polarization (MAeSTOSO): how far can we push residual spin polarization in solid-state NMR? *J Magn Reson* 267:1–8. doi:10.1016/j.jmr.2016.03.001
 23. Gopinath T, Veglia G (2016) Orphan spin polarization: a catalyst for high-throughput solid-state NMR spectroscopy of proteins. *Annu Rep NMR Spectrosc* 89: 103–121
 24. Mote KR, Gopinath T, Veglia G (2013) Determination of structural topology of a membrane protein in lipid bilayers using polarization optimized experiments (POE) for static and MAS solid state NMR spectroscopy. *J Biomol NMR* 57(2):91–102. doi:10.1007/s10858-013-9766-2
 25. Pines A, Gibby GM, Waugh JS (1973) Proton-enhanced NMR of dilute spins in solids. *J Chem Phys* 59:569–590
 26. Kupce E, Kay LE, Freeman R (2010) Detecting the “afterglow” of ^{13}C NMR in proteins using multiple receivers. *J Am Chem Soc* 132(51): 18008–18011. doi:10.1021/ja1080025
 27. Bellstedt P, Herbst C, Hafner S, Leppert J, Gorlach M, Ramachandran R (2012) Solid state NMR of proteins at high MAS frequencies: symmetry-based mixing and simultaneous acquisition of chemical shift correlation spectra. *J Biomol NMR* 54(4):325–335. doi:10.1007/s10858-012-9680-z
 28. Banigan JR, Traaseth NJ (2012) Utilizing after-glow magnetization from cross-polarization magic-angle-spinning solid-state NMR spectroscopy to obtain simultaneous heteronuclear multidimensional spectra. *J Phys Chem B* 116(24): 7138–7144. doi:10.1021/jp303269m
 29. Andreas LB, Le Marchand T, Jaudzems K, Pintacuda G (2015) High-resolution proton-detected NMR of proteins at very fast MAS. *J*

- Magn Reson 253:36–49.
doi:[10.1016/j.jmr.2015.01.003](https://doi.org/10.1016/j.jmr.2015.01.003)
30. Paulson EK, Morcombe CR, Gaponenko V, Dancheck B, Byrd RA, Zilm KW (2003) Sensitive high resolution inverse detection NMR spectroscopy of proteins in the solid state. *J Am Chem Soc* 125(51):15831–15836. doi:[10.1021/ja037315+](https://doi.org/10.1021/ja037315+)
 31. Buck B, Zamoon J, Kirby TL, DeSilva TM, Karim C, Thomas D, Veglia G (2003) Over-expression, purification, and characterization of recombinant Ca-ATPase regulators for high-resolution solution and solid-state NMR studies. *Protein Expr Purif* 30(2):253–261
 32. Bennett AE, Rienstra CM, Auger M, Lakshmi KV, Griffin RG (1995) Heteronuclear decoupling in rotating solids. *J Chem Phys* 103(16):6951–6958. doi:[10.1063/1.470372](https://doi.org/10.1063/1.470372)
 33. Hartmann SR, Hahn EL (1962) Nuclear double resonance in the rotating frame. *Phys Rev* 128(5):2042–2053
 34. Detken A, Hardy EH, Ernst M, Kainosho M, Kawakami T, Aimoto S, Meier BH (2001) Methods for sequential resonance assignment in solid, uniformly ^{13}C , ^{15}N labelled peptides: quantification and application to antamanide. *J Biomol NMR* 20(3):203–221.
 35. Franks WT, Kloepper KD, Wylie BJ, Rienstra CM (2007) Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins. *J Biomol NMR* 39(2):107–131. doi:[10.1007/s10858-007-9179-1](https://doi.org/10.1007/s10858-007-9179-1)
 36. Verel R, Ernst M, Meier BH (2001) Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme. *J Magn Reson* 150(1):81–99. doi:[10.1006/jmre.2001.2310](https://doi.org/10.1006/jmre.2001.2310)
 37. Takegoshi K, Nakamura S, Terao T (2001) ^{13}C - ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem Phys Lett* 344(5–6):631–637. doi:[10.1016/S0009-2614\(01\)00791-6](https://doi.org/10.1016/S0009-2614(01)00791-6)
 38. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293

Afterglow Solid-State NMR Spectroscopy

Gili Abramov and Nathaniel J. Traaseth

Abstract

Biomolecular solid-state NMR experiments have traditionally been collected through detection of ^{13}C or ^{15}N nuclei. Since these nuclei have relatively low sensitivity stemming from their smaller gyromagnetic ratios relative to ^1H , the time required to collect multi-dimensional datasets serves as a limitation to resonance assignment and structure determination. One improvement in the field has been to employ simultaneous or parallel acquisition techniques with the goal of acquiring more than one dataset at a time and therefore speeding up the overall data collection process. Central to these experiments is the cross-polarization (CP) element, which serves as a way to transfer magnetization between nuclei via magnetic dipolar couplings. In this chapter, we show how residual signal remaining after CP is a polarization source that can be used to acquire additional datasets. The setup of this class of experiments, referred to as *Afterglow* spectroscopy, is described and demonstrated using a membrane protein transporter involved in multidrug resistance.

Key words NMR spectroscopy, Magic-angle-spinning, Solid-state NMR, Sensitivity enhancement, Multiple receiver detection, Membrane proteins, Multidrug resistance

1 Introduction

Solid-state NMR (ssNMR) is a technique used to obtain atomic-scale information on a range of solid-like materials, including amorphous powders and materials, nanoparticles [1], and biomolecular assemblies, such as fibrous aggregates [2], membrane proteins [3–5], viruses [6], and intact cells [7]. One of the major advantages of ssNMR is the ability to probe macromolecules under native-like environments. NMR spectra of solid samples are dominated by anisotropic nuclear spin interactions, which provide insight into conformation and dynamics on a wide range of timescales. These interactions also complicate the spectra by giving broader spectral peaks, thus preventing site-specific resolution. Lowe and Andrew showed that spectra can be simplified by manipulating the angular part of these interactions, $P_2(\cos\beta_m)$ [8, 9] where P_2 is the 2nd Legendre polynomial. Spinning the sample faster than the size of the anisotropic interaction at the angle $\beta_m = 54.7^\circ$ relative to the

magnetic field led to a remarkable improvement in the observed linewidths. This popular technique is known as magic-angle spinning (MAS).

The ongoing developments in MAS ssNMR methodology continue to break new boundaries, raising the level of complexity of systems amenable to the technique. The introduction of selective isotope labeling schemes [10], ^1H detection [11], progresses in structure calculation protocols [12, 13], and access to high field and fast spinning instrumentation allows for the acquisition of high-resolution spectra of bio-macromolecular assemblies [14]. Nevertheless, challenges remain for increasing the sensitivity of detecting low-gyromagnetic nuclei such as ^{15}N and ^{13}C . Efforts toward speeding up data acquisition while maximizing the amount of information gained from an individual data set include the development of nonuniform sampling (NUS) techniques [15], improvements in MAS probe technology [16], and the usage of multiple receivers for simultaneous acquisition techniques [17].

Solution NMR methodology has also focused on improving sensitivity. On this front, it was shown by Kupčič et al. that the weak ^{13}C signal at the “tail” of a free induction decay can be transferred to ^1H and subsequently detected [18]. This residual or “afterglow” magnetization results in the collection of an additional dataset through the use of a second receiver configured for the ^1H channel. The advantage of this approach is that the recycle delay is much longer than the coherence transfer and detection steps and therefore two datasets can be obtained for the total experimental acquisition time of one. The presence of unused residual polarization was also recognized by Pines et al. in ssNMR experiments involving proton enhanced sequences, where multiple CP free induction decay signals were co-added [19]. However, the residual magnetization was largely ignored as pulse sequence development focused on transferred magnetization and not on signals left behind. In 2012, our lab proposed an approach to use residual signal to boost sensitivity by making use of ^{15}N polarization remaining after a frequency-selective cross-polarization period [20]. In this experiment, we were able to detect two multidimensional datasets that correlated ^{15}N with ^{13}CA and ^{15}N with ^{13}CO within proteins by making use of relatively long ^{15}N $T_{1\rho}$ and T_1 relaxation times in motionally restricted samples. The final result gave two complementary heteronuclear correlation datasets without any sensitivity loss in the first experiment and without the need for multiple receivers. It is important to note that a complementary but alternative approach to enhancing polarization was proposed by Gopinath and Veglia, referred to as DUMAS [21]. This technique makes use of simultaneous cross polarization from ^1H to both ^{15}N and ^{13}C , a shared acquisition period, and subsequent transfer of magnetization for ^{13}C detection. This powerful method has been combined with the afterglow detection approach to produce up to eight

datasets acquired at the same time [22, 23]. Additional efforts in the field have made use of residual polarization in combination with triple cross-polarization periods [24–26].

In the following sections, we describe the steps required for conducting afterglow N-CA/CO experiments in proteins. Since we cannot ignore the importance of all steps in acquiring afterglow spectra, we detail our step-by-step procedure of experimental setup and demonstrate the methodology on a membrane protein (EmrE) sample involved in multidrug resistance.

2 Materials

2.1 Sample Preparation

Reagents used for the EmrE MAS sample preparation have been described previously [27, 28] and were purchased from different sources as detailed below.

1. $^{13}\text{C}_6$ -glucose and $^{15}\text{NH}_4\text{Cl}$, both ~99%. For reverse labeling of specific amino acids, e.g., isoleucine and leucine, unlabeled amino acids were added to the growth medium.
2. *n*-dodecyl- β -D-maltoside (DDM).
3. EmrE was reconstituted into 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC).
4. *n*-octyl- β -D-glucoside (OG).
5. Bio-Beads™ SM-2 resin (Bio-Rad Laboratories).
6. Final sample preparation of EmrE involved centrifugation in an Optima™ MAX-XP Ultracentrifuge equipped with TLA-100 and TLA-110 rotors, Beckman Coulter.

2.2 NMR Instrumentation and Data Analysis

1. Samples were packed in 3.2 mm thin-walled zirconia MAS rotors with a volume capacity of ~36 μl . These rotors are capable of spinning up to a maximum rate of ~18 kHz (Agilent or Revolution NMR).
2. All MAS ssNMR spectra were acquired on an Agilent DD2 spectrometer operating at 14.1 T, corresponding to a ^1H Larmor frequency of 600 MHz. All experiments used a bio-MAS probe in triple resonance $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ configuration.
3. For data processing and analysis, NMRPipe [29] and Sparky [30] were used, respectively.

3 Methods

3.1 Adjustment of the Magic Angle

The experimental setup for MAS initiates with the adjustment of the magic angle using natural abundance ^{13}C glycine (*see Note 1*). Specifically, the linewidth of the ^{13}CO signal (~178 ppm) from powdered/crystalline glycine is minimized while adjusting the

magic angle (*see Note 2*). The carbonyl site has a large chemical shift anisotropy and is therefore sensitive to the rotor angle with respect to the magnetic field. The steps below are carried out after determination of the probe power limitations and the 90° pulse widths using isotopically enriched model compounds (*see Note 3*).

1. Place a full rotor of natural abundance glycine into the MAS probe.
2. Find the optimal contact time and power values for ^1H to ^{13}C cross-polarization (CP) to ensure sufficient signal-to-noise.
3. Place the spectrometer in “FID scan” mode (Agilent) to display the Fourier transformed 1D ^{13}C spectrum after each transient. This mode is synonymous with the “gs” command on Bruker spectrometers. Ensure no window function is applied prior to the Fourier transform that would broaden the spectrum. Set the zero-fill number to 65,536 or larger to provide sufficient digital resolution (“si” on Bruker; “fn” on Agilent).
4. Adjust the magic angle in an iterative fashion to obtain a linewidth on the ^{13}CO signal of less than ~40 Hz under a spinning frequency of 12.5 kHz. We also typically use 12.5 kHz MAS for protein backbone triple resonance experiments in our 3.2 mm bio-MAS probe [31] at a magnetic field of 14.1 T (600 MHz ^1H frequency).

3.2 Shimming and Chemical Shift Referencing

Analogous to solution NMR experiments, it is imperative to have a homogeneous magnetic field to achieve narrow spectral lines. Similarly, it is important to have a reliable chemical shift reference before proceeding with an unknown sample. Both of these steps for MAS are described below.

1. Acquire a 1D ^{13}C CP spectrum of a powder sample of adamantane. Due to the excellent linewidths achievable with adamantane, it is necessary to acquire the FID for ~150 ms and lower the power of the ^1H decoupling to a value compatible with the MAS probe.
2. Two signals should be observed in the ^{13}C spectrum. Reference the more deshielded peak to 40.48 ppm [32] (*see Note 4*).
3. Using the same adamantane sample, adjust the room temperature shim values to obtain homogeneous and narrow spectral lines. This is done in an iterative manner similar to that described for the glycine sample. A linewidth of 7 Hz or lower can be achieved using the CH_2 peak of adamantane.

3.3 Optimization of CP-MAS on the Protein Sample

Nearly all ssNMR experiments begin with a polarization transfer from the abundant ^1H spins to the more insensitive $^{13}\text{C}/^{15}\text{N}$ spins. Below is the way we optimize CP transfers using a one-dimensional (1D) CP-MAS experiment for ^1H to ^{15}N spins.

1. The ^1H carrier frequency is set on resonance with the water peak in a direct excitation experiment (4.5–5.0 ppm)
2. RF pulses on the X-channel are first calibrated using model compounds to obtain a full nutation curve for each nucleus. To determine optimal flip angles for the protein sample, a CP sequence followed by a 90° (or 270°) pulse on the X-channel under ^1H decoupling is applied. The pulse power is adjusted to give a null signal in this spectrum with the optimal value corresponding to the 90° (or 270°) pulse.
3. A ^1H nutation frequency of 100 kHz ($\omega_{1,\text{H}}/2\pi$) corresponding to a pulse width of 2.5 μs and a flip angle of 90° is optimized by varying the power level for a 360° pulse (i.e., 10 μs). This is done in a CP experiment by detecting on the X-channel, typically ^{13}C .
4. Optimize CP for ^1H to ^{15}N . The power-levels for the ^1H - ^{15}N CP pulses are adjusted by matching the Hartmann-Hahn condition under spinning $\omega_{1,^1\text{H}} = \omega_{1,^{15}\text{N}} + n\omega_r$, where ω_1 is the nutation frequency for each of the nuclei and ω_r is the MAS rate. The CP signal is optimized by varying the field strength associated with each channel around the calculated values, while keeping the other channel's power constant.
5. Optimize the CP contact time between ^1H and ^{15}N . The time for Hartmann-Hahn spin-lock is arrayed from 0.2 ms to approximately 2 ms. Typically, the optimal value for membrane proteins and crystalline soluble protein samples is 0.75–1.0 ms.

3.4 Detection of Rotary Resonance Conditions

The rotary resonance condition is the matching of a spin-lock nutation field with the rotational rate used for MAS [33]. This match leads to dephasing by means of chemical shift anisotropy (CSA) or dipolar coupling. Since spin-lock periods are used in CP transfers, a major goal of our setup is to avoid the rotary resonance condition to ensure the most efficient transfers from ^{15}N to ^{13}C or vice versa. The pulse sequence we use for detecting the rotary resonance conditions is shown in Fig. 1a [34]. Different than the previous uses of rotary resonance conditions for facilitating magnetization transfer, our goal of using this pulse sequence is to avoid conditions for the double CP transfer steps (*see* Subheading 3.5). Below are steps to find the optimal half-integer conditions of nutation frequencies relative to the spinning rate. The optimization is demonstrated in Fig. 1b, showing an array of 1D experiments applied to the membrane protein EmrE under different ω_1 irradiation powers.

1. Place a [^{13}C , ^{15}N] labeled protein sample into the MAS probe.
2. Find an optimal ^1H to ^{15}N CP condition by using the Hartmann-Hahn matching condition under MAS (*see* Subheading 3.3).

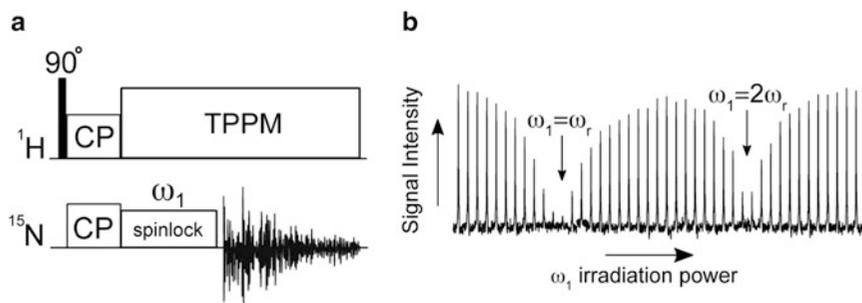


Fig. 1 Empirical detection of rotary resonance conditions. (a) Pulse scheme of the rotary resonance experiment. (b) Variation of the ω_1 amplitude around the two matching conditions ($\omega_1 = n\omega_r$, $n = 1$ or 2). The experiment is demonstrated on the ^{15}N amide signal in [U- ^{15}N , ^{13}C] labeled EmrE protein, using a 2 ms pulse length on ^{15}N

- Find conditions that meet the rotary resonance condition: $\omega_1 = n\omega_r$ ($n = 1, 2$). Set up an array where the continuous-wave (CW) ^{15}N RF powers are adjusted, as shown in Fig. 1b. This is a $T_{1\rho}$ experiment that should be applied with a ^{15}N spin-lock of 2–5 ms. The lowest signals shown in Fig. 1b correspond to the rotary resonance conditions, while the maximum signal intensities following the spin-lock are ideal conditions (i.e., half-integer) to be used in double CP experiments described in subsequent sections.

3.5 Optimization of the DCP Transfers ^{15}N - $^{13}\text{CA}/^{13}\text{CO}$

Most heteronuclear correlations in multi-dimensional spectra of solid-like biological samples originate from ^{15}N - ^{13}C dipolar couplings. These contacts are usually obtained by selective transfer techniques, correlating the ^{15}N chemical shift with its neighboring ^{13}CA and ^{13}CO chemical shifts. Selective N-CA *or* N-CO transfers are achieved by double cross polarization (DCP or “SPECIFIC CP”) Schemes [35, 36]. To achieve a band selective transfer, the irradiation conditions for both rare spin channels are optimized in a 1D N-CA/CO experiment as follows:

- For band selective excitation, set the carrier frequency to the desired spectral region, i.e., CA at ~ 55 ppm or CO at ~ 178 ppm.
- To achieve magnetization transfer between ^{15}N and ^{13}C nuclei, we initially set the RF amplitudes to $1.5\omega_r$ and $2.5\omega_r$ for ^{13}CA and ^{15}N channels, and $3.5\omega_r$ and $2.5\omega_r$ for ^{13}CO and ^{15}N channels, respectively (*see Note 5*). The DCP condition for ^{15}N is set using the steps outlined in Subheading 3.4, which is an empirical way to avoid rotary resonance conditions. In the case for ^{13}CO , it is preferable to apply a higher power on the ^{13}C channel due to the weaker ^1H - ^{13}CO dipolar couplings than those present for ^{13}CA [37, 38].

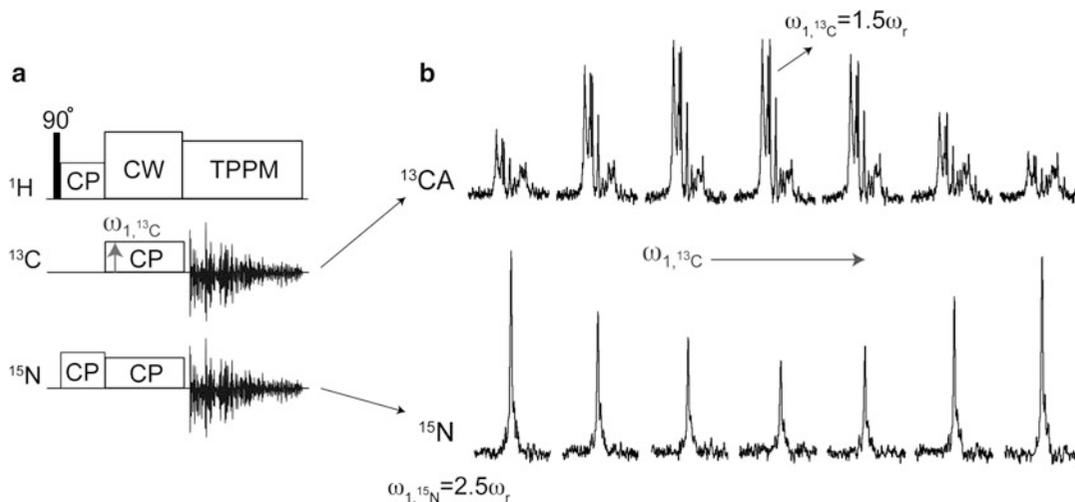


Fig. 2 Optimization of DCP ^{15}N to ^{13}CA transfer. **(a)** Pulse sequence used for optimizing the DCP from ^{15}N to ^{13}CA . The multiple receiver detection is not required for the setup but used to emphasize the reduction in ^{15}N signal as the ^{13}CA signal builds up. **(b)** The ^{13}C amplitude is varied around the calculated $1.5\omega_r$ condition. *Top:* ^{13}CA signal variation. *Bottom:* ^{15}N signal as the ^{13}CA RF amplitude is varied

- Once the initial power level for the DCP condition is calculated for ^{13}C , a careful optimization of the RF amplitude on the ^{13}C channel is carried out using a pulse sequence similar to that shown in Fig. 2a. An example for the optimization of the DCP condition in N-CA experiment is shown in Fig. 2b. In this experiment we have carried out acquisitions on both ^{15}N and ^{13}C in a simultaneous fashion using multiple receivers. This is not required for the set-up, but is used to emphasize that optimal transfer to ^{13}C gives the highest signal-to-noise in 1D datasets and the lowest signal intensities remaining on ^{15}N . Notably, even after finding optimal transfer conditions, residual polarization remains on ^{15}N and is the basis for our afterglow experiments (*see* Subheading 3.7). Lastly, when optimizing selective transfers, it is important to obtain the highest possible signal intensities while ensuring selectivity to either ^{13}CA or ^{13}CO .
- DCP contact time optimization. The time for which the ^{15}N and ^{13}C simultaneous pulses are applied is adjusted from 2 to 6 ms. We normally find optimal values from 4 to 5.5 ms.

3.6 Optimization of the Mixing Period for Optional N-CX Correlations

It is often desirable to establish correlations between ^{15}N and side-chain ^{13}C atoms. These contacts are obtained by applying a mixing period prior to the detection of the FID acquisition. Magnetization can be transferred efficiently from CA/CO to nearby ^{13}C atoms by utilizing the dense proton network, commonly present in protein samples (proton-driven spin-diffusion, PDS). The transfer efficiencies can be further improved upon application of a CW irradiation to

^1H during the mixing time at a frequency match corresponding to $\omega_{1,H} = n\omega_r$ ($n = 1$ or 2) (i.e., the DARR/RAD condition [39, 40]). The DARR condition is optimized by varying the CW ^1H power around the $n = 1$ or $n = 2$ DARR values in a 1D ^{13}C CP-MAS experiment to ensure efficient transfers among ^{13}C spins. The optimal conditions correspond to better dipolar recoupling characterized by a minimum signal in the array for ^{13}C nuclei receiving magnetization from ^{15}N (i.e., CA or CO).

3.7 Acquisition of Afterglow ^{15}N - ^{13}C CA/ ^{15}N - ^{13}C CO

The afterglow pulse scheme is shown in Fig. 3 and involves two acquisition periods for the simultaneous detection of 2D N-CA (first acquisition) and 2D N-CO (second acquisition, using residual ^{15}N polarization). To achieve selective CA- or CO-only detection, the DARR mixing period is set to zero. The optimal pulse widths/powers and initial CP and DCP conditions are found as described above, and are implemented into the current experiment. The two datasets are separated by a ^{13}C transverse magnetization dephasing period, followed by a 90° pulse on the ^{15}N channel to store the residual ^{15}N magnetization along the z-axis after the first DCP. The N-CO experiment is then initiated by a DCP selective transfer from the remaining ^{15}N signal to ^{13}C CO spins. The latter dataset is effectively “free of charge” and its sensitivity can be evaluated by comparing to the usual N-CA/CO datasets.

The experimental parameters for the afterglow experiment are essentially the optimized parameters described in Subheadings 3.3–3.6. Nevertheless, in practice, it is necessary to re-optimize the power levels on ^{13}C CO in the afterglow experiment in order to achieve maximal signal-to-noise. Figure 4 shows the comparison of

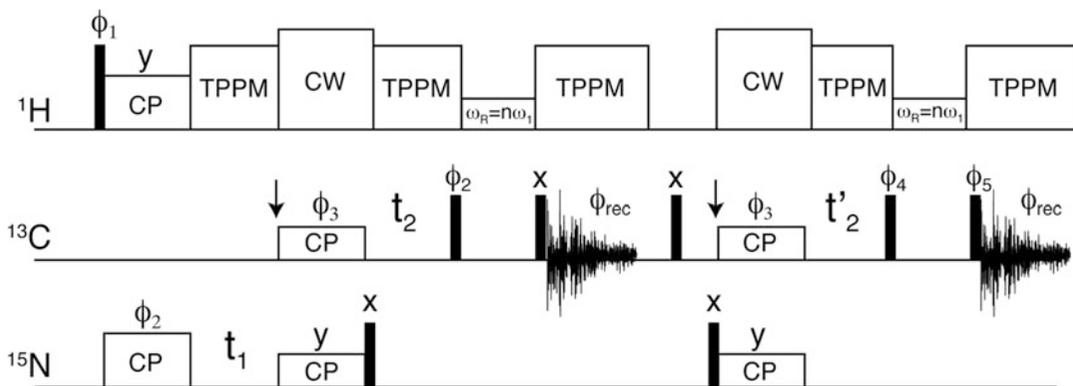


Fig. 3 Afterglow pulse sequence for acquiring N-CA/N-CO in a sequential fashion. The narrow rectangles correspond to 90° pulses. The left and right arrows within the ^{13}C channel signify the offset positioned on the ^{13}CA and ^{13}CO regions, respectively. Phases are: $\phi_1 = (x, -x)$, $\phi_2 = (y)$, $\phi_3 = (x, x, y, y)$, $\phi_4 = (-y, -y, -x, -x)$, $\phi_5 = (y, y, -x, -x)$, and $\phi_{\text{rec}} = (x, -x, -y, y)$. To obtain phase-sensitive data in t_1 and t_2 , ϕ_2 and ϕ_3 were phase-shifted by 90° , respectively. After the first FID acquisition, a 5 ms time was allowed to dephase residual ^{13}C magnetization. (Reprinted with permission from [20]. © 2012 American Chemical Society)

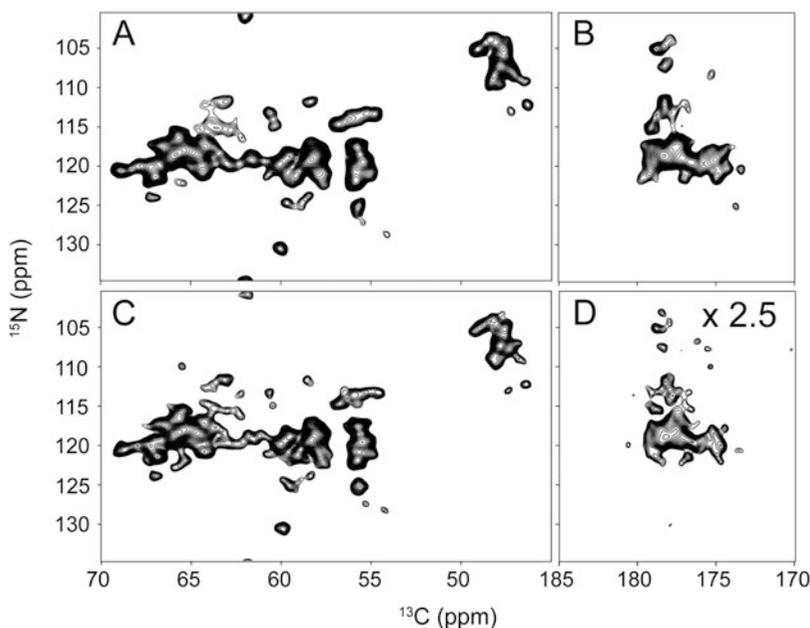


Fig. 4 Two-dimensional N-CA (a, c) and N-CO (b, d) spectra of uniformly ^{15}N and ^{13}C enriched EmrE. Standard N-CA spectrum (a) and N-CA as part of the Afterglow scheme (c) gave similar signal-to-noise ratios. Standard N-CO (b) and Afterglow N-CO (d). The latter is multiplied by 2.5 to compare with the standard experiment. Overall, the Afterglow N-CO results in about 30–35% of the signal of the standard N-CO for $[\text{U-}^{15}\text{N}, ^{13}\text{C}]$ labeled samples

standard 2D N-CA/CO with the afterglow experiments collected using the four transmembrane domain protein EmrE in DMPC liposomes.

3.8 Benefit of Spectroscopic Filtering with the Use of Afterglow

NMR is an inherently insensitive technique. Acquisition of multi-dimensional datasets on low gyromagnetic ratio nuclei presents additional challenges that further increase experimental times for protein samples. Thus, the development of parallel acquisition techniques is of high importance in the field. The afterglow scheme is advantageous for the following reasons:

1. It utilizes residual magnetization produced in the course of a “regular” heteronuclear experiment to obtain another dataset that can be complementary for the assignment process of biomolecules.
2. It is a simple scheme that does not require specific hardware modifications, and can be applied in any conventional ssNMR spectrometer.
3. The afterglow sequence acquires two ^{15}N - ^{13}C correlation spectra using a single recycle delay, which effectively maintains the same overall experimental time with respect to standard datasets.

We have described how Afterglow magnetization from residual ^{15}N signal at the end of a DCP period in 2D N-CA is utilized to obtain a high-quality 2D N-CO spectrum. The comparison between the standard N-CA/CO and the Afterglow N-CA/CO (Fig. 4) acquired on a membrane protein shows the applicability of the methodology to a complex biological system. The signal-to-noise in the afterglow N-CO experiment can be dramatically improved with the usage of sparse isotope labeling schemes (e.g., 1,3- ^{13}C or 2- ^{13}C glycerol labeling) due to the effect of spin-dilution [41]. Lastly, it is also possible to use the afterglow technique in combination with reverse labeling to further seed chemical shift assignments of congested spectra [41] and for measuring spin relaxation times [42].

4 Notes

1. It is also common to use potassium bromide (KBr) while detecting the ^{79}Br signal.
2. The state of glycine crystallinity or polycrystallinity can impact the observed linewidth.
3. Throughout the set-up process it is important to emphasize that high-power probe tuning is performed to minimize the ratio of forward and reflected powers using an oscilloscope. The ratio between the forward/reflected powers should be equal to or larger than $\sim 25/1$.
4. The ^{15}N chemical shift spectrum can be indirectly referenced from the ^{13}C adamantane by using the ratio of reference frequencies for ^{15}N and ^{13}C (0.402979954), as previously described [43, 44].
5. To match up the DCP condition, RF amplitudes must be applied that both satisfy the Hartmann-Hahn condition and avoid the rotary resonance condition. Therefore, frequency irradiation of half-integer multiples of the spinning rate is usually applied to the rare-spins channels, while a high-power CW pulse is simultaneously applied to the abundant ^1H channel.

Acknowledgments

This work was supported by NSF (MCB1506420) and NIH (R01AI108889).

References

1. Gutmann T, Grünberg A, Rothermel N et al (2013) Solid-state NMR concepts for the investigation of supported transition metal catalysts and nanoparticles. *Solid State Nucl Magn Reson* 55:1–11
2. Heise H, Hoyer W, Becker S et al (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length α -synuclein fibrils studied by solid-state NMR. *Proc Natl Acad Sci USA* 102:15871–15876
3. Opella SJ, Marassi FM (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chem Rev* 104:3587–3606
4. Hong M, Zhang Y, Hu F (2012) Membrane protein structure and dynamics from NMR spectroscopy. *Annu Rev Phys Chem* 63:1–24
5. Wylie BJ, Do HQ, Borcik CG, Hardy EP (2016) Advances in solid-state NMR of membrane proteins. *Mol Phys* 114(24):3598–3609
6. Quinn CM, Lu M, Suiter CL et al (2015) Magic angle spinning NMR of viruses. *Prog Nucl Magn Reson Spectrosc* 86:21–40
7. Takahashi H, Ayala I, Bardet M et al (2013) Solid-state NMR on bacterial cells: selective cell wall signal enhancement and resolution improvement using dynamic nuclear polarization. *J Am Chem Soc* 135:5105–5110
8. Andrew ER, Bradbury A, Eades GR (1958) Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* 182:1659–1659
9. Lowe IJ (1959) Free induction decays of rotating solids. *Phys Rev Lett* 2:285–287
10. Hong M (1999) Determination of multiple ϕ -torsion angles in proteins by selective and extensive ^{13}C labeling and two-dimensional solid-state NMR. *J Magn Reson* 139:389–401
11. Zhou DH, Shah G, Cormos M et al (2007) Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning. *J Am Chem Soc* 129:11791–11801
12. Loquet A, Sgourakis NG, Gupta R et al (2012) Atomic model of the type III secretion system needle. *Nature* 486:276–279
13. Wang S, Munro RA, Shi L et al (2013) Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat Methods* 10:1007–1012
14. Bertini I, Emsley L, Lelli M et al (2010) Ultrafast MAS solid-state NMR permits extensive ^{13}C and ^1H detection in paramagnetic metalloproteins. *J Am Chem Soc* 132:5558–5559
15. Paramasivam S, Suiter CL, Hou G et al (2012) Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *J Phys Chem B* 116:7416–7427
16. McNeill SA, Gor'kov PL, Shetty K et al (2009) A low-E magic angle spinning probe for biological solid state NMR at 750 MHz. *J Magn Reson* 197:135–144
17. Kupče Ě (2011) NMR with multiple receivers. In: *Mod. NMR Methodol.* Springer, Berlin, pp 71–96
18. Kupče E, Kay LE, Freeman R (2010) Detecting the “afterglow” of ^{13}C NMR in proteins using multiple receivers. *J Am Chem Soc* 132:18008–18011
19. Pines A, Gibby MG, Waugh JS (1973) Proton-enhanced NMR of dilute spins in solids. *J Chem Phys* 59:569–590
20. Banigan JR, Traaseth NJ (2012) Utilizing afterglow magnetization from cross-polarization magic-angle-spinning solid-state NMR spectroscopy to obtain simultaneous Heteronuclear multidimensional spectra. *J Phys Chem B* 116:7138–7144
21. Gopinath T, Veglia G (2012) Dual acquisition magic-angle spinning solid-state NMR-spectroscopy: simultaneous acquisition of multidimensional spectra of biomacromolecules. *Angew Chem Int Ed Engl* 51:2731–2735
22. Gopinath T, Veglia G (2013) Orphan spin operators enable the acquisition of multiple 2D and 3D magic angle spinning solid-state NMR spectra. *J Chem Phys* 138:184201
23. Gopinath T, Veglia G (2016) Multiple acquisitions via sequential transfer of orphan spin polarization (MAeSTOSO): how far can we push residual spin polarization in solid-state NMR? *J Magn Reson* 267:1–8
24. Mote KR, Gopinath T, Veglia G (2013) Determination of structural topology of a membrane protein in lipid bilayers using polarization optimized experiments (POE) for static and MAS solid state NMR spectroscopy. *J Biomol NMR* 57:91–102
25. Das BB, Opella SJ (2016) Simultaneous cross polarization to ^{13}C and ^{15}N with ^1H detection at 60kHz MAS solid-state NMR. *J Magn Reson* 262:20–26
26. Akbey Ü, Camponeschi F, van Rossum B-J, Oschkinat H (2011) Triple resonance cross-polarization for more sensitive ^{13}C MAS NMR spectroscopy of Deuterated proteins. *ChemPhysChem* 12:2092–2096

27. Banigan JR, Gayen A, Cho M-K, Traaseth NJ (2015) A structured loop modulates coupling between the substrate-binding and dimerization domains in the multidrug resistance transporter EmrE. *J Biol Chem* 290:805–814
28. Banigan JR, Gayen A, Traaseth NJ (2015) Correlating lipid bilayer fluidity with sensitivity and resolution of polytopic membrane protein spectra by solid-state NMR spectroscopy. *Biochim Biophys Acta* 1848:334–341
29. Delaglio F, Grzesiek S, Vuister GW et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
30. Goddard TD, Kneller DG SPARKY 3. <https://www.cgl.ucsf.edu/home/sparky/>
31. Stringer JA, Bronnimann CE, Mullen CG et al (2005) Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes. *J Magn Reson* 173:40–48
32. Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. *J Magn Reson* 162:479–486
33. Oas TG, Griffin RG, Levitt MH (1988) Rotary resonance recoupling of dipolar interactions in solid-state nuclear magnetic resonance spectroscopy. *J Chem Phys* 89:692–695
34. Gan Z (2006) Rotary resonance echo double resonance for measuring heteronuclear dipolar coupling under MAS. *J Magn Reson* 183:235–241
35. Schaefer J, McKay R, Stejskal E (1979) Double-cross-polarization NMR of solids. Academic Press, New York
36. Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol Phys* 95:1197–1207
37. Ishii Y, Ashida J, Terao T (1995) ^{13}C - ^1H dipolar recoupling dynamics in ^{13}C multiple-pulse solid-state NMR. *Chem Phys Lett* 246:439–445
38. Franks WT, Kloepper KD, Wylie BJ, Rienstra CM (2007) Four-dimensional heteronuclear correlation experiments for chemical shift assignment of solid proteins. *J Biomol NMR* 39:107–131
39. Takegoshi K, Nakamura S, Terao T (2001) ^{13}C - ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem Phys Lett* 344:631–637
40. Morcombe CR, Gaponenko V, Byrd RA, Zilm KW (2004) Diluting abundant spins by isotope edited radio frequency field assisted diffusion. *J Am Chem Soc* 126:7196–7197
41. Banigan JR, Gayen A, Traaseth NJ (2013) Combination of ^{15}N reverse labeling and after-glow spectroscopy for assigning membrane protein spectra by magic-angle-spinning solid-state NMR: application to the multidrug resistance protein EmrE. *J Biomol NMR* 55:391–399
42. Good D, Pham C, Jagas J et al. (2017) Solid-state NMR provides evidence for small-amplitude slow domain motions in a multispinning transmembrane α -Helical Protein. *J Am Chem Soc* 139(27):9246–9258.
43. Harris RK, Becker ED, Cabral de Menezes SM et al (2002) NMR nomenclature: nuclear spin properties and conventions for chemical shifts. IUPAC recommendations 2001. International Union of Pure and Applied Chemistry. Physical chemistry division. Commission on molecular structure and spectroscopy. *Magn Reson Chem* 40:489–505
44. Siemer AB, Ritter C, Steinmetz MO et al (2006) ^{13}C , ^{15}N resonance assignment of parts of the HET-s prion protein in its amyloid form. *J Biomol NMR* 34:75–87

Chapter 4

Filamentous Bacteriophage Viruses: Preparation, Magic-Angle Spinning Solid-State NMR Experiments, and Structure Determination

Omry Morag, Nikolaos G. Sgourakis, Gili Abramov, and Amir Goldbourt

Abstract

Filamentous bacteriophages are elongated semi-flexible viruses that infect bacteria. They consist of a circular single-stranded DNA (ssDNA) wrapped by a capsid consisting of thousands of copies of a major coat protein subunit. Given the increasing number of discovered phages and the existence of only a handful of structures, the development of methods for phage structure determination is valuable for biophysics and structural virology. In recent years, we developed and applied techniques to elucidate the 3D atomic-resolution structures of intact bacteriophages using experimental magic-angle spinning (MAS) solid-state NMR data. The flexibility in sample preparation – precipitated homogeneous solids – and the fact that ssNMR presents no limitation on the size, weight or morphology of the system under study makes it an ideal approach to study phage systems in detail.

In this contribution, we describe approaches to prepare isotopically carbon-13 and nitrogen-15 enriched intact phage samples in high yield and purity, and we present experimental MAS NMR methods to study the capsid secondary and tertiary structure, and the DNA-capsid interface. Protocols for the capsid structure determination using the Rosetta modeling software are provided. Specific examples are given from studies of the M13 and fd filamentous bacteriophage viruses.

Key words Solid-state NMR, Magic-angle spinning, Bacteriophages, Structure determination, Assignment, DNA-protein interaction, Structural virology, Rosetta

1 Introduction

Filamentous bacteriophages are elongated semi-flexible viruses that infect bacteria. They consist of a circular single-stranded DNA (ssDNA) wrapped by a capsid consisting of thousands of copies of a major coat protein subunit [1–3]. Bacteriophages are ubiquitous and have unique roles in molecular biology, biotechnology, and nanotechnology. Although filamentous bacteriophages have been the focus of many biochemical, biophysical, and biomedical studies for many years, and the number of discovered phages increases constantly, less than a handful of structures exist thus far. Structural

models have been obtained from fiber diffraction (e.g., Pfl [4], fd mutant Y21M [5]), from aligned-static NMR techniques [6], and recently, we have presented a 3D atomic-resolution model for the capsid of intact M13 bacteriophage using Rosetta model building guided by structure restraints obtained from magic-angle spinning (MAS) solid-state NMR experimental data [7]. MAS NMR has also been used to study in detail various filamentous viruses including hydration [8], dynamics [9] and structural transitions [10] in Pfl, comparison of M13 and fd [11, 12], both highly similar phages with a single amino acid replacement (aspartate-12 in the coat protein of fd is replaced with asparagine in M13). An additional study probed the interface between the capsid and the DNA in fd [13].

The advantage in using MAS solid-state NMR techniques to study filamentous viruses resides in their simple and efficient preparation and experimentation. Phage samples can be prepared in a precipitated form with no requirement for alignment or crystallization, the yields are high, experiments can be performed at a large range of temperatures and therefore mimic closely their natural condition, and the phage particles retain their infectivity. Any NMR method however requires isotopic labeling for several reasons: (1) For proteins, ^{13}C ($S = 1/2$) is the only NMR-active carbon isotope and its abundance in nature is 1.1%; (2) ^{15}N , also $S = 1/2$, is the nucleus of choice in biological systems since it can be detected to high resolution. However, ^{15}N has an abundance of 0.4% while ^{14}N with a spin $S = 1$ is broadened significantly by the quadrupolar interaction, and it has a smaller gyromagnetic ratio than ^{15}N . Complete isotopic labeling yields information on all ^{13}C and ^{15}N spins; however, linewidths are broadened by scalar couplings, spectral overlap limits the resolution, and distance measurement estimates are hampered by homonuclear interactions causing “relayed transfers.” In particular, spectral overlap can be severe for helical proteins such as those making the phage capsid. Many of these deficiencies can be resolved by sparse labeling [14, 15].

In this manuscript, approaches and new methodologies for the preparations, isotope labeling, MAS NMR experiments, and structure determination of such complex macromolecular systems are presented. These experimental approaches can be utilized to study different types of filamentous phages, bacteriophages of other forms (spherical, icosahedral, and cylindrical), and other filamentous structures. At least some of these systems have impact on health and disease, and can be utilized for the design of novel nano- and bio-materials.

We initially describe the preparation of isotopically labeled samples, which is a central step for obtaining high-resolution data while maintaining sample integrity; we discuss the purification steps, their biophysical characterization and NMR sample packing. Subsequently, we provide details on the NMR equipment, setup, experimental methodologies, and data analysis. In the end,

protocols for generating the quaternary model of the phage using the program *Rosetta* are described. We note that in the current protocol we do not discuss proton detection techniques, which have been made available recently with the advent of very fast spinning probeheads, up to 110 kHz and more [16]. Those techniques will certainly become extremely useful to study proteins, molecular assemblies, phage systems, and other biomolecules.

2 Materials

2.1 Materials for Sample Preparation and Purification

2.1.1 Strains

1. M13 bacteriophage: Wild-type (wt) M13 contains 6407 nucleotides [17] and is available from the ATCC (ATCC[®] 15669-B1[™]). We used the M13KO7 vector [18] at an initial concentration of 10^{14} particle forming units (pfu)/mL. This vector contains a Kanamycin antibiotic resistance gene inserted in the phage origin of replication. The DNA of this strain has 8669 nucleotides (nt), and ~3600 major coat protein subunits considering the theoretical value of 2.4 nucleotides to subunit [19] (nt/s). The molecular weight of unlabeled M13KO7 is therefore ~21.5 Mega Daltons (MDa).
2. fd bacteriophage: wt fd contains 6408 nucleotides [17] and is available from the ATCC (ATCC[®] 15669-B2[™]). The *fth1* vector [20], used in this study, has two insertion positions in its genome; a segment of transposon Tn10 coding for tetracycline antibiotic resistance and another gene (pVIIIISTSh), located between gene-3 and gene-7, which can incorporate another coat protein for the purpose of mutations and phage display. The vector *fth1* is genetically stable and produces high phage titers. This genome length is 8223 nt (3430 subunits, 20.5 MDa).
3. Both fd and M13 infect *E. coli* strains bearing incompatibility group F (*incF*). Examples are DH5 α F' strain, the tetracycline-resistant XL1 (XL1-tetR) strain, and Hfr D strain (ATCC-15669).
4. There are many additional filamentous phage strains, which are not described here. Examples are Ike [21] and I₂-2 [22], infecting *E. coli* bearing N-pili and I₂-pili, respectively; Pf3 and Pf4 infecting *Pseudomonas aeruginosa* strains O1 and K, respectively [2]; Xf infecting *Xanthomonas oryzae* [23]; and many others [2].

2.1.2 Buffers, Solutions, and Nutrient Media

1. *2 \times YT culture media*: Add 16 g/L Bacto[™] tryptone, 10 g/L g Bacto[™] yeast extract, and 5 g/L NaCl to deionized H₂O (ddH₂O) making up 90% of the total final volume. Adjust the pH to 7.0 with 5 M NaOH and dilute the solution (with ddH₂O) to the final volume. The media is sterilized by autoclaving. For the preparation of rich media with antibiotics, add

the appropriate amount to sterilized media at no higher than 50 °C.

2. *YT-agar plates*: add 15 g/L of agar to 2×YT culture media after pH adjustment and prior to final volume dilution. Pour into sterile petri dishes and allow to cool.
3. *LB culture media*: Prepared as 2×YT, with the following contents; 10 g/L Bacto tryptone, 5 g/L g Bacto yeast extract, and 10 g/L NaCl.
4. *LB-agar plates*: add 15 g/L of agar to LB media after pH adjustment and prior to final volume dilution.
5. *Top agar plates*: Prepared as LB, with the addition of 7 g/L agar.
6. *10 mM Tris-HCl buffer*: Dilute tenfold a 100 mM stock solution, which is prepared as follows: Weigh the corresponding mass for 100 mM Tris (hydroxymethyl) aminomethane (MW = 121.4 g/mol) and dissolve in ddH₂O (in 80% of the final volume). Adjust the pH to 8.0 with 3 M HCl and dilute the solution with ddH₂O to the final volume. Sterilize the stock solution by an autoclave (121 °C) and store at 4 °C.
7. *50 mg/mL kanamycin stock*: Add 0.5 g kanamycin into 5 mL of ddH₂O. Mix thoroughly and add another 5 mL of ddH₂O. Filter-sterilize (0.2 μm) and store at -20 °C. A concentration of 50 μg/mL (10⁻³ dilution) is required for M13 phage growth.
8. *12.5 mg/mL tetracycline stock (50% v/v ethanol/ddH₂O)*: Add 0.125 g tetracycline to 5 mL of absolute ethanol. Mix thoroughly and add 5 mL of ddH₂O. Filter-sterilize (0.2 μm) and store at -20 °C. A concentration of 12.5 μg/mL (10⁻³ dilution) is required for fd phage growth.
9. *M9 Minimal media solution for growing uniformly enriched ¹³C/¹⁵N bacteriophages*:
 Prepare the following stocks
 - Minimal salts (M9 salts) stock × 10: 0.478 M Na₂HPO₄, 0.220 M KH₂PO₄, 0.086 M NaCl. Adjust to pH 7.0 and sterilize by autoclaving. Store at 4 °C. It is best to keep at room temperature 24 h prior to use.
 - 1 M MgSO₄ (sterilize by autoclaving and store at 4 °C).
 - 1 M CaCl₂ (sterilize by autoclaving and store at 4 °C).
 - 0.025 M FeSO₄ (0.2 μm filter sterilized, light sensitive, store as long as the solution is not yellow).
 - 1.48 mM Thiamine-HCl (0.2 μm filter-sterilized, prepare fresh).

Mix the following amounts of the solutions to create 1 L of the M9 media: 47.8 mM Na_2HPO_4 , 22.0 mM KH_2PO_4 , 8.6 mM NaCl (M9 salts), 2 mM MgSO_4 , 0.1 mM CaCl_2 , 0.5 mM FeSO_4 , 0.296 mM Thiamine-HCl. Dilute the solution with sterile ddH₂O to final volume.

For growth, add to the M9 media 4 g/L $^{13}\text{C}_6$ -glucose and 0.5–1.5 g/L $^{15}\text{NH}_4\text{Cl}$.

10. Minimal Media Solution for Growing 1,3- ^{13}C (1,3-gly) and 2- ^{13}C (2-gly) Bacteriophages:

The minimal nutrient medium contains the following:

- M9 solution at pH = 6 (*see Note 1*).
- ME vitamins×100 solution (we use Sigma-B6891).
- Metal trace [24] solution (sterilized) containing: 0.2 mM FeSO_4 , 0.4 mM CaCl_2 , 0.6 mM MnCl_2 , 30 μM CoCl_2 , 20 μM ZnSO_4 , 20 μM CuCl_2 , 3 μM H_3BO_3 , 2 μM $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, 0.1 mM EDTA.
- 2 mM MgSO_4 (autoclaved).
- 12.5 $\mu\text{g}/\text{mL}$ tetracycline (sterilized). This step is optional and applied if antibiotic resistance exists, as for fth1 strain of fd.
- For the preparation of 2-gly-M13: 1 g/L $^{15}\text{NH}_4\text{Cl}$, 2 g/L $\text{NaH}^{13}\text{CO}_3$, 2 g/L [2- ^{13}C]-glycerol.
- For the preparation of 1,3-gly-M13: 1 g/L $^{15}\text{NH}_4\text{Cl}$, 2 g/L $\text{NaH}^{12}\text{CO}_3$, 2 g/L [1,3- ^{13}C]-glycerol.

2.2 NMR Materials and Equipment

1. An NMR spectrometer consisting of a three-channel console and a high-field magnet. Our lab has a Bruker Avance-III wide-bore solid-state NMR spectrometer operating at a magnetic field of 14.1 T. Larmor frequencies are 600.0 MHz for ^1H , 150.9 MHz for ^{13}C , 60.8 MHz for ^{15}N , and 242.9 MHz for ^{31}P .
2. A triple-resonance probe operating at HCN mode for 2D and 3D experiments involving ^{13}C and ^{15}N . To date, available rotor diameters range from 4 to 0.6 mm. They provide maximal spinning speeds of 15 kHz (4 mm), 24 kHz (3.2 mm), 35 kHz (2.5 mm), 67 kHz (1.3 mm), >110 kHz (0.7, 0.6 mm). In regular solenoid coils, heating can be induced by radio-frequency pulses making them unsuitable for high-salinity samples. Efree (Bruker) and scroll-coil [25] technologies minimize the heating effect by directing the electric field away from the sample. Such coils are available for 3.2 and 4 mm probes.
3. A triple-resonance probe operating at HPN/HPC mode for protein-DNA interaction studies.

4. $\text{Pb}(\text{NO}_3)_2$ packed in a rotor for sample temperature measurements (*see* **Note 2**).
5. Adamantane/KBr powder packed in a rotor for setup.
6. Glycine powder (labeled or natural abundance) packed in a rotor for setup.
7. Uniformly (^{13}C , ^{15}N) labeled phage samples (*see* Subheading **3.1.1**).
8. Sparsely labeled phage samples (*see* Subheading **3.1.2**).

3 Methods

3.1 Sample Preparation

3.1.1 Preparation of Fully (^{13}C , ^{15}N) Labeled M13 and fd Samples

1. Prepare fresh 2×YT (rich media) agar plates for the growth of the bacterial host (*E. coli* DH5αF' strain or XL1-tetR strain or any other suitable strain) in a sterile environment.
2. Spread the bacteria and incubate the plates overnight (37 °C), and obtain single colonies.
3. Pick out a single colony and incubate with shaking (37 °C, 240 rpm) in a sterile tube containing 5 mL 2×YT medium for ~16 h.
4. Transfer 200 μL of the dense bacterial solution to a 5 mL rich medium tube and incubate with shaking for ~4 h, allowing the cells to reach the log phase of the growth curve (Optical Density, O.D. ~ 0.6). Growth is monitored by light-scattering measurements (wavelength of 600 nm).
5. At log phase, reduce shaking speed to 60 rpm for 30 min. This is done in order to allow optimal growth of F-pili and thereby enhance infectivity.
6. Add 50 μL of the phage stock to the 5 mL bacterial solution and incubate for another 30 min (37 °C, 60 rpm). For effective infection, the concentration of the phage stock in pfu/cfu, plaque forming units to colony forming units, should be 100/1. At this point, the cell density is ~10⁸ cfu/mL.
7. Transfer an aliquot of the infected culture (at a ratio of 1/100) to a solution of 500 mL minimal nutrient medium at 37 °C, 240 rpm (*see* **Note 3**).
8. After 1 h, add kanamycin (for M13) or tetracycline (for fd) antibiotics to the mixture. This allows only infected cells to further replicate.
9. After 24 h of incubation and shaking, centrifuge the bacteria at 8000 rpm (9800 × *g*) for 45 min (fixed-angle rotor, we use F14S-250X6 FiberLite Piramoon Technologies Inc.).
10. Decant the supernatant phage solution.

11. Precipitate the phage by making the solution 5% (w/v) polyethylene glycol (PEG) 8000 and 0.5 M NaCl.
12. Collect the PEG-precipitated phages at 8000 rpm (9800 $\times g$) for 45 min at 4 °C and resuspend in 10 mM Tris buffer, pH 8.0.
13. Centrifuge the phage solution once again at 14000 rpm (22800 $\times g$) for 2–3 h to discard leftover bacterial cells. Here we used 50 ml tubes in a Fiberlight F15-8 \times 50cy fixed angle rotor.
14. The total yield of fully labeled M13 and fd phages after this step of purification should be ~50 mg for 1 L of culture but can vary between different preparations.

3.1.2 Preparation of Sparsely Labeled M13 Using Partially Enriched Glycerol

Labeling proteins with 1,3-¹³C₂-glycerol or 2-¹³C₁-glycerol was first presented by LeMaster et al. [26] and produces the “checkerboard labeling” pattern (*see Note 4*).

1. Prepare fresh tetracycline-resistant (for the XL1-tetR strain) YT-agar plates and grow *E. coli* colonies overnight.
2. Follow the procedure described in Subheading 3.1.1 from **step 3** onward using the minimal nutrient medium from Subheading 2.1.2, **item 10**. The pH is adjusted to 6 (*see Note 1*).
3. Expect yields of ~20–30 mg of glycerol-based M13 growth for 1 L of culture before purification by ultracentrifugation.

3.1.3 Preparation of Aromatic-Unlabeled Samples

This protocol is used for observing DNA resonances in a protein-DNA complex. The spectral distinction results from the elimination of the aromatic signals of the amino acids tryptophan, phenylalanine and tyrosine.

1. Follow the protocol from Subheading 3.1.1 for growth of fully enriched samples using the minimal nutrient medium from Subheading 2.1.2, **item 9**, and add 0.2 g/L of natural abundance Phe, Tyr and Trp.
2. For YFW^{unlab}-¹³C/¹⁵N-fd, yields of ~30 mg for 1 L of culture can be obtained.

3.1.4 Purification of Filamentous Phage Using CsCl Gradient Ultracentrifugation

This technique separates components based on their buoyancy. It requires an ultracentrifuge, a swinging bucket rotor and tubes suitable for high-speed centrifugation. The protocol below is described for Ultra-Clear™ 13.2 mL Beckman tubes and a SW-41 Ti swinging bucket rotor.

1. Weigh 4.83 g of CsCl to a 50 mL tube and calibrate the scale to zero.
2. Add dropwise ~12 mL of a ~1 mg/mL clarified phage solution to a tube to a final mass of 10.75 g. This gives a solution density of 1.29–1.31 g/cm³, which positions the phage band in the middle of the tube after centrifugation, *see step 6* below (Fig. 1).

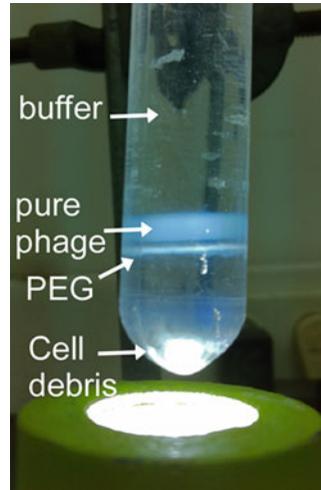


Fig. 1 Sample tube after CsCl gradient ultracentrifugation

3. Mix the phage-CsCl solution thoroughly and zero the scales again.
4. In order to assess the density of the final solution, take 1 mL with a pipette tip, while keeping the tube on the scale. The scale should show values ranging from negative 1.28 g to negative 1.31 g, which indicates the final density of the solution.
5. Transfer the solution to Ultra-Clear™ 13.2 mL Beckman tubes and insert well-balanced tubes to their positions in a SW-41 Ti swinging bucket rotor (or similar) (*see Note 5*).
6. Centrifuge to equilibrium at 37,000 rpm ($234700 \times g$), 4 °C, for 48 h (e.g., Optima XE 100K Ultracentrifuge).
7. Remove the tubes and assemble on the top of a bottom light to view clearly the viscous phage band (Fig. 1).
8. Remove the top clear phase (buffer) carefully with a pipette.
9. Remove the viscous central band of the phage by a pipette (*see Note 6*).
10. Dilute the recovered phage solution to ~1 mg/mL.
11. The solution can be stored at this stage at 4 °C or cleared from ~2 M Cs⁺ ions (**step 12**).
12. Precipitate the phage solution with 5% w/v PEG8000, decant the supernatant, and resuspend in Tris buffer (10 mM).

3.1.5 M13 and fd Virus Characterization Using UV Spectroscopy Measurements

Pure phages exhibit a typical UV absorption spectrum having a broad plateau at 260–280 nm, with a shallow maximum at 269 nm and a minimum at 245 nm resulting from both proteins and DNA [19, 27]. Phage purity is determined by the ratio between maximum and minimum absorbance, and by the ratio between the baseline and maximum absorbance. The baseline appears due to scattering effects [28].

1. Transfer a virus solution and a buffer solution into two 1 cm path quartz cells.
2. Measure the buffer solution to obtain the blank (220–380 nm).
3. Measure the phage solution (220–380 nm) and obtain the difference spectrum.
4. Calculate the maximum absorbance (269 nm), the minimum absorbance (245 nm), and the baseline (350 nm). The maximum may shift with phage type and purity. Avoid measurements beyond the linear regime of the UV spectrophotometer.
5. Pure phage is obtained if $OD_{269}/OD_{245} \sim 1.37$ and $OD_{350}/OD_{269} \sim 0.02$.
6. The concentration of the phage is given by the Beer-Lamber law $A = \epsilon bC$ where for M13 and fd, the extinction coefficient is [19] $\epsilon = 3.84 \text{ cm}^2/\text{mg}$. Note that a concentration of 1 mg/mL may be too high and the solution needs to be diluted prior to measurement.
7. Figure 2 shows a typical UV spectrum of M13 bacteriophage after ultracentrifugation purification.

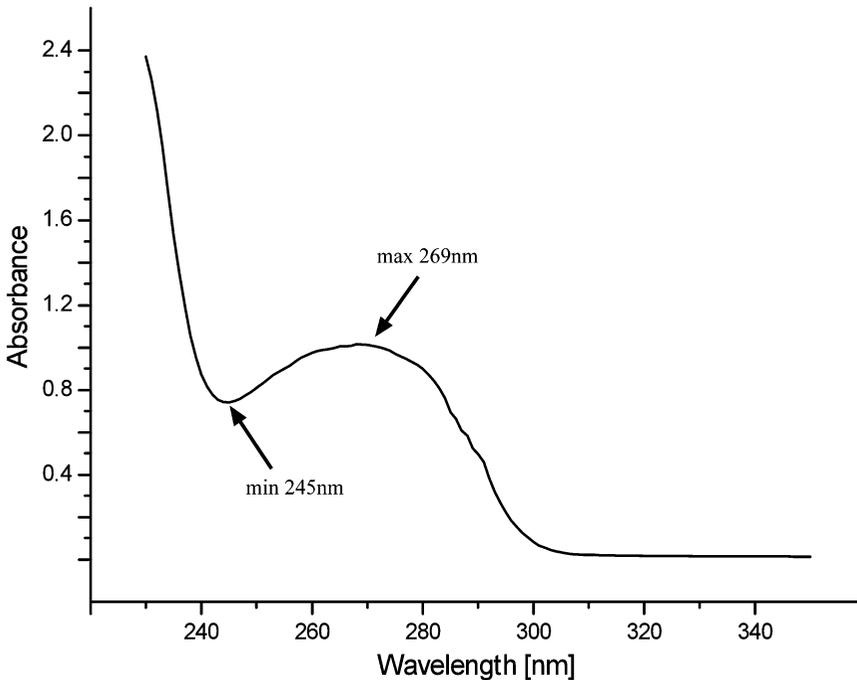


Fig. 2 Spectrum of a tenfold diluted M13 phage after ultracentrifugation. The virus has a UV absorbance spectrum with a maximum at 269 nm and a minimum at 245 nm. The purity is assessed by two ratios: max to min absorbance and baseline to max absorbance. Here, the 269/245 ratio is 1.37 and the 350/269 ratio is 0.02, all indicating that the phages are viable and pure. The absorbance of the pure sample in this case corresponds to 16 mg

3.1.6 Quantifying Phage Infectivity with a Plaque Assay

1. Grow a 5 mL culture of the *E. coli* strain TG1 to log phase.
2. Using autoclaved tips, transfer aliquots of 90 μL log-phase bacteria to a series of wells in a sterile 96-well cell culture plate (e.g., Greiner bio-one).
3. Transfer 10 μL of purified phage to the first well and mix gently using a pipette (*see* **Note 7**).
4. Replace the tip and transfer 10 μL from the first well to the second well. Mix gently and discard the tip.
5. Replace the tip and transfer 10 μL from the second well to the next well. Mix gently and discard the tip.
6. Repeat the process described in **step 5** where each time the new dilution is used to dilute the following well. Dilutions should be performed until $\sim 10^{-14}$ or until the last well contains < 10 phage particles. Dilutions of 100-fold can also be done (by transferring 1 μL at each step) but with care.
7. Always retain one well with no phage particles (only cells) as control.
8. Incubate the plate at 37 °C for 1 h, allowing the phage to infect the cells.
9. Take 10 μL from each well and drop into different slices of LB-agar plates containing kanamycin (for M13) or tetracycline (for fd).
10. Incubate the plates at 37 °C overnight. Only infected cells will form colonies.
11. In addition, it is advised to incubate overnight an LB-agar plate containing kanamycin without *E. coli* for control.
12. At the high dilutions, it is possible to count the number of isolated plaques and therefore to determine pfu/mL in the original phage solution.
13. A typical titration plate is shown on Fig. 3a.

3.1.7 Top-Agar Assay for Wild-Type Phage with No Antibiotic Resistance

Wild-type phage particles do not have antibiotic resistance. In order to quantify the amount of pfu/mL and assess their infectivity, a top agar assay is performed, which is equivalent to the plaque assay described above.

1. Prepare LB-agar plates and let them cool down.
2. Prepare a series of phage dilutions up to $\sim 10^{14}$ in 1 mL Eppendorf tubes.
3. Grow target host cells to their log phase (OD 0.6–0.8).
4. Prepare a top agar solution (Subheading 2.1.2, **item 5**) and let it cool to ~ 50 °C (we found that the procedure works best using a fresh top agar solution but this is not mandatory).

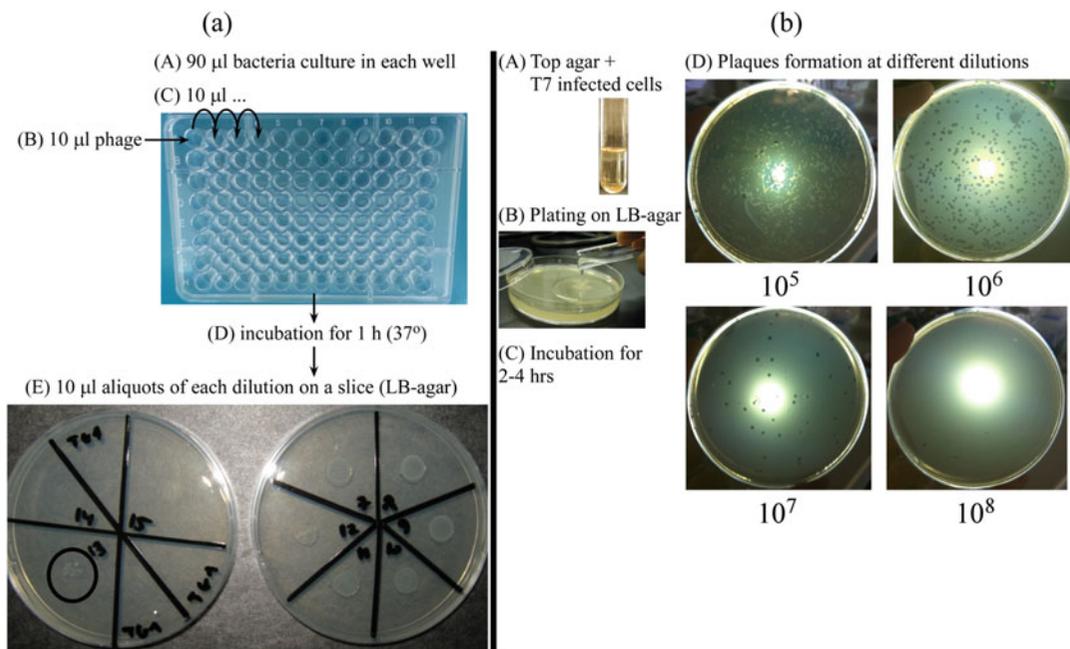


Fig. 3 (a) fd phage infectivity assay. The assaying steps for fd are denoted by A–E. 10 μL of infected phage solutions are dropped onto slices in the LB-agar plate. Each slice of infected cells represents a dilution count from 1 to 12. In the most diluted slice (#12) the plaques can be counted and subsequently the number of phage particles per milliliter (pfu/mL) in the initial phage sample can be calculated. (b) T7 phage top-agar infectivity assay. The assaying steps for T7 are denoted by A–D. In D, each plate represents a dilution count, which is indicated on the *bottom*. Note: T7 is not a filamentous bacteriophage. It is a tailed bacteriophage with a 60 nm icosahedral head that belongs to the family Podoviridae. Yet, the infectivity assay procedure is similar to that of filamentous viruses

5. Pre-warm a dry bath incubator with 15 mL glass tubes to a temperature of 45 $^{\circ}\text{C}$. Transfer 3 mL of the top agar solution and 250 μL of the host cells to the glass tubes. Leave one glass tube without any host cells. The number of glass tubes should equal the number of phage dilutions and two for additional controls.
6. Add 100 μL of each phage dilution to a different glass tube.
7. Remove each glass tube out of the incubator, vortex gently and quickly, and overlay homogeneously on one of the LB-agar plates.
8. Overlay one plate with no phage particles (only top-agar and cells) as a first control.
9. Overlay one plate with no cells and no phage particles (only top-agar) as a second control.
10. Incubate the agar plates at 37 $^{\circ}\text{C}$. The incubation time depends on the phage and host strains. For filamentous bacteriophages we usually incubate overnight.
11. Observe and count the turbid plaques. Calculate the pfu/mL accordingly (*see Fig. 3b*).

3.1.8 Packing Purified PEG-Precipitated Phage Samples into MAS Rotors

1. Dissolve 5% w/v PEG8000 in a 1 mg/mL phage solution (50 mL tubes).
2. Add 5 mM MgCl₂. Instantaneous precipitation can be recognized by the appearance of a white cloudiness.
3. Centrifuge the phage solution for 15 min at 5000 rpm (BIOShield swing rotor fitted with 15 or 50 ml tubes, 5900 × *g*) at 4 °C. Discard most of the supernatant, leaving a layer of about 200 µL solution above the pellet.
4. Prepare a sufficient amount of flame-sealed 200 µL tips.
5. Gently mix the phage pellet with the remaining solution. Transfer the aliquots to several flame-sealed 200 µL tips.
6. Mount the tips on 1.5 mL Eppendorf tubes (remove caps), cover with parafilm, and centrifuge at 14,000 rpm (21900 × *g*) and 4 °C for 15 min (e.g., Sigma 1-15PK centrifuge with a swing-bucket rotor 11124-H or any other centrifuge suitable for 1.5 mL tubes). A swing-bucket at this point is recommended but not absolutely essential.
7. Remove the supernatants from each tip, and combine pairs of tips by centrifugation of one into the other for about 1 min.
8. Repeat **steps 6–7** until a single tip contains all the phage. Leave a small layer of the precipitating solution above.
9. Centrifuge the final precipitate for several hours, until the level of condensation does not change in the tip (the time required for this procedure may change from sample to sample).
10. Mount the final tip on a ZrO₂ (4 or 3.2 mm) MAS rotor and transfer the phage precipitate by centrifugation (~1 min). This can be done with a home-made device as shown in Fig. 4 or any other device.
11. Centrifuge the rotor containing the sample under the same conditions (14,000 rpm, 4 °C, 15 min, using a swing-bucket), and seal with a spacer according to the vendor's instruction. (Bruker HRMAS rotors have a spacer, a spacer-screw, and a cap. The small hole in the spacer allows a small fraction of the liquid to transfer through showing that the spacer is in place).
12. The approximate mass of the phage in a 4 mm rotor is 8–10 mg occupying ~30 µL (~250–300 mg/mL).

3.2 NMR Experiments

3.2.1 General Setup

1. *Matching and tuning.* After the probe has been installed, and for each sample independently, tune each channel to the required frequency and match the impedance to 50 Ω.
2. *Set up the magic angle using KBr:* Insert a rotor containing KBr powder, set the spinning speed (ν_R) to 5 kHz, match and tune to ⁷⁹Br frequency, and conduct a repetitive single pulse experiment (calibration of pulses is detailed in Subheading 3.2.2) with the

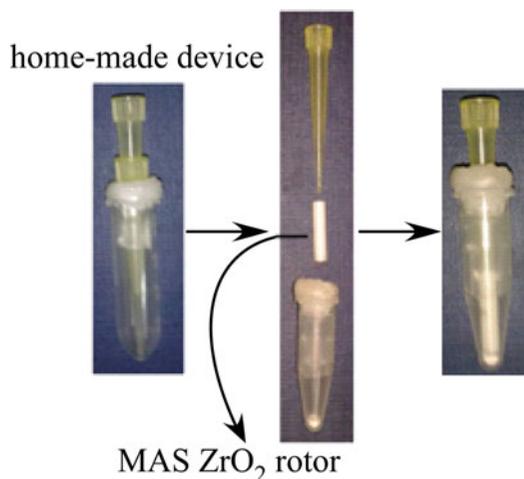


Fig. 4 NMR sample packing with a home-made device. First, the hydrated phage precipitate is transferred from a yellow flame-sealed tip, which was cut in the bottom, to another tip through a series of centrifugations at 14,000 rpm, as shown in the *left photo*. The tip containing the sample collected from all tips is placed above the MAS ZrO_2 rotor (*mid and right photos*) and the phage is transferred to the rotor by centrifugation (the phage precipitate inside the yellow tip is not shown here)

carrier frequency on the central line while observing the free induction decay (FID). Gently adjust the magic angle using the MAS knob until the rotational echoes of the FID extend to their maximal time ($\sim 8\text{--}9$ ms on a 600 MHz spectrometer at $\nu_R = 5$ kHz). Stop the acquisition and collect a single pulse spectrum. Measure the intensity ratio between the first spinning side band (ssb) and the central peak in the Fourier-transformed (FT) spectrum. A ratio of 13–14% and a linewidth (FWHH) of ~ 110 Hz indicate an optimal magic angle of $\text{atan}(\sqrt{2}) = 54.736^\circ$.

3. *Setting the ^1H carrier frequency using powder Adamantane.* Match and tune ^1H , set $\nu_R = 5$ kHz, collect the ^1H spectrum using a short excitation pulse (FID ~ 100 ms), set the carrier frequency on resonance.
4. *High-resolution shimming using Adamantane.* Match and tune ^{13}C , set $\nu_R = 5$ kHz, collect a repetitive single pulse ^{13}C spectrum (*see* Subheading 3.2.2, **step 1** for power calibration) with low power ($\sim 30\text{--}40$ kHz) ^1H decoupling and an acquisition time of 500 ms (*see* **Note 8**). Change the current in the shimming coils until a linewidth of 2–3 Hz or less for the CH_2 line is obtained.
5. *Referencing ^{13}C ppm scale using powder Adamantane.* Conduct a single pulse experiment with decoupling as in **step 4** above

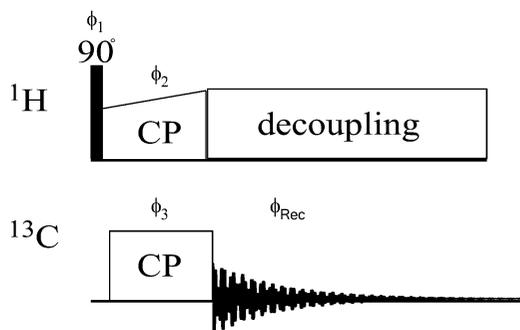


Fig. 5 Cross-polarization (CP) pulse sequence [30]. The phases of the proton CP pulse and excitation pulse are perpendicular to each other to spin-lock the magnetization. The complete phase cycle is as follows (in degrees): $\phi_1 = 0, 180$; $\phi_2 = 90$; $\phi_3 = 270, 270, 180, 180$; $\phi_{\text{rec}} = 180, 0, 90, 270$

(reduce acquisition time to ~ 200 ms to avoid heating). Calibrate the downfield ^{13}C resonance (CH_2 line) to 40.48 ppm [29].

6. *High-power tuning.* Perform a CPMAS experiment as shown in Fig. 5 on the phage sample according to Subheading 3.2.2, **step 4**. Observe on a scope the forward (F) and reflected (R) voltages (this can be done by putting a directional coupler between the preamplifier and the probe). Adjust the matching and tuning knobs until the ratio F/R is equal to or larger than 10 (*see Note 9*).
7. *Referencing ^{15}N ppm scale using powder $^{15}\text{NH}_4\text{Cl}$.* Perform a single pulse ^1H -decoupled ^{15}N experiment and adjust the ^{15}N resonance to 39.27 ppm [31].

3.2.2 Basic Calibration of ^1H and ^{13}C Power Levels, Polarization Transfer, and Decoupling

1. *^1H and ^{13}C power calibration (Adamantane choice, single pulse):* Set up a single pulse experiment as described in Subheading 3.2.1, **step 3**. Set the length of the pulse to $1 \mu\text{s}$ with a power level allowed by your hardware. Fourier transform and phase the spectrum. Set the pulse length to $5 \mu\text{s}$. Run consecutive experiments with varying values of the power level from low to high power. When the signal becomes zero the power level is 100 kHz. All other values can be calculated from the linearity of the transmitter (*see also Shi and Ladizhansky [32]*). For ^{13}C repeat the same procedure but using a $10 \mu\text{s}$ pulse to obtain a power level of 50 kHz. If the transmitter is not linear, a complete calibration curve needs to be obtained.
2. *^1H power calibration (glycine or phage sample, via CPMAS):* The power can also be set from a CPMAS experiment (Fig. 5), especially in a phage sample (and if the probe is not of the Efree type) that may be affected by RF heating thereby altering the ^1H power levels obtained from powder samples. Set the ^1H

power of the excitation pulse to 100 kHz as described in **step 1**. Set ^{13}C carrier frequency on resonance with the Glycine C α line (or phage aliphatic lines). Set the ^{13}C power level during CP to 50 kHz (**step 1**). Set the ^1H power level during CP to be the ^{13}C power level plus twice the spinning speed (Hartman Hahn matching condition under MAS [33]) and the CP length to 1 ms. Check for a signal, then set the length of the ^1H pulse to 5 μs , and vary the power of the ^1H pulse until the signal disappears. The power level is equivalent to 100 kHz.

3. ^{13}C power calibration (glycine or phage sample, via CPMAS): Add to the CPMAS experiment (Fig. 5) an additional $\pi/2$ pulse with a phase of 90° with respect to the carbon CP pulse. Set the ^1H flip angle to $\pi/2$ (2.5 μs if set to 100 kHz in **step 2**). Set the power of the ^{13}C additional pulse to zero and its length to 0.1 μs . Fourier transform and phase the resulting spectrum. Set the length of the ^{13}C pulse to 5 μs and vary the power until the signal disappears (or attains a minimum). The ^{13}C power is now to 50 kHz.
4. ^1H - ^{13}C ramped-CPMAS optimization (glycine or phage sample choice, via CPMAS): Change the pulse sequence to CPMAS (Fig. 5), with a 10% linear ramp on the ^1H channel during CP. Set the ^{13}C power level to 50 kHz and set the carrier frequency to the middle of the ^{13}C spectrum. Vary the ^1H power level between 50 and 100 kHz to hit the match condition (the difference between the power levels is once or twice the spinning speed ν_{R}). Search for a maximum intensity of a selected signal or for the entire spectrum. Repeat the optimization for the length of the CP pulse up to a few milliseconds.
5. ^1H - ^{15}N ramped-CPMAS optimization ($^{15}\text{NH}_4\text{Cl}$ or phage sample choice, via CPMAS): Repeat the procedure of **step 4** with the carrier frequency set to the observe ^{15}N channel.
6. ^1H Decoupling: There are many choices for decoupling sequences, examples of which are TPPM [34], SPINAL [35], XiX [36], and others [37]. All of them require the optimization of one, or all, of the pulse lengths, pulse phases, and pulse power levels. The swf-TPPM decoupling sequence [38] is a more robust variant of TPPM. Set the power level to the maximum allowed by the probe and vary the basic decoupling pulse length between 1 and 10 μs . Choose the value corresponding to the maximum intensity. The flip angle of the pulse should be $\sim 180 \pm 20^\circ$. The phases of alternating pulses should be switched by $\sim 20^\circ$.

3.2.3 Setting Up a 2D ^{13}C - ^{13}C DARR Correlation Experiment

1. The experiment is suitable for spinning speeds < 20 kHz, i.e., for 3.2 or 4 mm probes.

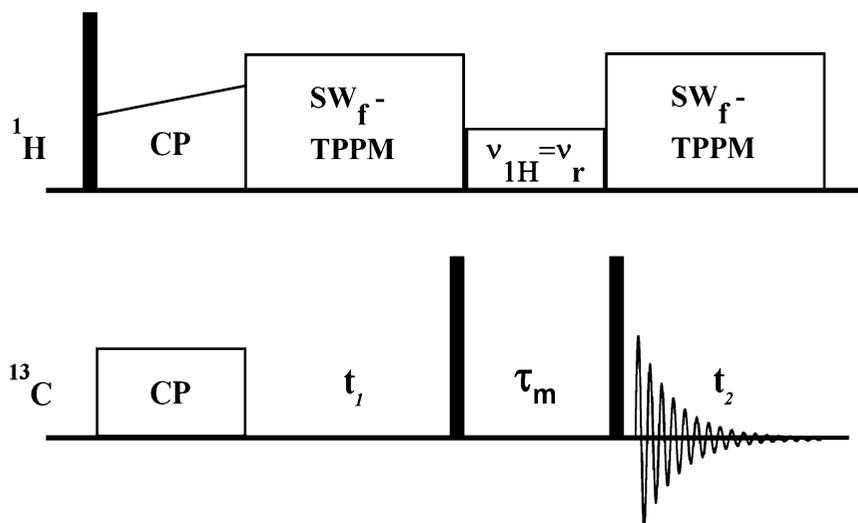


Fig. 6 The dipolar-assisted rotational resonance (DARR) experiment [39]. Setup of the $\nu_{1H} = \nu_R$ condition is described in **step 3**, the mixing time τ_m is discussed in **step 4**. CP is discussed in Subheading 3.2.2, **step 4**. Decoupling is described in Subheading 3.2.2, **step 6**. *Filled rectangles* correspond to 90° pulses. Details of the phase cycle can be found elsewhere [32, 40]. The CORD (COmbined $R2_n^y$ -Driven) experiment [41], described in Subheading 3.2.4, is performed by replacing the ^1H mixing pulse with blocks of R-symmetry pulses. It is efficient for spinning speeds >20 kHz (but also for lower spinning speeds) and effective for aromatic signal recoupling

2. Set the spinning frequency and avoid overlap of spinning sidebands with the main spectrum (usually the equivalent of ~ 80 ppm).
3. *DARR condition*: Set up a CP experiment with cw decoupling (constant irradiation). Vary the ^1H decoupling power between the equivalent of $4\nu_R$ and zero until the CH_2 signal has a maximum broadening and a minimum intensity. At this point the rotary resonance recoupling (R^3) condition $\nu_{1H} = \nu_R$ is satisfied and should be used for the ^1H irradiation during the mixing time of the DARR sequence shown in Fig. 6.
4. *DARR mixing time*: To a good approximation mixing times of ~ 1 – 20 ms correspond to correlations between carbons separated by distances of 1 – 2 Å, 20 – 100 ms correspond to distances of 2 – 5 Å, and longer mixing times permit long-ranged transfers (2 – 8 Å). The exact correlation between distance and the mixing time is sample dependent.

3.2.4 Setting Up 2D CORD_{xy4} Experiment

1. Repeat the setup for the DARR experiment from Subheading 3.2.3.
2. Find the power levels for $\nu_{1H} = \nu_r$ (in kHz) for $R2_1^y$ and $\nu_{1H} = \nu_r/2$ for $R2_2^y$.

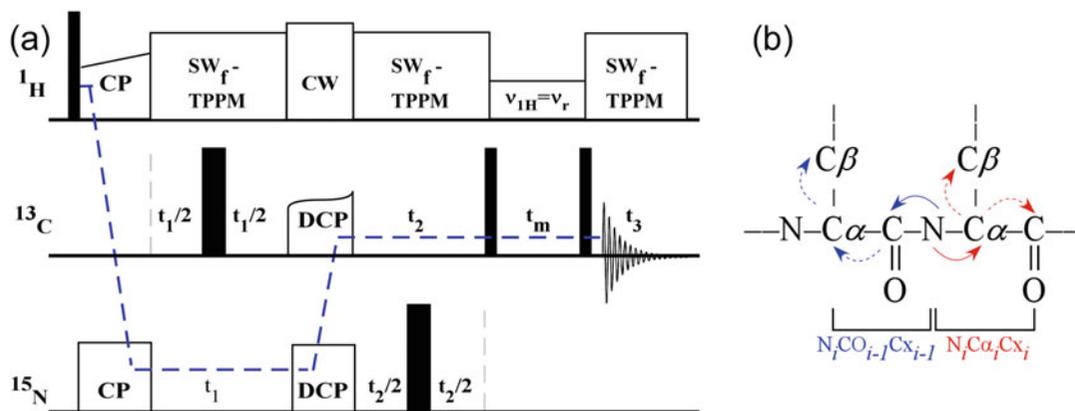


Fig. 7 (a) A basic pulse sequence for 3D NCACX or NCOCX and (b) the corresponding forward and reverse magnetization transfer pathways. Phases and setup are detailed in Shi and Ladizhansky [32]. The DCP (double cross polarization [42]) has a tangential ramp that can be optimized (slope and maximum to minimum difference)

3. Program the following four blocks for ^1H irradiation: (a) $R2_1^1$ and $R2_2^2$ —length one rotor period (T_R), power level $\nu_{1\text{H}} = \nu_R$; (b) $R2_1^1$ and $R2_2^2$ —length $2T_R$, power level $\nu_{1\text{H}} = \nu_R/2$. The total length for one basic cycle is $6T_R$.
4. Repeat the four-block cycle of **step 3** three more times, each consecutive $6T_R$ block phase shifted by 90° . **Steps 3** and **4** comprise the basic CORDxy4 unit.
5. Choose a desired total mixing time (multiple of $24T_R$) and perform the experiment.

3.2.5 Setting Up 3D NCOCX and NCACX Experiments

1. Figure 7 outlines the basic scheme for performing 3D experiments and the corresponding polarization transfer pathways. For a detailed description for the setup the reader is referred to the protocol described in the article by Shi and Ladizhansky [32].

3.2.6 Setting Up 2D PHHC Experiment: Observing Protein-DNA Contacts

1. Set up the probe configuration to H-P-C mode (this will normally be a regular, not an Efree, probe).
2. Tune and match the X channel to ^{31}P and the Y channel to ^{13}C .
3. Insert a rotor with 85% (14.6 M) phosphoric acid. Retune ^{31}P . Collect a single pulse ^{31}P spectrum, adjust the power, and set the observed signal to zero ppm—this is the reference chemical shift for ^{31}P .
4. The following steps can be applied first to a model ^{31}P sample such as phosphoserine or directly to the phage sample.
5. Insert a new sample and retune/rematch all three channels. Set $\nu_R = 12$ kHz (or 80 ppm) to avoid ^{13}C carbonyl overlap with aromatic and DNA signals. For the phage sample, perform high-power tuning.

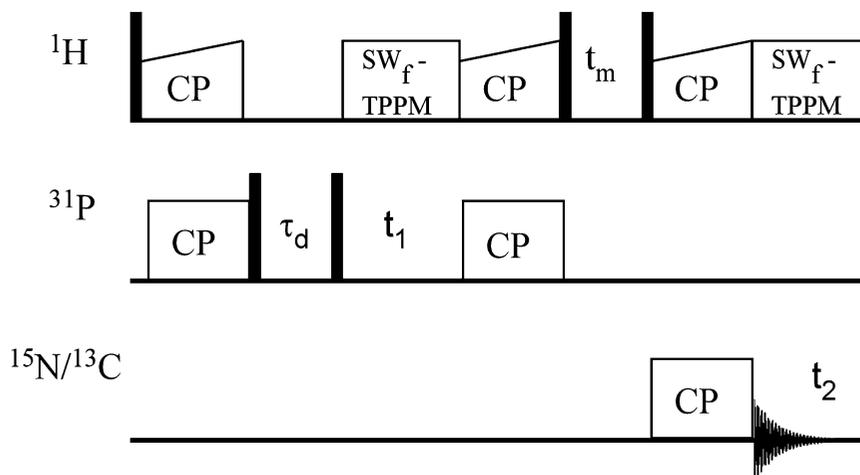


Fig. 8 PHHC 2D correlation experiment. *Filled rectangles* are 90° pulses. Details of the phase cycle can be found elsewhere [13].

6. Calibrate ^1H and ^{13}C power levels (**steps 1–4** in Subheading 3.2.2). Note that ^{13}C frequency is on the low (Y) channel now.
7. Calibrate ^{31}P power in a similar way (preliminary calibration can be performed on H_3PO_4 , or on phosphoserine, similarly to Adamantane, Subheading 3.2.2, **step 1**).
8. Optimize a ^1H - ^{31}P CPMAS experiment as in Subheading 3.2.2, **step 4**.
9. Optimize a ^1H - ^{13}C CPMAS experiment as in Subheading 3.2.2, **step 4**. Use a short CP contact time to selectively transfer the magnetization only to the attached carbons (contact times of 100–300 μs).
10. Upload the PHHC pulse sequence [13, 43] shown in Fig. 8.
11. Set all the pulse powers. For the second CP step ($^{31}\text{P} \rightarrow ^1\text{H}$) use the same values obtained in **step 8**.
12. Set the z-filter time, τ_d (*see* Fig. 8), to be equal to one or two rotor periods (to dephase transverse ^1H magnetization). This parameter can also be optimized up to 1–3 ms to obtain better lineshapes and to reduce artifacts.
13. Conduct 1D versions of the PHHC experiment using different ^1H - ^1H mixing times (~ 100 – 500 μs).
14. Conduct a 2D version of the experiment with the desired or several mixing times.

3.2.7 Signal Processing

1. Signal processing can be performed using the free software nmrPipe (original version here: <https://spin.niddk.nih.gov/NMRPipe/install/legacy.html> and an updated version here: <https://www.ibbr.umd.edu/nmrpipe/install.html>). The

software website contains all instructions for installation and for writing scripts.

2. Convert the spectrometer files to nmrPipe-compatible files (or to any other software of choice).
3. For Fourier transform processing, set the following parameters for each dimension: apodization function, zero-fill, linear prediction.
4. Fourier transform the spectrum. Make sure you use the proper frequency discrimination methods from the acquisition in the indirect dimension of 2D and 3D experiments—TPPI, STATES, STATES-TPPI.
5. Phase correct the spectrum and adjust the baseline.
6. Ensure proper referencing in all dimensions.

3.3 Data Analysis

3.3.1 Major Coat

Protein Site Specific

Resonance Assignment

Resonance assignment is the process in which every chemical shift (resonance) is matched to a specific atom. The capsid of fd and M13 (and also other phages) comprises about 85% of the total virion mass; hence, the NMR spectra are dominated by the major coat protein resonances. Yet, DNA signals can be readily observed since they mostly resonate at different frequencies (*see* Subheading 3.3.2). Chemical shifts of proteins are reported in the Biological Magnetic Resonance Bank, BMRB [44], and $^{13}\text{C}/^{15}\text{N}$ assignment procedures for MAS NMR can be viewed in the corresponding articles. We provide the main guidelines here. Commonly used free software suitable for resonance assignment and analysis are CCPN (<http://www.ccpn.ac.uk/>), Sparky (<https://www.cgl.ucsf.edu/home/sparky/>), CARA (<http://www.cara.nmr.ch/doku.php>).

1. Collect and process data from 2D ^{13}C - ^{13}C chemical shift correlation experiments (Subheadings 3.2.3, 3.2.4, and 3.2.7).
2. Collect and process data from 3D NCACX and NCOCX experiments (Subheadings 3.2.5 and 3.2.7), and optionally from CONCA experiments [32].
3. Optionally complement by additional experiments such as RFDR [45–47] and J-based INADEQUATE [48, 49]. These experiments are not described here but have been used for example to complement the assignment of a small portion of the M13 resonances [11].
4. In 2D spectra, identify the spin systems for amino acids using well-isolated chemical shifts such as the $\text{C}\alpha/\text{C}\beta$ signals of Ala (~55 ppm/~20 ppm), Ser (60/63), Asp (57/41), and Thr (65/69), the side-chain patterns of Ile (CH_3 groups at ~13/17) and Val (CH_3 at ~22 ppm), and the backbone carbons of Gly ($\text{C}\alpha$ at ~46 ppm). The spectral region shown in Fig. 9 is a portion of the complete Fourier-transformed 2D DARR

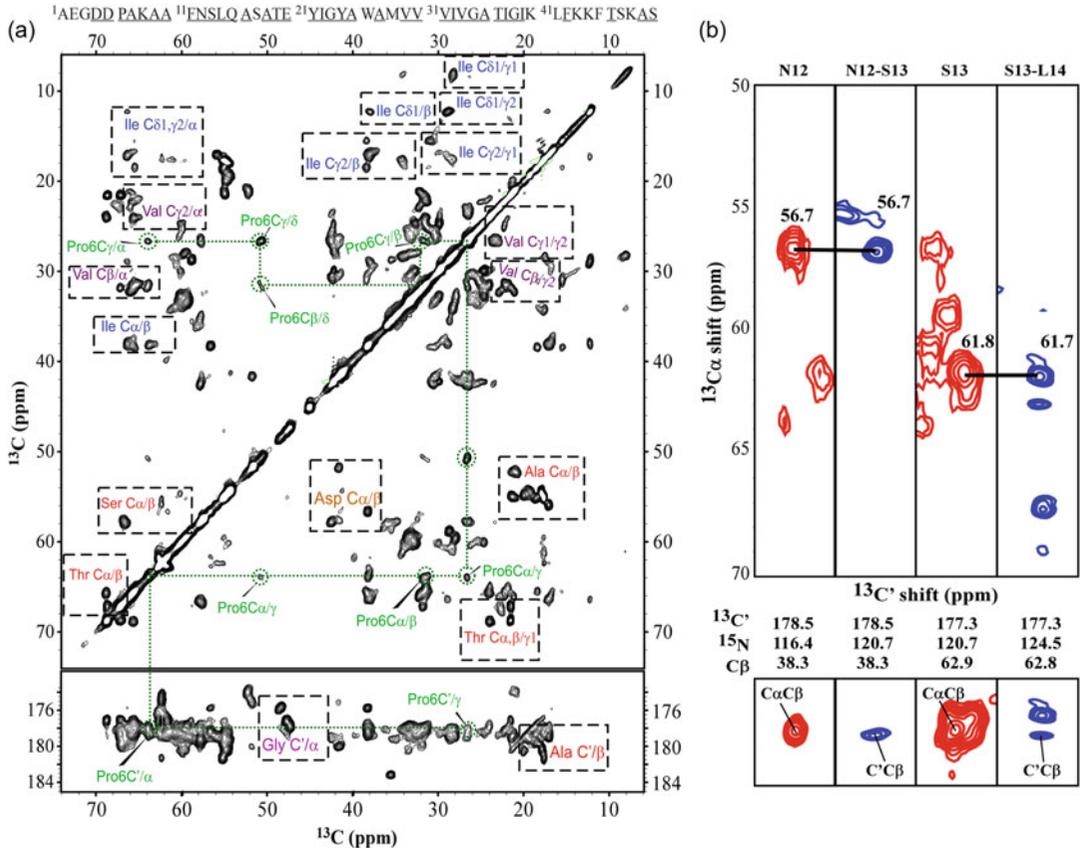


Fig. 9 (a) Aliphatic (*top*) and carbonyl (*bottom*) regions of the complete 2D ^{13}C - ^{13}C DARR15 spectrum of M13 virion with the amino acid sequence of the coat protein shown on *top*. The complete amino acid pattern of Pro6, which appears once in the sequence, is marked with a *dotted green line* connecting all intra-residue peaks. The spectrum was recorded using a spinning rate of 13.5 kHz and acquisition times of 12.8 and 25 ms in t_1 and t_2 , respectively. Data were acquired for 12 h using a recycle delay of 2.7 s, corresponding to $5T_1$ (T_1 is the longitudinal relaxation time measured by a saturation recovery experiment). A Lorentz-to-Gauss apodization function was used in both dimensions followed by zero filling to $4096 (t_1) \times 8192 (t_2)$ prior to Fourier transformation. Contour levels are shown from 8σ (σ being the noise root-mean-square determined by SPARKY), with each additional level multiplied by 1.4. (b) Strip plots from 3D heteronuclear correlation experiments showing residues 12–14 of the M13 capsid protein. *Red* and *blue spectra* correspond to intra-residue NCA CX and sequential NCO CX experiments, respectively. *Horizontal bars* link the strips from ^{15}N planes of residues j and $j - 1$ that share the C' - $\text{C}\alpha$ peaks of residue $j - 1$. The *bottom* sub-spectra correspond to 2D planes at the same ^{15}N shift and display $\text{N}-\text{C}'-\text{C}\beta$ or $\text{N}-\text{C}\alpha-\text{C}\beta$ crosspeaks

^{13}C - ^{13}C homonuclear correlation experiment acquired using a mixing time of 15 ms (DARR15) and collected for the intact uniformly labeled M13 phage. Spin-systems with well-resolved chemical shifts are marked as well as the full pattern of the single proline residue.

- In 3D experiments, identify clear ^{15}N chemical shifts of Gly, Ser, and Thr, all having shifts <115 ppm (Gly mostly

~100–110 ppm, Ser and Thr ~115 ppm). With their unique $C\alpha/C\beta$ shifts, they serve as good starting points for the sequential assignment process (*see* also **Note 10**).

6. With anchor assignments obtained, identify and link additional amino acids by matching $C\alpha$ -CO and $C\alpha$ - $C\beta$ resonance pairs from NCACX to CO- $C\alpha$ and CO- $C\beta$ pairs from the NCOCX experiment. If the linkage is unique, record the assignment. If there is overlap (common for helical capsids), record several options and link to the next residue. Most of these linkages will be unique. Strip plots from the 3D experiments (for residues 12–14 in M13 phage coat protein) are shown in Fig. 9.
7. The assignment of N, CO, $C\alpha$, and $C\beta$ is aided using the known average shifts of amino acids either from the BMRB, or from average shifts of specific secondary structure elements [50].
8. Link the backbone resonances obtained in **step 6** with the 2D data sets to complete the assignment of all residues, including sidechains (the “side-chain walk”).
9. Use 2D ^{13}C - ^{13}C correlation experiments acquired with longer mixing times (~100 ms) to obtain inter-residue correlations such as $C\alpha_i$ - $C\alpha_{i\pm 1}$ and $C\beta_i$ - $C\beta_{i\pm 1}$ and verify the assignments.
10. Assignment of aromatic crosspeaks can be hindered by spectral congestion, scalar couplings, and internal dynamics. Yet, *aromatic signals are extremely important due to their role in phage inter-subunit packing*. The structure cannot be solved without their assignment. These assignments most probably require the analysis of 2D and 3D spectra acquired with sparsely labeled M13 samples (1,3-gly and 2-gly, Subheading 3.1.2, or other types of labeling schemes).
11. The N-terminus of the coat protein is mobile and either undetected or very weakly detected in DARR spectra. Conduct a 2D RFDR experiment or DARR with long mixing times. Optionally record the through-bond (zfr)-INADEQUATE experiment [51, 52].

3.3.2 ssDNA Nucleotide Assignments

1. Prepare an aromatic-unlabeled sample (Subheading 3.1.3).
2. Perform a 2D ^{13}C - ^{13}C correlation experiment (DARR—Subheading 3.2.3 or CORD—Subheading 3.2.4) and process the data.
3. Perform a 2D ^{15}N - ^{13}C TEDOR experiment [53, 54] or a 2D ^{15}N - ^{13}C DCP experiment [32] and process the data.
4. Locate anchor signals of the DNA (T: deoxythymidine; C: deoxycytidine; A: deoxyadenine; G: deoxyguanine): Individual signals of methyl TC7 (14.4 ppm) and the quaternary carbon TC5 (113.6 ppm); CC5 (~100 ppm); the well-resolved

6. Link the nucleobase signals to the ribose moieties (C1'–C5').
7. Use the results and additional spectral cross-peaks to observe protein-DNA contacts.

3.4 Structure Determination

3.4.1 NMR Restraints

1. Collect 2D ^{13}C - ^{13}C NMR experiments (Subheadings 3.2.3 and 3.2.4) using sparsely labeled samples (Subheading 3.1.2) at various mixing times ranging from 50 to 500 ms. Process the data (Subheading 3.2.7).
2. Use the assignment table (Subheading 3.3.1) to identify non-ambiguous contacts. These are signals in the 2D spectrum that have unique chemical shifts in the assignment list, i.e., no other shift can be found within 0.2–0.3 ppm (user choice according to the goodness of the assignment and estimate of linewidths).
3. Differentiate intra-residue (between the atoms of the same residue i), medium-range ($i \leftrightarrow i + n$; $n \leq 4$), long-range ($i \leftrightarrow i + n$; $n > 4$), and inter-subunit contacts ($i \leftrightarrow j + n$; $n > 4$; i and j are the same residue in a different coat protein subunit). The latter two are difficult to differentiate. Since filamentous phages are helical, long-range contacts are almost exclusively inter-subunit. Exceptions are for the non-helical N-terminus, which may have contacts longer than four residues (*see* **Note 11**).
4. Identify ambiguous contacts with a small degree of ambiguity—not more than two options for each peak.
5. Set for all these restraints a distance limit of 2–8 Å. Feed them into Rosetta for a first round of calculations (Subheading 3.4.2).
6. After Rosetta calculations of the first round are performed, use the preliminary model to reduce ambiguity—rule out contacts with distances > 8 Å.
7. The residue Trp can be used as a scale bar if it is rigid (this can be determined by DIPSHIFT [56] measurements for example or by observing strong $\text{C}\alpha$ - $\text{C}\gamma$ contacts at short mixing times). Signals that appear in 2D experiments where no contacts between $\text{C}\alpha$ and the six-membered indole ring exist can be limited to 5 or 5.5 Å [7]. These data can further improve the results. Alternative methods can also be used, e.g., classification to weak, medium, and strong peaks [57].
8. Repeat **steps 5–7** until convergence is obtained.

3.4.2 Rosetta Calculations

Rosetta symmetric modeling of molecular assemblies (fold-and-dock [58]) uses a predefined set of explicit symmetry operations to model the rigid-body degrees of freedom, together with Monte Carlo refinement of the backbone torsion angles within each polypeptide chain of the system. The calculations consist of two stages; a low-resolution (centroid) sampling stage, which is followed by a

full-atom refinement stage. While in the centroid stage the side-chains are represented by centroid atoms, during full atom refinement the side-chain rotamer conformations are sampled explicitly toward optimizing a physically realistic energy function. Rosetta uses several input files that define the symmetry degrees of freedom, the NMR restraints, and the chemical shift information.

1. Symmetry

The symmetry degrees of freedom that define the position of each monomer subunit relative to its neighbors are contained within a Rosetta symmetry definition file (e.g., C5S2.symm). The commands for generating this file are described elsewhere [59]. In summary, the user needs to define the degree rotation per layer, which, in the case of perfect C_5S_2 symmetry, should be close to 36° . All other degrees of freedom (tilt, rise per pentamer, radius, subunit coordinates) are sampled according to the NMR restraints. In a general case, a rise per subunit, or per n -subunit fragment, is defined. For class-I phages, $n = 5$.

For example, among the top scoring models for the filamentous M13 phage calculations, the radius of the capsid, as measured from center of mass of the phage to the center of mass of one subunit, was 21.9–22.6 Å, the rise between pentamers was 16.6–16.7 Å between subunits and the tilt between pentamers was 36.1 – 36.6° , indicating a symmetry very close to C_5S_2 . This value can be validated using mass-per-length measurements from scanning transmission electron microscopy (STEM).

2. NMR Constraints

For the ssNMR distance constraints a flat-bottom potential with an upper limit of 5.0–8.2 Å (depending on the constraint type) and an exponential penalty function is used. The constraints applied in the two-sampling stages are defined in the *centroid.cst* and *fullatom.cst* files, respectively. Within the files, each line defines a distance constraint, according to the format:

```
AtomPair CG2 769 C 766 BOUNDED 1.500 7.000 0.300 NOE;
```

where the potential parameters are defined as follows:

CG2: Atom 1

769: Residue 1

C: Atom 2

766: Residue 2

1.500: lower bound (Å)

7.000: upper bound (Å)

0.300: potential steepness

For the definition of centroids, all side-chain atoms are mapped to a single atom called “CEN” and a 1.5 Å padding per C-C bond will be added to the upper limit.

3. Chemical shift-derived fragments

The following input files are required:

frags.score.200.3mers, *frags.score.200.9mers*:

Prior to executing the structure calculations, the user needs to select 3mer and 9mer backbone fragments according to the chemical shift data, as outlined in detail by Vernon et al. [60].

4. Fold-and-dock structure refinement

Typical calculations are carried out in parallel in a commodity LINUX cluster, using the Rosetta3 command:

```
rosetta_scripts.default.linuxgccrelease -parser:protocol
fold_and_dock.xml -nstruct 100 -out:file:silent models.out
-out:file:silent_struct_type binary
```

where,

nstruct: number of models to be computed (typically set to 50 batches of 100 models each).

models.out: Output file, in silent binary format.

fold_and_dock.xml: Rosetta scripts file in XML format, as described below.

```
<ROSETTASCRIPTS>
  <TASKOPERATIONS>
</TASKOPERATIONS>
  <SCOREFXNS>
    <score0 weights="score0" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=5.0/>
</score0>
    <score1 weights="score1" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=5.0/>
</score1>
    <score2 weights="score2" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=5.0/>
</score2>
    <score5 weights="score5" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=5.0/>
</score5>
    <score3 weights="score3" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=1.0/>
</score3>
    <talaris2013_symm weights="talaris2013" symmetric=1>
      <Reweight scoretype=atom_pair_constraint weight=0.5/>
</talaris2013_symm>
```

```

<talaris2013cart_symm weights="talaris2013_cart" symmetric=1>
  <Reweight scoretype=atom_pair_constraint weight=0.5/>
</talaris2013cart_symm >
</SCOREFXNS>
<FILTERS>
  <ScoreType name=cst_check scorefxn=score3
score_type=atom_pair_constraint threshold=3000/>
</FILTERS>
<MOVERS>
  <SetupForSymmetry name=symm definition="C5S2.symm"/>

  Fragment movers
  <SingleFragmentMover name="frag9"
fragments="frags.score.200.9mers" policy="uniform"/>
  <SingleFragmentMover name="frag3"
fragments="frags.score.200.3mers" policy="smooth"/>

  Fold-and-dock specific movers
  <SymFoldandDockRbTrialMover name="rbtrial" rot_mag=8.0
trans_mag=3.0 rotate_anchor_to_x=1/>
  <SymFoldandDockRbTrialMover name="rbtrial_smooth"
rot_mag=1.0 trans_mag=0.1 rotate_anchor_to_x=1/>
  <SymFoldandDockMoveRbJumpMover name="rbjump"/>
  <SymFoldandDockSlideTrialMover name="slidetrials"/>

  Random movers
  <RandomMover name=early_stage_moveset
movers="frag9,rbtrial,rbjump,slidetrials" weights="1.0,0.2,1.0,0.1" repeats=1/>
  <RandomMover name=final_stage_moveset
movers="frag3,rbtrial_smooth,rbjump,slidetrials" weights="1.0,0.2,1.0,0.1" repeats=1/>

  Monte Carlo Movers
  <GenericMonteCarlo name="stage1" scorefxn_name="score0"
mover_name="early_stage_moveset" temperature=2.0 trials=200 recover_low=1/>
  <GenericMonteCarlo name="stage2" scorefxn_name="score1"
mover_name="early_stage_moveset" temperature=2.0 trials=200 recover_low=1/>
  <GenericMonteCarlo name="stage3a" scorefxn_name="score2"
mover_name="early_stage_moveset" temperature=2.0 trials=20 recover_low=1/>
  <GenericMonteCarlo name="stage3b" scorefxn_name="score5"
mover_name="early_stage_moveset" temperature=2.0 trials=20 recover_low=1/>
  <GenericMonteCarlo name="stage4" scorefxn_name="score3"
mover_name="final_stage_moveset" temperature=2.0 trials=400 recover_low=1/>

  Special stage 3 logic
  <ParsedProtocol name="stage3_cyc">
    <Add mover=stage3a/>
    <Add mover=stage3b/>
  </ParsedProtocol>
  <LoopOver name=stage3 mover_name=stage3_cyc iterations=5 drift=1/>

```

```

Converts the centroid-level pose to fullatom for scoring
<SwitchResidueTypeSetMover name=fullatom set=fa_standard/>
<FastRelax name=relax scorefxn=talaris2013_symm repeats=4 />
<FastRelax name=relaxCart cartesian=1 repeats=1 scorefxn=talaris2013cart_symm/>
<ConstraintSetMover name=add_fullatom_csts cst_file="fullatom.cst"/>
<ConstraintSetMover name=add_centroid_csts cst_file="centroid.simple.cst"/>
<ClearConstraintsMover name=remove_csts />
</MOVERS>
<PROTOCOLS>
  <Add mover=symm/>
  <Add mover=add_centroid_csts/>
  <Add mover=stage1/>
  <Add mover=stage2/>
  <Add mover=stage3/>
  <Add mover=stage4/>
  <Add filter_name = cst_check/>
  <Add mover=remove_csts/>
  <Add mover=fullatom/>
  <Add mover=add_fullatom_csts/>
  <Add mover=relax/>
  <Add mover=relaxCart/>
</PROTOCOLS>
</ROSETTASCRIPTS>

```

4 Notes

1. We found that when growing XL1-tetR or DH5 α F' strains on minimal media solution for the production of 1,3-¹³C and 2-¹³C-glycerol labeled bacteriophages the pH increases at least by two units probably due to the presence of sodium bicarbonate. When using the regular minimal salts (at pH 7) the pH of the solutions after adding sodium bicarbonate increases to over 8 and the efficiency of growth decreases dramatically. The starting pH for such a growth is therefore set to 6.
2. The NMR probe temperature is controlled in the entry to the probe. The effective sample temperature can increase up to 10–25 °C due to frictional heating induced by sample spinning. RF-induced heating can also occur, in particular for non-Efree probes. The exact temperature therefore depends on the type of probe, on the gas flow rate, on the spinning rate, and on the decoupling strength. Yet, phage samples are relatively stable over a wide range of temperatures and setting the controller between –15 and 5 °C is sufficient to get reproducible spectra, and sample stability and infectivity are retained.

3. The amount of 500 mL should be grown in a 2 L Erlenmeyer at least. In general, a ratio of 1:4 or 1:5 between the culture volume and flask volume should be maintained, and we have found that it is best to use the largest ratio possible to allow sufficient aeration.
4. “Checkerboard labeling” is an expression that describes a situation, in which alternating carbons in a protein are ^{13}C labeled. There are variations to this rule since for some amino acids types labeling also occurs at adjacent carbon positions and for some there is scrambling (only a certain percentage of a particular carbon is labeled). A scheme describing the labeling pattern can be found in the work by Oschkinat and coworkers for example [61].
5. The ultracentrifuge tubes must be filled completely and balanced very carefully; otherwise they may fail. Balancing should be done separately for each pair of tubes positioned exactly across each other in the rotor.
6. If the band containing the phage is not sufficiently isolated, change the density of the solution to move the band up or down. The position will change for different labeling schemes and for different types of phages. The tube after the removal of the phage may still contain some phage, or in case where the PEG and phage bands are mixed, the entire procedure has to be repeated.
7. Mixing the solution of 10 μL phage/90 μL bacteria is done by continuous gentle pipetting of the solution, while the pipette is set to a volume of 10 μL . Do not press the pipette down all the way (complete expulsion of liquid) to avoid production of air bubbles.
8. Acquisition times of 500 ms may damage the probe. Care should be taken to ensure that the power is sufficiently low and that the duty cycle is not too high. Approximately 10–14 s between scans is satisfactory for Adamantane.
9. Watch the incoming and outgoing power on the scope screen and by matching and tuning set a concave shape of power coming out of the probe on the ^1H channel with a minimum in the middle of the acquisition block (*see* Fig. 11). Note that for some samples the outgoing signal will not reach a concave shape but a flat shape.
10. Resonances in 3D spectra can be tracked easily by observing the projections over the three orthogonal planes (e.g., $\text{NC}\alpha$, $\text{C}\alpha\text{CX}$, NCX from NCACX). These projections are helpful for assessing the location of spinning sidebands, for assessing the completeness of the experiment and the sensitivity. Comparison to the corresponding 2D data sets (DCP and DARR/CORD) is extremely helpful.

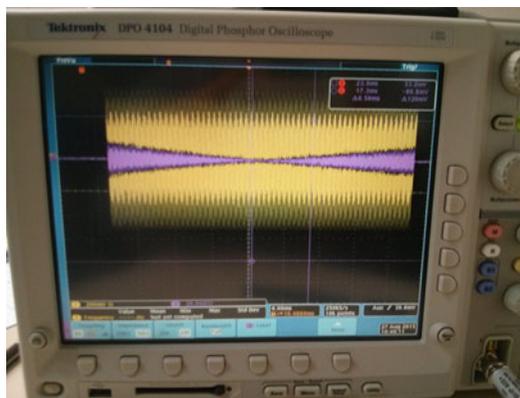


Fig. 11 High-power tuning as observed on the scope. The flat signal is the forward input power coming directly from the pre-amplifier and directed to the probe. The concave signal is the reflected power from the probe

11. If the coat protein is not entirely helical, then differentiating intra- from inter-subunit contacts is complicated. This is the case for general filamentous structures or for other types of bacteriophages and requires additional strategies. See, for example, the NMR study of the type-III secretion needle by Lange and coworkers [62].

References

1. Russel M (1991) Filamentous phage assembly. *Environ Microbiol* 5:1607–1613
2. Day LA (2008) Inoviruses. In: Mahy BWJ, van Regenmortel MHV (eds) *Encyclopedia of virology*. Academic, Oxford, UK, pp 117–124
3. Marvin DA (1998) Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol* 8:150–158
4. Welsh LC, Symmons MF, Marvin DA (2000) The molecular structure and structural transition of the α -helical capsid in filamentous bacteriophage Pfl. *Acta Crystallogr D Biol Crystallogr* 56:137–150
5. Marvin DA, Welsh LC, Symmons MF, Scott WRP, Straus SK (2006) Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. *J Mol Biol* 355:294–309
6. Zeri AC, Mesleh MF, Nevzorov AA, Opella SJ (2003) Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy. *Proc Natl Acad Sci USA* 100:6458–6463
7. Morag O, Sgourakis NG, Baker D, Goldbourn A (2015) The NMR–Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc Natl Acad Sci USA* 112:971–976
8. Sergeyev IV, Bahri S, Day LA, McDermott AE (2014) Pfl bacteriophage hydration by magic angle spinning solid-state NMR. *J Chem Phys* 141:22D533
9. Lorieau JL, Day LA, McDermott AE (2008) Conformational dynamics of an intact virus: order parameters for the coat protein of Pfl bacteriophage. *Proc Natl Acad Sci USA* 105:10366–10371
10. Goldbourn A, Day LA, McDermott AE (2010) Intersubunit hydrophobic interactions in Pfl filamentous phage. *J Biol Chem* 285:37051–37059
11. Morag O, Abramov G, Goldbourn A (2011) Similarities and differences within members of the Ff family of filamentous bacteriophage viruses. *J Phys Chem B* 115:15370–15379
12. Abramov G, Morag O, Goldbourn A (2011) Magic-angle spinning NMR of a class I filamentous bacteriophage virus. *J Phys Chem B* 115:9671–9680

13. Morag O, Abramov G, Goldbourt A (2014) Complete chemical shift assignment of the ssDNA in the filamentous bacteriophage fd reports on its conformation and on its interface with the capsid shell. *J Am Chem Soc* 136:2292–2301
14. Lian L-Y, Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog Nucl Magn Reson Spectrosc* 39:171–190
15. Hoogstraten C, Johnson J (2008) Metabolic labeling: taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concepts Magn Reson A* 32:34–55
16. Stanek J, Andreas LB, Jaudzems K, Cala D, Lalli D, Bertarello A et al (2016) NMR spectroscopic assignment of backbone and side-chain protons in fully protonated proteins: microcrystals, sedimented assemblies, and amyloid fibrils. *Angew Chem Int Ed* 55:15504–15509
17. van Wezenbeek P, Hulsebos TJM, Schoenmakers JGG (1980) Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* 11:129–148
18. Vieira J, Messing J (1987) Production of single-stranded plasmid DNA. *Methods Enzymol* 153:3–11
19. Day LA, Marzec CJ, Reisberg SA, Casadevall A (1988) DNA packing in filamentous bacteriophages. *Annu Rev Biophys Biomol Struct* 17:509–539
20. Enshell-Seiffers D, Smelyanski L, Gershoni JM (2001) The rational design of a “type 88” genetically stable peptide display vector in the filamentous bacteriophage fd. *Nucleic Acids Res* 29:E50
21. Bradley DE (1979) Morphology of pili determined by the N incompatibility group plasmid N3 and interaction with bacteriophages PR4 and IKe. *Plasmid* 2:632–636
22. Coetzee JN, Bradley DE, Hedges RW (1982) Phages I α and I $_2$ -2: IncI plasmid-dependent bacteriophages. *J Gen Microbiol* 128:2797–2804
23. Kuo T-T, Huang T-C, Chow T-Y (1969) A filamentous bacteriophage from *Xanthomonas oryzae*. *Virology* 39:548–555
24. Cai M, Huang Y, Sakaguchi K, Clore GM, Gronenborn AM, Craigie R (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in shape *Escherichia coli*. *J Biomol NMR* 11:97–102
25. Stringer JA, Bronnimann CE, Mullen CG, Zhou DH, Stellfox SA, Li Y et al (2005) Reduction of RF-induced sample heating with a scroll coil resonator structure for solid-state NMR probes. *J Magn Reson* 173:40–48
26. LeMaster D, Kushlan D (1996) Dynamical mapping of *E. coli* thioredoxin via ^{13}C NMR relaxation analysis. *J Am Chem Soc* 7863:9255–9264
27. Tomar S, Green MM, Day LA (2007) DNA-protein interactions as the source of large-length-scale chirality evident in the liquid crystal behavior of filamentous bacteriophages. *J Am Chem Soc* 129:3367–3375
28. Kostrikis LG, Liu DJ, Day LA (1994) Ultraviolet absorbance and circular dichroism of Pfl virus: nucleotide/subunit ratio of unity, hyperchromic tyrosines and DNA bases, and high helicity in the subunits. *Biochemistry* 33:1694–1703
29. Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. *J Magn Reson* 162:479–486
30. Pines A, Gibby MG, Waugh J (1973) Proton-enhanced NMR of dilute spins in solids. *J Chem Phys* 59:569–590
31. McDermott AE, Gu Z (1996) Carbon and nitrogen chemical shifts: applications to solid state proteins. In: Grant DM, Harris RK (eds) *Encyclopaedia of nuclear magnetic resonance*. Wiley, Chichester, pp 1137–1147
32. Shi L, Ladizhansky V (2012) Magic angle spinning solid-state NMR experiments for structural characterization of proteins. *Meth Mol Biol* 895:153–165
33. Schaefer J, McKay R, Stejskal E (1979) Double-cross-polarization NMR of solids. *J Magn Reson* 34:443–447
34. Bennett AE, Rienstra CM, Auger M, Lakshmi KV, Griffin RG (1995) Heteronuclear decoupling in rotating solids. *J Chem Phys* 103:6951
35. Fung BM, Khitrin AK, Ermolaev K (2000) An improved broadband decoupling sequence for liquid crystals and solids. *J Magn Reson* 142:97–101
36. Detken A, Hardy EH, Ernst M, Meier BH (2002) Simple and efficient decoupling in magic-angle spinning solid-state NMR: the XiX scheme. *Chem Phys Lett* 356:298–304
37. Madhu PK (2014) Heteronuclear spin decoupling in solid-state nuclear magnetic resonance: overview and outlook. *Isr J Chem* 54:25–38
38. Thakur RS, Kurur ND, Madhu PK (2006) Swept-frequency two-pulse phase modulation for heteronuclear dipolar decoupling in solid-state NMR. *Chem Phys Lett* 426:459–463
39. Takegoshi K, Nakamura S, Terao T (2001) ^{13}C - ^1H dipolar-assisted rotational resonance

- in magic-angle spinning NMR. *Chem Phys Lett* 344:631–637
40. Goldbourn A (2009) Magic-angle spinning solid-state nuclear magnetic resonance: application to structural biology. In: Meyers RA (ed) *Encyclopedia of analytical chemistry*. Wiley, Chichester, pp 1–27
 41. Hou G, Yan S, Trébosc J, Amoureux J-P, Polenova T (2013) Broadband homonuclear correlation spectroscopy driven by combined R_2^V sequences under fast magic angle spinning for NMR structural analysis of organic and biological solids. *J Magn Reson* 232:18–30
 42. Baldus M, Petkova AT, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol Phys* 95:1197–1207
 43. Wilhelm M, Feng H, Tracht U, Spiess HW (1998) 2D CP/MAS ^{13}C isotropic chemical shift correlation established by ^1H spin diffusion. *J Magn Reson* 134:255–260
 44. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J et al (2008) Biological magnetic resonance data bank. *Nucleic Acids Res* 36:D402–D408
 45. Bennett AE, Griffin RG, Ok JH, Vega S (1992) Chemical shift correlation spectroscopy in rotating solids: radio frequency-driven dipolar recoupling and longitudinal exchange. *J Chem Phys* 96:8624
 46. Ishii Y (2001) ^{13}C – ^{13}C dipolar recoupling under very fast magic angle spinning in solid-state nuclear magnetic resonance: Applications to distance measurements, spectral assignments, and high-throughput secondary-structure determination. *J Chem Phys* 114:8473–8483
 47. Pauli J, van Rossum B, Forster H, de Groot HJM, Oschkinat H (2000) Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the alpha-spectrin SH3 domain. *J Magn Reson* 143:411–416
 48. Lesage A, Bardet M, Emsley L (1999) Through-bond carbon-carbon connectivities in disordered solids by NMR. *J Am Chem Soc* 121:10987–10993
 49. Lesage A, Auger C (1997) Determination of through-bond carbon-carbon connectivities in solid-state NMR using the INADEQUATE experiment. *J Am Chem Soc* 7863:7867–7868
 50. Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* 11:852–861
 51. Grasso G, de Swiet T, Titman JJ (2002) Electronic structure of the polymer phase of CsC60: refocused INADEQUATE experiments. *J Phys Chem B* 106:8676–8680
 52. Cadars S, Sein J, Duma L, Lesage A, Pham TN, Baltisberger JH et al (2007) The refocused INADEQUATE MAS NMR experiment in multiple spin-systems: interpreting observed correlation peaks and optimising lineshapes. *J Magn Reson* 188:24–34
 53. Hing A, Vega S, Schaefer J (1993) Measurement of heteronuclear dipolar coupling by transferred-echo double-resonance NMR. *J Magn Reson A* 103:151–162
 54. Abramov G, Goldbourn A (2014) Nucleotide-type chemical shift assignment of the encapsulated 40 kbp dsDNA in intact bacteriophage T7 by MAS solid-state NMR. *J Biomol NMR* 59:219–230
 55. Sergeev IV, Day LA, Goldbourn A, McDermott AE (2011) Chemical shifts for the unusual DNA structure in Pfl bacteriophage from dynamic-nuclear-polarization-enhanced solid-state NMR spectroscopy. *J Am Chem Soc* 133:20208–20217
 56. DeAzevedo ER, Saalwachter K, Pascui O, de Souza AA, Bonagamba TJ, Reichert D (2008) Intermediate motions as studied by solid-state separated local field NMR experiments. *J Chem Phys* 128:104505
 57. Franks WT, Wylie BJ, Schmidt HLF, Nieuwkoop AJ, Mayrhofer R-M, Shah GJ et al. (2008) Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR. *Proc Natl Acad Sci USA* 105:4621–4626
 58. Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S et al. (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci USA* 106:18978–18983
 59. DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I (2011) Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 6:e20450
 60. Vernon R, Shen Y, Baker D, Lange OF (2013) Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J Biomol NMR* 57:117–127
 61. 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
 62. Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C et al (2012) Atomic model of the type III secretion system needle. *Nature* 486:276–279

Spherical Nanoparticle Supported Lipid Bilayers: A Tool for Modeling Protein Interactions with Curved Membranes

Erin R. Tyndall and Fang Tian

Abstract

Mechanistic studies of protein-membrane interactions can be complicated by the limitations of the membrane model system chosen. Many of these limitations can be overcome by using a spherical silica nanoparticle to support the membrane. In this chapter, we present a detailed protocol for the construction of spherical nanoparticle supported lipid bilayers (SSLBs), with discussion of methods to improve production.

Key words Spherical nanoparticle supported lipid bilayers (SSLBs), Membrane curvature, Curvature recognition, NMR

1 Introduction

Protein-membrane interactions are the key to a variety of essential cellular processes including autophagy, mitosis, endocytosis, and vesicular trafficking [1, 2]. However, the molecular mechanisms of how proteins interact with their membrane environment, and how the membrane environment influences their activities, remain elusive in most cases. This is particularly true for an emerging number of proteins that depend upon membrane geometry for localization or activity [3–7]. Force measurements, as well as structural and mechanistic studies of these proteins, strongly benefit from a modeling system where membrane curvature can be stably and precisely defined [8, 9].

Common membrane mimics include micelles, bicelles, nanodics, small unilamellar vesicles (SUVs), large or giant unilamellar vesicles (LUV/GUV), and planar supported lipid bilayers (Fig. 1). There are however some caveats to these systems when they are applied to simulate the native curved membrane environment; micelles are very small, with a diameter of ~5 nm, and are extremely curved, while bicelles and nanodics are most commonly used to provide a planar membrane surface [10–12]. SUVs and LUVs can

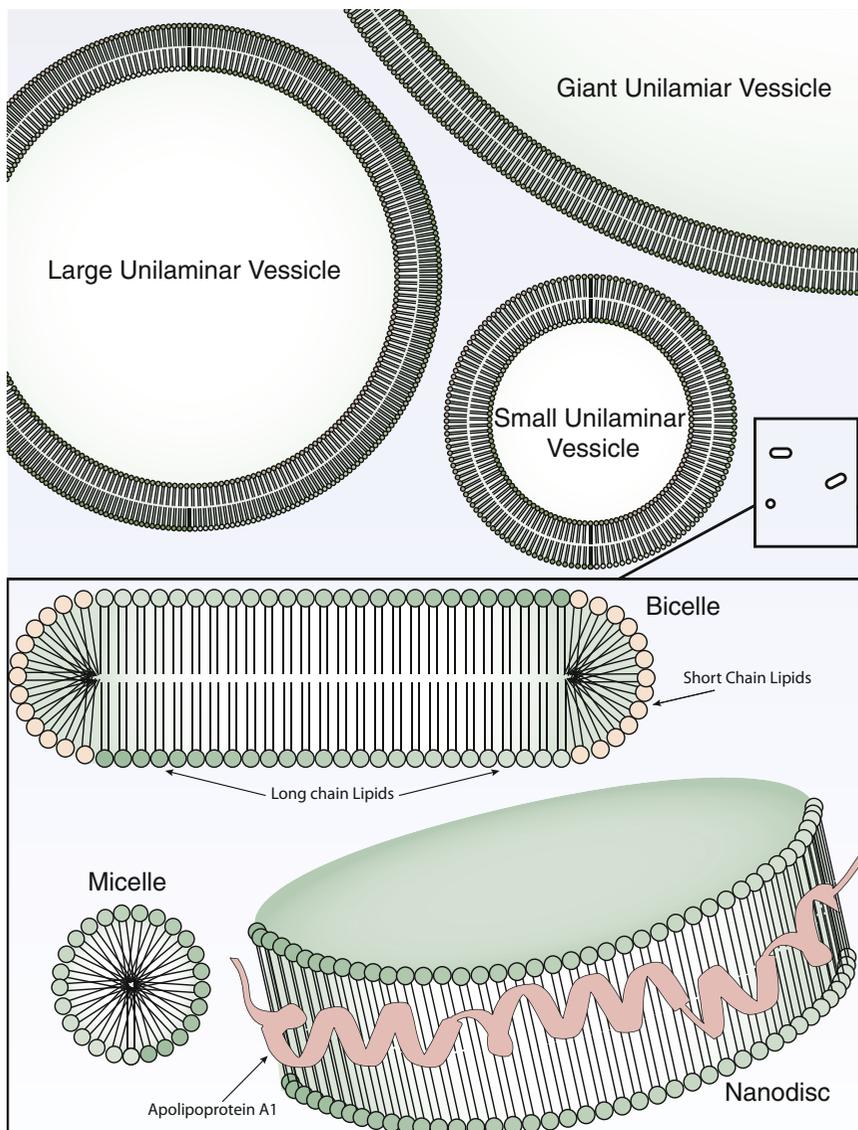


Fig. 1 Pictorial representations of commonly used membrane mimics

be produced with a diameter distribution within a certain range, but their sizes are difficult to adjust and control rigorously [13, 14]. And as the radius increases, as in GUVs, so do stability problems, which can lead to membrane blebbing, vesicle fusion, and other undesirable changes in membrane structure [15].

However, spherical nanoparticle supported lipid bilayers (SSLBs) [16], unilamellar vesicles supported by a silica bead, overcome many of these limitations [17, 18]. In the prevailing model for SSLBs (supported by studies using an array of techniques including fluorescent microscopy, differential scanning calorimetry, NMR, and electron microscopy) the adsorbed lipids follow the

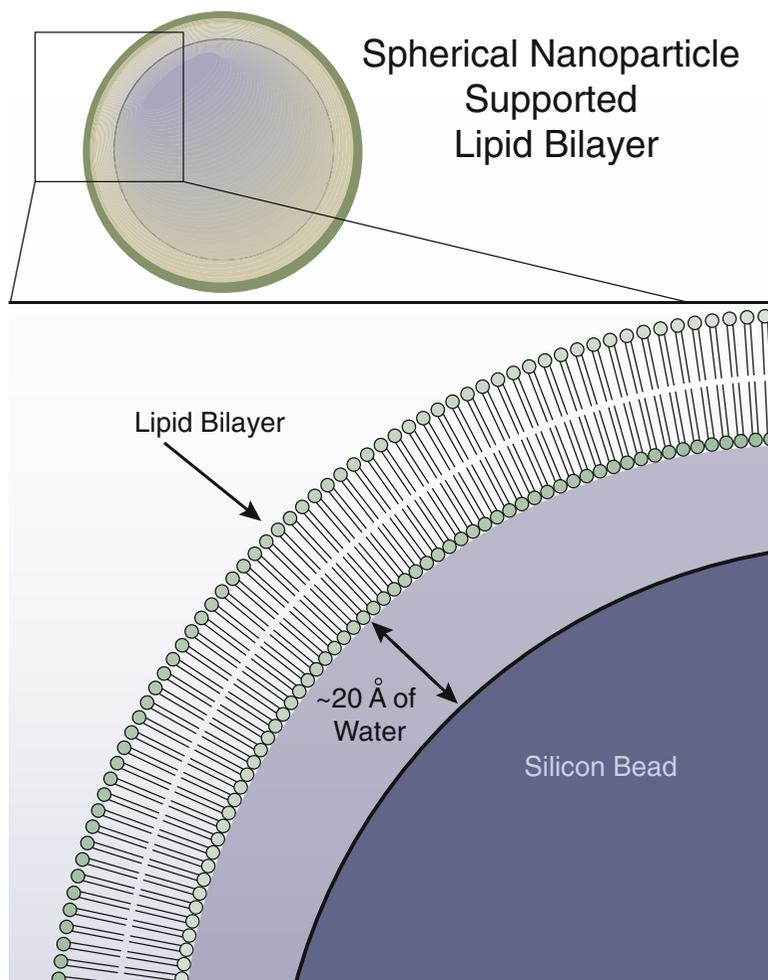


Fig. 2 Spherical nanoparticle supported lipid bilayers. The size of the SSLB is dictated by the size of the bead. As seen in the lower portion, a layer of water molecules separates the lipids and the supporting silicon bead

surface of the silicon beads and form a single bilayer [19–23]. The bilayer is not in direct contact with its support, and is instead supported by a buffer of water molecules (Fig. 2). Consequently, this solid but cushioned support allows SSLBs to retain many of the essential properties of cellular membranes, including a well-defined geometry of the bilayer. The mechanical stability provided by the interior support allows the bilayer to be more resistant to potential changes in membrane shape when interacting with proteins, while imposing few constraints on the fluidity of the lipids themselves. These unique advantages of SSLBs have been explored to study protein-lipid interaction in the past [24–28]. We have recently employed SSLBs to model different curved membranes and mechanistically and structurally studied their interactions with a membrane curvature-recognizing peptide using fluorescent microscopy

and NMR [8, 9]. In this chapter, we will describe protocols for the preparation of SSLBs.

2 Materials

2.1 Materials for Lipid Preparation

1. A SpeedVac concentrator with condenser.
2. A lyophilizer.
3. Glass tubes between 12 and 15 mm in diameter, and compatible with your SpeedVac.
4. A laboratory fume hood.
5. Either the requisite lipids as powdered lipids, or dissolved in chloroform. *See* Subheading 2.3 for details.
6. Chloroform.
7. Chloroform-safe pipette tips such as S8064 from Sigma.
8. Ultra-pure water.
9. A vortex machine.
10. Parafilm.
11. Desktop centrifuge and compatible tubes.
12. A water bath at 42 °C.
13. A dry ice and methanol bath.
14. A sonication bath.

2.2 Materials for SSLB Preparation

1. Silica beads in aqueous solution or powdered silica beads of the desired diameter. *See* Subheading 2.3 for details.
2. Ultra-pure water.
3. Methanol.
4. Desktop centrifuge for beads larger than 50 nm, ultracentrifuge for beads smaller than 50 nm, with compatible tubes.
5. 15 mL centrifuge tubes.
6. 1 mM CaCl₂ in ultra-pure water.

2.3 Additional Resources

1. Resources for lipids.

Avanti Polar Lipids sells a wide range of lipids, including phospholipids and lipid extracts. Lipids can be obtained both in powder form and in chloroform solutions.

A smaller number of lipids may be obtained from Sigma-Aldrich or Anatrace. The latter has a similar range of lipids as Sigma-Aldrich that can be obtained in powder form.

2. Resources for silica beads.
 - (a) Nissan Chemical: Snowtex product line. This supplier produces a variety of beads that are provided in an aqueous solution. Some specially shaped beads are also available if your application requires them.
 - (b) Fisher: Size standards. These products are rigorously screened to the National Institute of Standards and Technology's standards for particle size and size distribution. A variety of sizes are available, though some product lines have fewer size options.
 - (c) Cospheric Nano: Offers both nanospheres and microspheres; products are shipped as a dry powder and must be rehydrated.

3 Methods

3.1 Preparation of SUVs

1. Begin by selecting the lipids you wish to use for your SSLBs. If your targeted protein is lipid-makeup insensitive, we recommend 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) as its phase transition temperature allows the resulting SSLBs to be stable at room temperature. If you are using high percentage of charged lipids or lipids of disparate length, you will need to take into consideration the various phase transition temperatures and possible interactions with the negatively charged silica beads, *see Note 1*.
2. Weigh out your lipids in a glass tube and add enough chloroform to dissolve them. If your lipids are already in chloroform, simply add them to the tube. When using chloroform, be sure to not allow it to contact plastic for any length of time. Our laboratory's standard preparation size is 15 mg of DMPC in 200 μ L of chloroform. *See Fig. 3b, c* for a visual reference, and **Note 2** for caveats.
3. Use the SpeedVac to extract the solvent and deposit the lipids on glass. Usually, this will require 2–4 h, longer if you used a larger volume of chloroform. Once the solvent has been removed, the lipids should appear completely dry, as in *Fig. 3d*. At this point, transfer the glass tube with your lipids to the lyophilizer overnight. This will result in a completely dried lipid film. If non-DMPC lipids are used, the dried film may appear more crystalline.
4. To re-hydrate the lipids add ultra-pure water to the tube to bring the lipid concentration to 15 mg/mL. Seal with parafilm and allow the lipids to sit at 42 °C for at least an hour with vortexing every 15–20 min (*see Note 3*). The lipids appear as a particulate suspension in water, and are pictured in *Fig. 3e*. At this point the mixture will separate if left to sit.

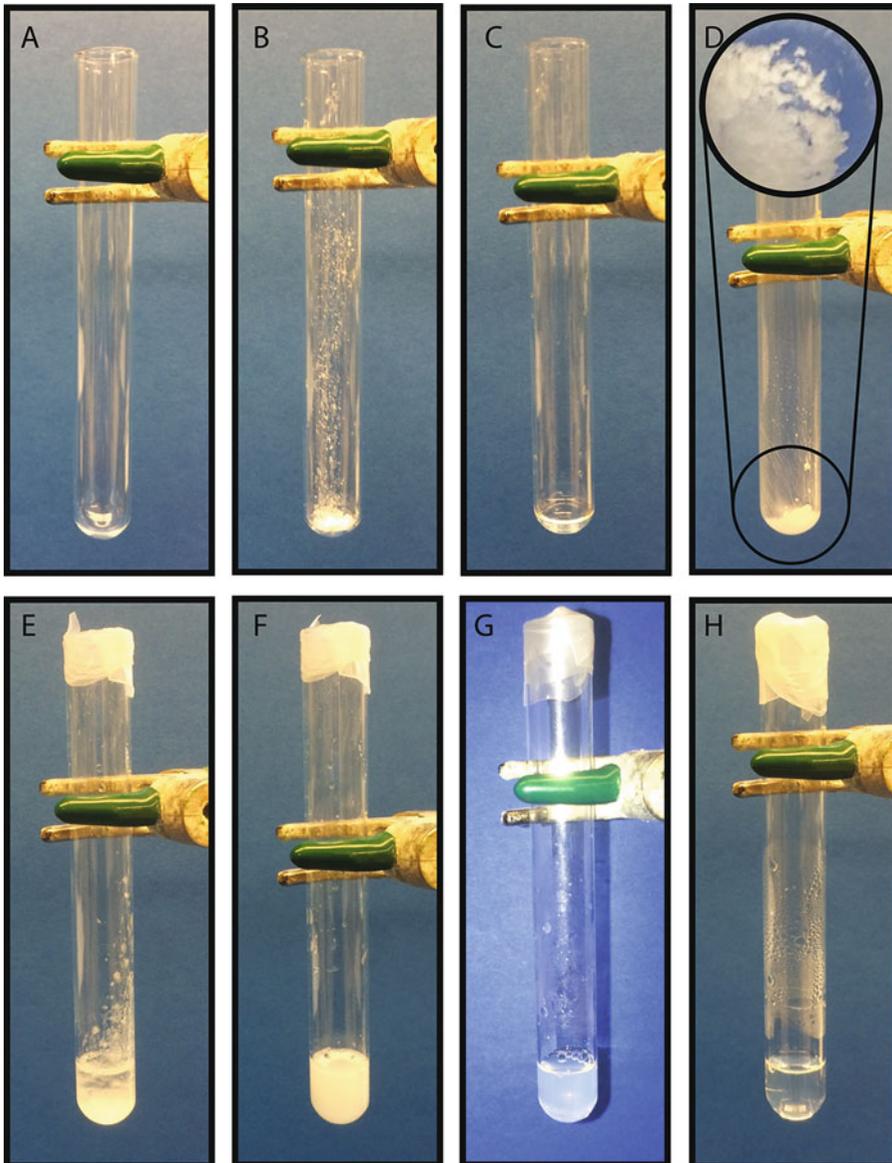


Fig. 3 SUV Preparation. Liposome preparations begin with a glass tube (a). Add lipids (b) and dissolve in chloroform (c). After depositing on glass, the lipids should be completely dried (d and *inset*). Rehydrate lipids with water (e), and vortex repeatedly to resuspend (f). After a freeze-thaw cycle, return to the glass tube for sonication; lipids should retain appearance. Sonicate until clear (h), (g) shows an example of incompletely sonicated liposomes, which retains opacity

5. Freeze-thaw the lipid-water mixture. Briefly spin the tube to recover and collect everything at the bottom, and then transfer the suspension to a plastic centrifuge tube. Incubate the tube in the methanol-dry ice bath for 15 min, and allow to thaw, while frequently vortexing. Repeat four times, to fully freeze-fracture the lipids into vesicles. At this point the lipids will appear

translucent, and should remain suspended in solution longer, this is exemplified in Fig. 3f.

- For the formation of the SUVs, transfer the lipid mixture into a clean glass tube. Seal the tube with copious amounts of parafilm and insert into a float and add to the sonication bath. Sonicate until the solution becomes clear, or at least translucent; compare Fig. 3g, h (*see Note 4*). Place your SUVs in the 42 °C water bath until needed; they should remain stable in the solution for a few days. We recommend not storing re-hydrated lipids for more than a week before use.

3.2 Preparation of SSLBs

- Select the desired size of your SSLB; this should correspond to the membrane geometry your protein is known to interact with. If there are a range of sizes applicable to your protein, consider selecting the largest, as it will result in an easier preparation. *See Fig. 4a* for an example of 50 and 1000 nm beads.
- If your beads are in powder form, you will need to hydrate them before they can be used. Weigh out the powdered beads, and add 2–5 volumes of water. Allow to sit at 42 °C for at least 16 h, with vortexing. If the beads are not completely hydrated, it will lead to clumping.
- Shake bead solution to homogenous suspension and add desired volume of bead to a microcentrifuge tube. We usually use a total of 40 mg of beads in each tube.
- Spin the solution to pellet beads (*see Table 1* for spinning times associated with each bead size). Discard the supernatant, and add 1 mL of ultrapure water to resuspend the pellet by aggressive vortexing, *see Fig. 4b, c* for an example of how the pellet should appear when resuspending (*see Note 5*).

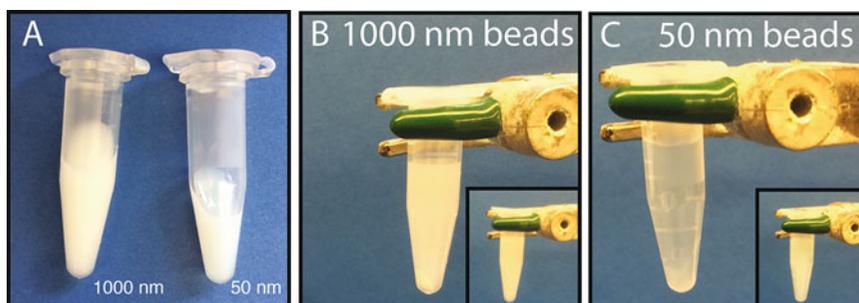


Fig. 4 Silicon Beads. 1000 and 50 nm beads aliquoted from commercial stock (a). The beads can be difficult to resuspend, and the smaller they are the harder it will be. (b) shows 1000 nm beads immediately after re-addition of water, and the increased opacity can clearly be seen; the *inset* shows how they appear in suspension. (c) shows the same for 50 nm beads, and it is clear that the beads remain largely in the pellet, and that the pellet is more difficult to see

Table 1
SSLB preparation

Bead size (nm)	Theoretical lipid required to coat beads (per mg beads) (mg)	Experimentally recommended amount (mg)	Time and speed needed to pellet
20–30	0.4	1.6	15 min at $200,000 \times g$
50	0.2	0.8	10 min at $16,000 \times g$
100	0.1	0.4	5 min at $16,000 \times g$
1000	0.01	0.04	2 min at $16,000 \times g$

5. Repeat **step 4** and wash beads with 1 mL methanol once and 1 mL water three times. For final resuspension, target 40 mg/mL in ultra-pure water. Beads should now be stable at room temperature (*see Note 6*).
6. To prepare SSLBs, first ensure all components are at 42 °C. Next, take desired volume of beads and bring volume to 2 mL with ultra-pure water in a 15 mL conical tube. To a second tube, add lipids according to the value in Table 1, and bring to 2 mL with ultra-pure water. Vigorously pipette the bead mixture into the lipid mixture, and quickly add 8 μ L of the 1 mM CaCl₂ solution. Vortex for at least 60 s.
7. Move the mixture solution to the 42 °C water bath for an hour with vortexing every 15 min. An example of how 1000 and 50 nm beads appear during this process is in Fig. 5a, b respectively.
8. Spin the SSLB solution according to Table 1. Once pelleted, discard the supernatant and wash pellet with 1 mL ultra-pure water. Vortex to suspend the pellet (*see Note 7*). Repeat three times. Figure 5c, d demonstrate both the pelleted and resuspended SSLBs.
9. Resuspend the pellet in 200 μ L water. SSLBs are now ready for use as seen in Fig. 5e.

4 Notes

1. If more than 15% of your lipids are negatively charged you may have to alter the protocol to account for this. A spacer can be used to loosely tether the lipids out further from the bead, such as using avidin as a linker between the lipids and the bead [29].
2. Some lipids and lipid mixtures are not soluble in chloroform. While this will not lead to immediate problems, it is usually

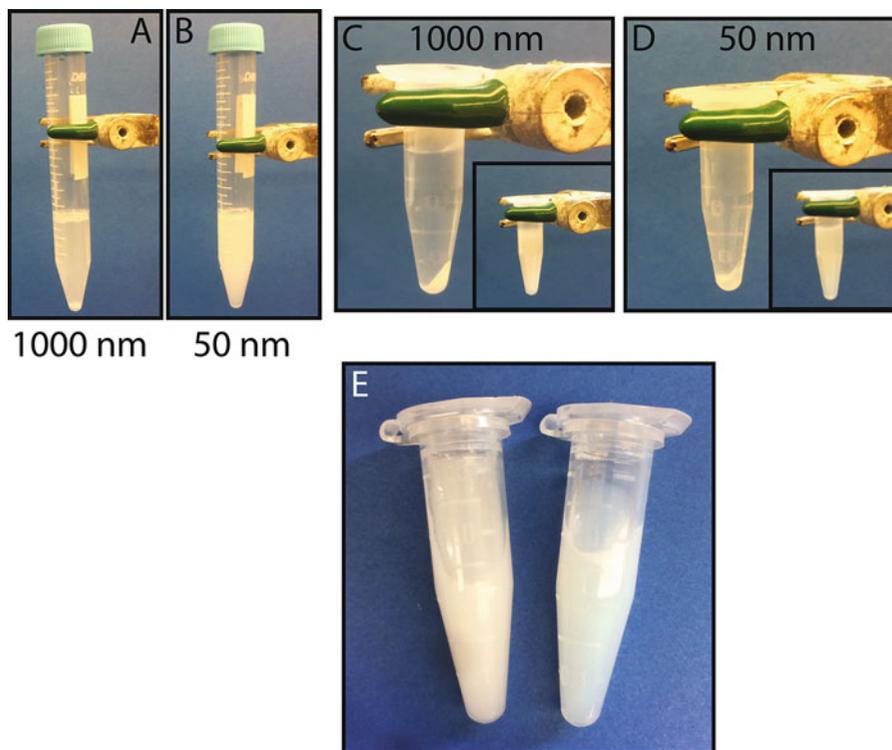


Fig. 5 SSLB preparation. (a) and (b) demonstrate how the SSLBs should appear after vortexing during incubation with the CaCl_2 . (c) The pellet for 1000 nm beads appears the same as before coating, and the resuspended mixture is largely opaque. (d) The pellet for 50 nm beads will appear more opaque, and the resuspension should appear iridescent, but not entirely opaque. (e) A side-by-side comparison of completed 1000 and 50 nm SSLBs

indicative of atypical behavior later: including SUVs not being stable at lower temperatures, or incomplete transition to SUV state. You may consider trying to optimize your lipid selection for those of similar lengths and phases transition temperatures. Additionally, be careful using whole lipid extracts, as they often are not pure, and are more difficult to completely form into SUVs.

3. The better hydrated the lipids are, the easier it will be to transition them into SUVs. Leaving the lipids at 42°C overnight with hour of intermittent vortexing the next day can be helpful for difficult lipid mixtures.
4. For optimal sonication, it is necessary to check on the tube every 15–20 min. It is best to try and keep the tube “pinned” on one of the nodes using other floats. It is easy to tell when you are in the correct place because the sound will shift slightly, and the surface of your lipid mixture becomes very volatile. However, maintaining position can be difficult, and even if correctly managed, the heat of sonication often melts the

parafilm. Thus, we recommend checking the position and parafilm integrity every 15 min. If after 1 h you have seen no change in opacity of the solution, try dividing it into two tubes. Figure 2g shows incompletely sonicated DMPC liposomes, as compared to h. Some lipid makeups will never become more translucent than g.

5. It can be difficult to resuspend the bead pellets. Allowing them to sit in a warm water bath for 10 min, or sonicating them may help loosen the pellet. It is also best to resuspend in an excess of water. For the preparation of a large amount of beads it is best to do it in small lots and combine at the end. The smaller the bead, generally the harder it is to resuspend. Do not pipette up and down; you will lose beads stuck in the tip.
6. Small beads, such as 20–100 nm, should remain in the solution rather than pelleting after resuspension. Larger beads will sink if left to sit on the bench. If your beads are not staying in the solution, they may need to be further hydrated before proceeding to the coating stage.
7. The SSLB pellet should appear similar to the bead pellet, but with a slight increase in opacity and a colored sheen on the smaller sizes. If you have trouble resuspending the SSLB pellet, allow the solution to sit in a water bath for 10–15 min. We do not recommend sonication once lipids have been added. Using excess SUVs during coating will make resuspension easier. If the pellet has turned milky white or if there are obvious chunks, then the SSLB coating has failed. It is best at this juncture to start over with another set of beads, and consider doubling the amount to CaCl_2 to stabilize the silica-lipid interactions.

Acknowledgment

We thank Nissan Chemical Industries, Ltd. for 50 and 100 nm silicon beads. This work was supported by the National Institutes of Health NIGMS (R01GM105963 to F.T.) and the Four Diamonds Fund. This project also is funded, in part, under a grant with the Pennsylvania Department of Health using Tobacco CURE Funds. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

References

1. Maxfield FR, McGraw TE (2004) Endocytic recycling. *Nat Rev Mol Cell Bio* 5:121–132
2. Nakatogawa H, Suzuki K, Kamada Y, Ohsumi Y (2009) Dynamics and diversity in autophagy mechanisms: lessons from yeast. *Nat Rev Mol Cell Bio* 10:458–467
3. Bigay J, Gounon P, Robineau S, Antony B (2003) Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. *Nature* 426:563–566
4. Nath S et al (2014) Lipidation of the LC3/GABARAP family of autophagy proteins relies

- on a membrane-curvature-sensing domain in Atg3. *Nat Cell Biol* 16:821–821
5. Ramamurthi KS, Lecuyer S, Stone HA, Losick R (2009) Geometric cue for protein localization in a bacterium. *Science* 323:1354–1357
 6. Lenarcic R et al (2009) Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J* 28:2272–2282
 7. Larsen JB et al (2015) Membrane curvature enables N-Ras lipid anchor sorting to liquid-ordered membrane phases. *Nat Chem Biol* 11:192–U176
 8. Gill RL et al (2015) Structural basis for the geometry-driven localization of a small protein. *Proc Natl Acad Sci USA* 112: E1908–E1915
 9. Fu R et al (2015) Spherical nanoparticle supported lipid bilayers for the structural study of membrane-curvature sensitive molecules. *J Am Chem Soc* 137:14031–14034
 10. Kallick DA, Tessmer MR, Watts CR, Li CY (1995) The use of dodecylphosphocholine micelles in solution NMR. *J Magn Reson B* 109:60–65
 11. Sanders CR, Hare BJ, Howard KP, Prestegard JH (1994) Magnetically-oriented phospholipid micelles as a tool for the study of membrane associated molecules. *Prog NMR Spectrosc* 26:421–444
 12. Nath A, Atkins WM, Sligar SG (2007) Application of phospholipid bilayer nanodiscs in the study of membrane and membrane proteins. *Biochemistry* 46:2059–2069
 13. Cruciani O et al (2006) An improved NMR study of liposomes using 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine as model. *Molecules* 11:334–344
 14. Larsen J, Hatzakis NS, Stamou D (2011) Observation of inhomogeneity in the lipid composition of individual nanoscale liposomes. *J Am Chem Soc* 133:10685–10687
 15. Walde P, Cosentino K, Engel H, Stano P (2010) Giant vesicles: preparations and applications. *Chembiochem* 11:848–865
 16. Bayerl TM, Bloom M (1990) Physical properties of single phospholipid bilayers adsorbed to micro glass beads. *Biophys J* 58:357–362
 17. Buranda T et al (2003) Biomimetic molecular assemblies on glass and mesoporous silica microbeads for biotechnology. *Langmuir* 19:1654–1663
 18. Savarala S, Ahmed S, Ilies MA, Wunder SL (2010) Formation and colloidal stability of DMPC supported lipid bilayers on SiO₂ nano-beads. *Langmuir* 26:12081–12088
 19. Mornet S, Lambert O, Duguet E, Brisson A (2005) The formation of supported lipid bilayers on silica nanoparticles revealed by cryoelectron microscopy. *Nano Lett* 5:281–285
 20. Naumann C, Brumm T, Bayerl TM (1992) Phase transition behavior of single phosphatidylcholine bilayers on a solid spherical support studied by DSC, NMR and FT-IR. *Biophys J* 63:1314–1319
 21. Dolainsky C, Unger M, Bloom M, Bayerl TM (1995) Two-dimensional exchange ²H NMR experiments of phospholipid bilayers on a spherical solid support. *Phys Rev E* 51:4743–4750
 22. Bunge A et al (2009) Characterization of lipid bilayers adsorbed on spherical LbL-support. *Soft Matter* 5:3331–3339
 23. Nordlund G, Lonneborg R, Brzezinski P (2009) Formation of supported lipid bilayers on silica particles studied using flow cytometry. *Langmuir* 25:4601–4606
 24. Baksh MM, Jaros M, Groves JT (2004) Detection of molecular interactions at membrane surfaces through colloid phase transitions. *Nature* 427:139–141
 25. Picard F, Paquet MJ, Dufourc EJ, Auger M (1998) Measurement of the lateral diffusion of dipalmitoylphosphatidylcholine adsorbed on silica beads in the absence and presence of melittin: A ³¹P two-dimensional exchange solid-state NMR study. *Biophys J* 74:857–868
 26. Tanaka M, Sackmann E (2005) Polymer-supported membranes as models of the cell surface. *Nature* 437:656–663
 27. Reintl HM, Bayerl TM (1993) Interaction of myelin basic protein with single bilayers on a solid support: an NMR, DSC and polarized infrared ATR study. *Biochim Biophys Acta* 1151:127–136
 28. Galneder R et al (2001) Microelectrophoresis of a bilayer-coated silica bead in an optical trap: application to enzymology. *Biophys J* 80:2298–2309
 29. Gopalakrishnan G, Rouiller I, Colman DR, Lennox RB (2009) Supported bilayers formed from different phospholipids on spherical silica substrates. *Langmuir* 25:5455–5458

Rapid Prediction of Multi-dimensional NMR Data Sets Using FANDAS

Siddarth Narasimhan, Deni Mance, Cecilia Pinto, Markus Weingarth, Alexandre M.J.J. Bonvin, and Marc Baldus

Abstract

Solid-state NMR (ssNMR) can provide structural information at the most detailed level and, at the same time, is applicable in highly heterogeneous and complex molecular environments. In the last few years, ssNMR has made significant progress in uncovering structure and dynamics of proteins in their native cellular environments [1–4]. Additionally, ssNMR has proven to be useful in studying large biomolecular complexes as well as membrane proteins at the atomic level [5]. In such studies, innovative labeling schemes have become a powerful approach to tackle spectral crowding. In fact, selecting the appropriate isotope-labeling schemes and a careful choice of the ssNMR experiments to be conducted are critical for applications of ssNMR in complex biomolecular systems. Previously, we have introduced a software tool called FANDAS (Fast Analysis of multidimensional NMR DAta Sets) that supports such investigations from the early stages of sample preparation to the final data analysis [6]. Here, we present a new version of FANDAS, called FANDAS 2.0, with improved user interface and extended labeling scheme options allowing the user to rapidly predict and analyze ssNMR data sets for a given protein-based application. It provides flexible options for advanced users to customize the program for tailored applications. In addition, the list of ssNMR experiments that can be predicted now includes proton (^1H) detected pulse sequences. FANDAS 2.0, written in Python, is freely available through a user-friendly web interface at <http://milou.science.uu.nl/services/FANDAS>.

Key words Biomolecular NMR, Labeling schemes, Spectral prediction, Spectral analysis and proton detection

1 Introduction

NMR represents a powerful tool for studying protein structure and dynamics. Thus, there is a growing need to make it more accessible to the community by providing analysis tools from an early level of the project to the final data analysis stage. We previously introduced a web application termed Fast Analysis of multidimensional NMR DAta Sets (FANDAS) [6] which aids in spectral analysis by producing peak lists for a variety of multidimensional solid-state NMR (ssNMR) experiments. FANDAS has the unique ability to

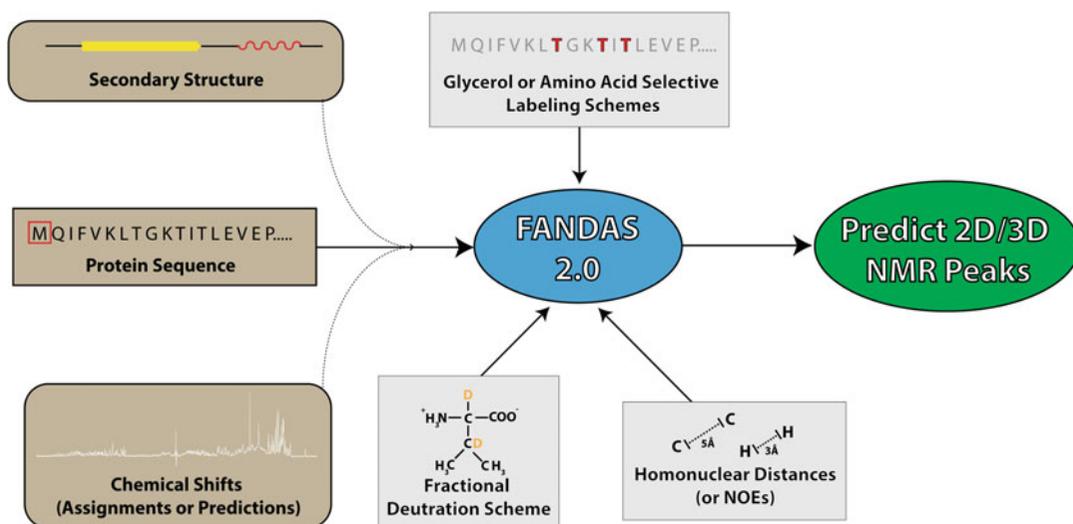


Fig. 1 Overall scheme of FANDAS, describing the different types of data used to refine the final prediction of peaks for multi-dimensional NMR experiments. The different types of inputs that FANDAS accepts are given in *brown boxes* and input modification options such as labeling schemes and distances are given in *grey boxes*

accommodate a variety of inputs and input modifications to accurately predict the peak pattern for given NMR experiments (Fig. 1). This property has been particularly useful in predicting NMR peaks in multidimensional NMR experiments and has, for example, enabled rapid identification of the most suitable labeling schemes even for studies of proteins in native cellular environments [3–5, 7–9].

Here, we present an improved version of FANDAS, called “FANDAS 2.0” (simply referred to as FANDAS in this chapter). FANDAS is now built on Python, and extended to a web interface using the Flask micro-framework (<http://flask.pocoo.org>). Therefore, vast improvements could be done to the user interface by introducing a user-friendly web-form that integrates all inputs at once. Another major improvement aids in choosing labeling schemes by offering improved choices for forward labeling (*see*, ref. 10 for a further discussion of forward and reverse isotope labeling) such as specific ^{13}C -only or ^{15}N -only labeling which were not available before. These new options speed up the process of predicting NMR correlation data for different forward labeling schemes, in particular with respect to inter-residue sequential correlation experiments such as N-Co-Cx [11]. Additionally, it is now possible to predict correlations for proton-detected ssNMR experiments that are of increasing use in biological ssNMR [12, 13]. Finally, additional features have been implemented, which will be discussed in detail in the following sections. For the users with command line experience, we provide a standalone python application (<https://github.com/siddarthnarasimhan/>

FANDAS_2.0), which is also the engine for the webserver and can be run on a local personal computer. This version provides more flexibility to the user allowing customization of the program for advanced studies such as those involving unique or unnatural amino acids, or to extend the program to predict additional ssNMR experiments currently not implemented in FANDAS 2.0.

This chapter provides protocols describing the use of both the webserver and command line versions of FANDAS. In the following sections, a step-by-step description is given for each feature in FANDAS, using a sample protein as an example.

2 Materials and Prior Knowledge

2.1 Introduction to SPARKY Peak Lists

FANDAS has been geared toward users of the NMR data analysis program, SPARKY [14]. Thus, a prior knowledge of SPARKY is highly recommended. SPARKY allows for easy NMR spectral visualization, assignment, and analysis. The peaks identified in a spectrum are usually stored as space-delimited plain text files known as peak lists. Each line in the peak file, corresponding to a single peak, consists of three elements: (a) The peak label that appears as an annotation on the spectrum, (b) the peak coordinates that represent chemical shifts of the correlating nuclei, and (c) the notes section, where the user is free to add any notes regarding the peak (Fig. 2). The output of FANDAS is composed of peak files that the user can then superimpose on the spectra.

A) Peak Labels				C) Notes Section			
???	120.79	54.4	M1N-M1CA	??-?-?	120.19	55.12	175.93 M1N-M1CA-M1C
???	125.08	54.9	Q2N-Q2CA	??-?-?	120.19	55.12	32.93 M1N-M1CA-M1CB
???	116.08	59.5	I3N-I3CA	??-?-?	120.19	55.12	175.01 M1N-M1CA-Q2C
???	118.58	54.6	F4N-F4CA	??-?-?	120.19	55.12	32.45 M1N-M1CA-Q2CB
???	127.98	60.2	I13N-I13CA	??-?-?	123.52	55.55	175.01 Q2N-Q2CA-Q2C
B) Peak Coordinates							

Fig. 2 Description of SPARKY 2D (*left*) and 3D (*right*) peak lists generated using FANDAS. (a) The peak labels, which would appear on the spectrum when the peak list is loaded. The default “? -?” for 2D peak list and “?-?-?” for 3D peak list imply that the peaks do not have any labels. (b) Peak coordinates, which are essentially the chemical shifts of the correlating nuclei. (c) The notes section- notes do not appear on the spectrum, but can be accessed too in the peak list. By default, FANDAS stores the peak information in the notes section

2.2 FANDAS Webserver

The FANDAS webserver has been made user friendly to enable its use in the wider scientific community. Thus, no prior computational knowledge is required to be able to operate it and there are no restrictions on the operating system or the browser. However, we recommend usage of any version of Google Chrome or Mozilla Firefox. The server is freely available at: <http://milou.science.uu.nl/services/FANDAS>.

2.3 FANDAS Local Installation

To use a local version of FANDAS, the user is required to have at least some elementary knowledge on working with the command line (bash environment is recommended) like navigating through folders and opening text files using editors. Thus, a basic knowledge of using text editors like vim or nano is recommended (alternatively GUI-based text editors like Notepad + or TextEdit could also be used). FANDAS has been created on a computer running MacOS X El Capitan (10.11) with Anaconda 2.4.1 running Python 2.7.12. However, FANDAS can be readily run on a different platform, provided that the following computational requirements are met:

1. Windows/Mac/Linux operating systems that can preferably run a BASH shell environment (the protocol we describe here uses this environment, but the user is free to use something else).
2. Python 2.7 (<https://www.python.org/downloads/>) or higher with pip (<https://pypi.python.org/pypi/pip>) to install python packages and the python package numpy (preferably 1.11.2) (<https://www.scipy.org/scipylib/download.html>).
3. git (<https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>): to clone the FANDAS package onto the local desktop.
4. DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>) or STRIDE (<http://webclu.bio.wzw.tum.de/stride/>) for secondary structure assignments (Optional).
5. SHIFTX2 (<http://www.shiftx2.ca/>) for estimating chemical shifts from existing structural models (Optional).

Detailed instructions to install the above (required) packages on your operating system of choice are given in the **readme.md** on the GitHub repository https://github.com/siddarthnarasimhan/FANDAS_2.0.

3 Methods

This section presents a general tutorial, with instructions to use the FANDAS webserver and the command line version of FANDAS. For this purpose, a small post translational modifier protein—Ubiquitin,

has been used as a test case. The test dataset used for this demonstration (**test_dataset.tar**) can be downloaded from the GitHub repository: https://github.com/siddarthnarasimhan/FANDAS_2.0. This demonstration will summarize a typical FANDAS workflow and give the user a glimpse of all the features that the program offers.

3.1 Using the FANDAS Webserver

To run the FANDAS webserver, open a web browser of your choice and navigate to: <http://milou.science.uu.nl/services/FANDAS/>. The webpage consists of a web form that is divided into blocks, containing fields corresponding to different types of input. Instructions to fill in particular fields are mentioned in the webpage and will be elucidated below in a step-by-step manner.

1. **Input Sequence and Secondary Structure Block:** This block contains three fields of which two are mandatory:
 - (a) **Project Name (mandatory):** The user may choose any arbitrary name with more than four letters or numbers. No special characters (except underscore, “_”) or spaces are permitted.
 - (b) **Protein Sequence (mandatory):** The input sequence must be in single letter amino-acid code (not case sensitive). To this end, inputs are restricted to the 20 naturally occurring amino acids and any other input values will be changed to “A,” i.e., alanine. If the user uses a raw FASTA sequence, all lines that do not contain the sequence need to be removed.
 - (c) **Secondary Structure:** The secondary structure assignments will be used to accurately assign chemical shifts to the backbone heavy atoms and protons by using pre-determined average chemical shift statistics for every amino acid existing in different distinct conformations [15]. If the protein structure is available, it is possible to obtain these assignments by using programs such as DSSP [16] (<http://swift.cmbi.ru.nl/gv/dssp/>) or STRIDE [17] (<http://webclu.bio.wzw.tum.de/stride/>) and if the structure is not known, prediction tools such as JPRED4 [18] (<http://www.compbio.dundee.ac.uk/jpred/>) or PSIPRED [19] (<http://bioinf.cs.ucl.ac.uk/psipred/>) may be used. The user must ensure that the secondary structure assignments/predictions are simplified to the assignments that are permitted, namely: alpha-helix “a,” beta-sheet “b,” and random coil “c” in addition to NMR calculated averages “n” which is assigned to the protein by default. In the example shown in Fig. 3, the secondary structure assignments were done using command line version of STRIDE on the PDB structure 1UBQ and

 >> INPUT SEQUENCE AND SECONDARY STRUCTURE

Project Name*: No spaces or special characters (except "_") → Field 1

Protein Sequence (single letter amino acid code)*

MQIFVKLTGKITLEVEPSDTIENVKAKIQDKEGIPDPQQLRIFAGKQLEDGRTLSDYNIQKESTLHLVLRLL
 RGG

 → Field 2

cbbbbcccccbbbbcccccaaaaaaaaaaaccaabbbbbcbbcccccaaacccccbbbbbcccc

Fig. 3 Example input of the “Input Sequence and Secondary Structure” block

simplified to **a**, **b**, and **c**. If the secondary structure assignments are not available for a part of the protein, it is recommended that the user fills in “**c**” or “**n**” at the appropriate sequence sites. For example, if for the sequence “**APAPMLQSMVSLQLSLV**” the secondary structure of the first four residues is not known, the secondary structure input could be “**nnnn**” or “**cccc**” followed by the assignments for the remaining residues e.g.,: **nnnnaaaaaaaaaaaaa**.

An example of a completed form is given in Fig. 3.

2. **Input Chemical Shifts as BMRB Table Block:** FANDAS also accepts user defined chemical shift assignments that would be used in lieu of the default assignments discussed in the previous section. In the absence of chemical-shift assignments, it is possible to generate (predict) chemical shifts on the basis of a 3D protein structure to a reasonable level of accuracy for every atom using programs such as SHIFTX2 [20] (<http://www.shiftx2.ca/cgi-bin/shiftx2.cgi>). The following are the descriptions of the fields present in this input block:

- (a) **BMRB Tables:** It is recommended for the assignments to be in a tabular NMR-STAR format (for a detailed description, see <http://www.bmrwisc.edu/formats.html>). FANDAS also accepts any table-like text file with rows (lines) corresponding to each nucleus, with columns consisting of atom name, residue number, and chemical shift (Fig. 4). In the example (Fig. 5), the command line version of SHIFTX2 was used to generate the BMRB tables using the crystal structure of Ubiquitin (1UBQ).

1	C	170.52
1	CA	54.40
1	N	120.79
2	C	175.84
2	CA	54.90
2	CB	30.80
2	N	125.08
↑	↑	↑
Residue Numbers	Atom Names	Chemical Shifts

Fig. 4 An example of BMRB tables that is accepted in FANDAS

>> INPUT CHEMICAL SHIFTS AS BMRB TABLES

Recommended format- space delimited NMR-STAR (any version)

Hint: You can use packages such as SHIFTX2 to predict chemical shifts for a given PDB structure

Format example: 1 1 TRP HD1 H 7.33 0.01 1

1 M C 170.5200
1 M CA 54.4000
1 M CB 33.4599
1 M CE 17.6847
1 M CG 31.8754
1 M H 8.2666
1 M HA 4.2200
1 M HB2 2.0700
1 M HB3 2.1390
1 M HE 1.8684

→ Field 1

Provide column number (starting with 1) for:

Residue number: Atom name: Chemical shift: → Field 2

Provide the residue number for the first entry (if the residue numbers don't match the input sequence): → Field 3

Fig. 5 Filled out sample of “Input Chemical Shifts as BMRB Tables” block

- (b) **Provide column numbers for residue number, atom name, and chemical shift:** The user is required to provide the column numbers for residue number, atom name, and chemical shift corresponding to the tables input in the previous field.
- (c) **Provide the residue number for the first entry:** This field should be left blank if the residue numbers in the BMRB tables match the residue numbers in the input sequence. If not, the residue number for the first entry (line) in the assignment table is to be entered in this field. This value is used to offset all residue numbers in the table to match the sequence.

3. **Amino-Acid Selective Labeling Schemes Block:** By default, it is assumed that the protein of interest is fully isotope labeled (i.e., 100% enriched at the ^{13}C and ^{15}N positions). The options provided in this section allow for the incorporation of amino-acid selective forward or reverse labeling schemes (Fig. 6). If forward or reverse labeling schemes are used, the user must specify a list of amino acids for the chosen labeling scheme. This section is particularly useful for assessing the result of different labeling schemes on the resulting ssNMR spectra. A typical application for features in this section is discussed section 4.
4. **Glycerol Labeling Schemes and Fractional Deuteration:** To reduce spectral crowding among the various carbon positions, glycerol-based labeling schemes have been introduced [21–23]. Such schemes have been implemented in FANDAS and choosing any of the glycerol labeling schemes would disable amino acid selective (forward or reverse) labeling schemes or fully labeled schemes. Additionally, this section is set up to include fractional deuteration (dashed box in Fig. 7) scheme, which is intended to fully deuterate $^{13}\text{C}\alpha$ nuclei as well as specific side chain carbons [13, 24].

>> AMINO ACID SELECTIVE LABELLING SCHEMES

Select a labelling scheme

Reverse labelling scheme (would remove the amino acids entered below)

^{12}C & ^{14}N - List: ^{12}C - List: ^{14}N - List:

Forward labelling scheme (would label only the amino acids entered below)

^{13}C & ^{15}N - List: ^{13}C - List: ^{15}N - List:

Fully labelled (Default)

Fig. 6 “Amino-Acid Selective Labeling Schemes” block consists of options to choose the appropriate labeling scheme and a provision to enter the desired labeled amino acids in double or single labeled forms

>> GLYCEROL LABELLING SCHEMES & FRACTIONAL DEUTERATION

NOTE: Glycerol labelling cannot be coupled with amino acid selective labelling schemes, choosing the former would disable the latter and vice-versa

1,3-Glycerol 2-Glycerol [Follow this link for description](#)

Fully labelled (Default)

Fractionally Deuterated [Follow this link for description](#)

Fig. 7 Highlighting how different labeling schemes are handled in FANDAS and the option for fractional deuteration (green box)

>> DISTANCE LIST BETWEEN HOMONUCLEAR (H-H OR C-C) PAIRS

Hint: You can use our Python Script to create such a list from a PDB

Format syntax: resi_num_1, atm_name_1, resi_num_2, atm_name_2, dist (in Å)

Format example: 1, CA, 3, CB, 15

```
75,C,75,CA,1.516
75,C,76,CA,2.473
75,C,76,C,3.713
76,CA,75,CA,3.794
76,CA,75,C,2.473
76,CA,76,C,1.537
76,C,75,CA,4.762
76,C,75,C,3.713
76,C,76,CA,1.537
```

Distance cut-off (in Å): 5

CC- Spin Diffusion Distance Edited Cutoff: 5Å					CC- Spin Diffusion Distance Edited Cutoff: 2Å				
?-?	54.4	170.52	M1CA-M1C_1.548		?-?	54.4	170.52	M1CA-M1C_1.548	
?-?	170.52	54.4	M1C-M1CA_1.548		?-?	170.52	54.4	M1C-M1CA_1.548	
?-?	54.4	33.46	M1CA-M1CB_1.506		?-?	54.4	33.46	M1CA-M1CB_1.506	
?-?	33.46	54.4	M1CB-M1CA_1.506		?-?	33.46	54.4	M1CB-M1CA_1.506	
?-?	54.4	31.88	M1CA-M1CG_2.528		?-?	170.52	54.4	M1C-M1CA_1.548	
?-?	31.88	54.4	M1CG-M1CA_2.528		?-?	54.4	170.52	M1CA-M1C_1.548	
?-?	54.4	54.9	M1CA-Q2CA_3.804		?-?	33.46	54.4	M1CB-M1CA_1.506	
?-?	54.9	54.4	Q2CA-M1CA_3.804		?-?	54.4	33.46	M1CA-M1CB_1.506	
?-?	54.4	175.84	M1CA-Q2C_4.51		?-?	33.46	31.88	M1CB-M1CG_1.505	
?-?	175.84	54.4	Q2C-M1CA_4.51		?-?	31.88	33.46	M1CG-M1CB_1.505	

Fig. 8 Example outputs (*dashed boxes*) suited for the analysis of distance-dependent CC correlation experiments (such as protein-driven spin diffusion) for two different cut-off values. Distance lists were calculated using the crystal structure of Ubiquitin (PDB: 1UBQ)

5. Distance List Between Homonuclear (H-H or C-C) Pairs

Block: Predicting peaks for distance-edited experiments requires a list of distances between H-H or C-C groups depending on the experiment. These distances can essentially be through-space distance restraints obtained previously from NMR experiments or they can be generated from PDB coordinates. To generate distances from PDB coordinates, we have generated a dedicated python script “[distance_calculator.py](#)” which can be accessed from the GitHub repository: https://github.com/siddarthnarasimhan/FANDAS_2.0. Along with the distances provided, the user must provide a cutoff distance (in Å). To exemplify such a distance cutoff, a sample output for a distance edited CC Spin diffusion experiment is shown for two cut-off values in (Fig. 8).

6. Predict Peaks for NMR Experiments (SPARKY Format)

Block: This is the final section of FANDAS webserver where the experiments for which predictions are to be made, are listed. This section consists of a list of experiments that the user can simply select using the appropriate check boxes (Fig. 9).

- Peak Labels: Along with the list of experiments, this section features an option to turn on peak labels (highlighted with a dashed box in Fig. 9). As described in Subheading 2.1, the default peak labels are null: “?-?” “?-?-?”. By turning on the “Peak Labels” option, the peak list would now be incorporated with the labels. The labels would inform the user of the nuclei involved in the correlation observed (Fig. 10).

>> PREDICT PEAKS FOR NMR EXPERIMENTS (SPARKY FORMAT)

On Off

Offset residue numbers in the peak file to start with residue number

Atoms within the brackets represent that they are not seen in the spectrum

2D NMR Experiments

<input checked="" type="checkbox"/> N-H	<input type="checkbox"/> H-N	<input type="checkbox"/> C-H
<input type="checkbox"/> H-C	<input type="checkbox"/> H-H Spin Diffusion	<input type="checkbox"/> C-C DQ-SQ Correlation
<input type="checkbox"/> C-C Spin Diff. intra residue	<input type="checkbox"/> C-C Spin Diff. (residues i, i+1 & i-1)	<input type="checkbox"/> N-Co
<input type="checkbox"/> N-Co	<input type="checkbox"/> N-(Ca)-Cx	<input type="checkbox"/> N-(Ca)-Cx (residues i, i+1 & i-1)
<input type="checkbox"/> N-(Co)-CaCb	<input type="checkbox"/> N-(Co)-Cx	<input type="checkbox"/> Ca-(N)-H
<input type="checkbox"/> Co-(N)-H	<input type="checkbox"/> Ca-(Co)-(N)-H	<input type="checkbox"/> Co-(Ca)-(N)-H
<input type="checkbox"/> N-(Ca)-H		

3D NMR Experiments

<input checked="" type="checkbox"/> N-Ca-Cx	<input type="checkbox"/> N-Ca-Cx (residues i, i+1 & i-1)	<input type="checkbox"/> N-Co-Cx
<input type="checkbox"/> N-Co-CaCb	<input type="checkbox"/> SQSQSQ (residues i, i+1 & i-1)	<input type="checkbox"/> DQSQSQ intra residue
<input type="checkbox"/> DQSQSQ (residues i, i+1 & i-1)	<input type="checkbox"/> Ca-N-H	<input type="checkbox"/> Co-N-H
<input type="checkbox"/> Ca-(Co)-N-H	<input type="checkbox"/> Co-(Ca)-N-H	<input type="checkbox"/> N-Ca-Ha

2D NMR Experiments (Distance Encoded)

<input checked="" type="checkbox"/> C-C Spin Diffusion	<input type="checkbox"/> H-H Spin Diffusion	<input type="checkbox"/> C-(HH)-C
<input type="checkbox"/> N-(HH)-C	<input type="checkbox"/> C-(H)-H	<input type="checkbox"/> N-(H)-H
<input type="checkbox"/> N-(Ca)-Cx	<input type="checkbox"/> N-(Co)-Cx	

3D NMR Experiments (Distance Encoded)

<input checked="" type="checkbox"/> N-Ca-Cx	<input type="checkbox"/> N-Co-Cx	<input type="checkbox"/> C-H-H
<input type="checkbox"/> N-H-H		

Submit Reset

Results will be stored for upto a day

Fig. 9 The final input block, where the user can select experiments for which predictions are to be made, specify if peak labels are required and if the residue numbers need to be offset (highlighted using *dashed boxes*)

- (b) Residue number offset: This field (highlighted with a dashed box in Fig. 9) takes in numerical values to offset the residue number in the final peak lists. If the user wishes that the residue numbers start from a different number than one, it can be specified here. This is particularly useful if the prediction is being done for a specific segment or a domain and not the whole protein, wherein the starting residue number may not be one.

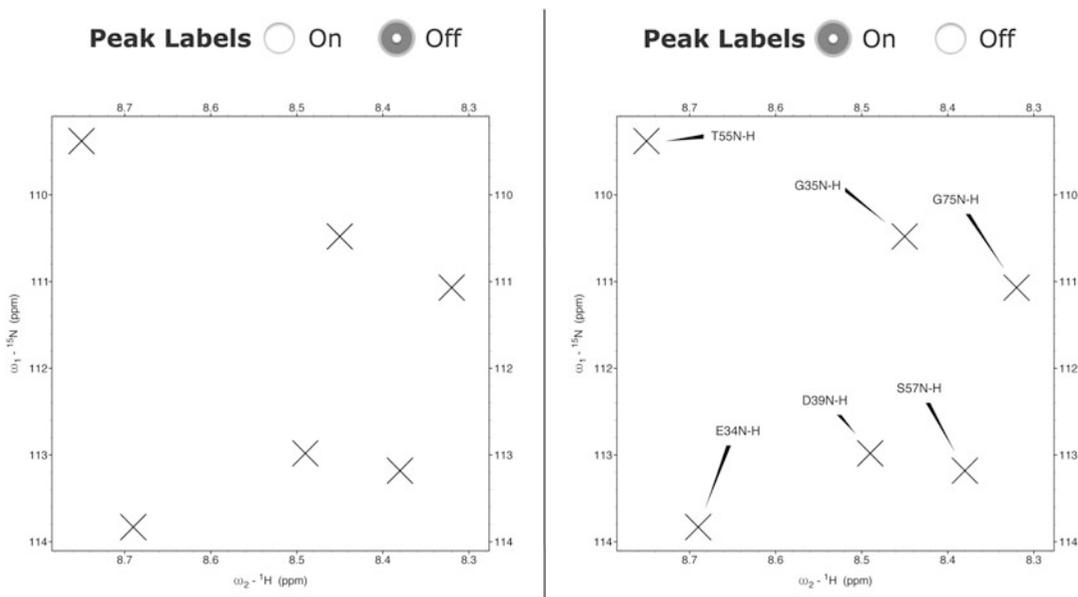


Fig. 10 Sample peak outputs overlaid on empty N-H Spectra in SPARKY. The left panel shows the default output when Peak Labels are inactive and the right panel shows the output when Peak Labels are turned on

HERE YOU GO!

>> RESULTS

Note the URL in your browser and access the following files for the next 24 hours

- Ubiquitin_Demo-all-predictions.zip → If more than one experiment is chosen, a zip file containing all the predictions would be made available
- Ubiquitin_Demo-hn.txt
- Ubiquitin_Demo-ncacx_inter_3d.txt

Fig. 11 Prediction results as seen on the results page

7. **Accessing the Results:** After the form is completed, the job would be submitted to the server and the user will be redirected to the results page where the predictions will be available for download (Fig. 11). The user can access the peak files by simply clicking on the appropriate file that is listed. Currently, results are stored for up to one day on the server and can be accessed by simply navigating to the URL of the results page that is shown in the browser window.

3.2 Running the FANDAS Application Locally

The features of this version are the same as the webserver, since the webserver uses this python application as an engine. Thus, this section will explain how to clone the application along with the supporting files onto a local computer and how to use the interface rather than discuss-specific features. Every command line input and output is shown in a black background; input lines begin with a \$ sign.

1. **Cloning the git repository onto the local computer:** For a quick introduction to git, we refer the interested reader to <https://git-scm.com/book/en/v2/Getting-Started-Git-Basics>. To start with, open a BASH environment with a terminal emulator of your choice and type the following:

```
$ git clone https://github.com/siddarthnarasimhan/FANDAS_2.0
```

The user should see the following output if the git cloning was successful (if not, please check if git has been installed):

```
Cloning into 'FANDAS_2.0'...
remote: Counting objects: 7, done.
remote: Compressing objects: 100% (6/6), done.
remote: Total 7 (delta 0), reused 7 (delta 0), pack-reused 0
Unpacking objects: 100% (7/7), done.
```

Navigate to the FANDAS_2.0 folder and list the files present. All files are required and the demo folder (ubiquitin_demo) is shown below:

```
distance_calculator.py fandas.py readme.md standard.dat test_
dataset.tar
```

2. **Description of the contents in the FANDAS_2.0 Package:**
 - (a) **distance_calculator.py:** This script is used to calculate the distance between homonuclear pairs from the supplied PDB coordinates.
 - (b) **fandas.py:** This is the FANDAS_2.0 application.
 - (c) **README.md:** A github mandated readme file for the program.
 - (d) **standard.dat:** A file containing the standard chemical-shift tables for all amino acids.

- (e) **test_dataset.tar**: Contains a test dataset that has been used to demonstrate the use of FANDAS in this manual. It contains a readme file named “**readme.txt**” that describes all the files in the folder. In the following sections, the files in this folder would be referred to for demonstration.

3. **Testing the application and the help flag**: This is important, as the user would already know if all the modules (*see* **readme.md** file for more information) necessary to run FANDAS are available. Besides, considering the number of arguments that FANDAS accepts, the user may have to invoke the help feature each time the program is used, to ensure that arguments are used correctly. Execute the script with the help “-h” flag:

```
$ python fandas.py -h
```

If the usage message followed by a list of arguments is printed (shown in the next section), then there are no compilation errors and thus the application is ready to use.

4. **Usage message and description of the arguments**: The usage message essentially lists all arguments that FANDAS takes as flags. The flags that have been listed within square brackets are optional arguments. The ones that are not listed within square brackets are the obligatory arguments, which in our case are only -i that corresponds to the input sequence:

```
usage: fandas.py [-h] -i I [-ss SS] [-bt BT] [-btc BTC BTC BTC] [-ls LS]
               [-dl DL [DL ...]] [-cl CL [CL ...]] [-nl NL [NL ...]] [-fd]
               [-exp_2d EXP_2D [EXP_2D ...]]
               [-exp_2dd EXP_2DD [EXP_2DD ...]]
               [-exp_3d EXP_3D [EXP_3D ...]]
               [-exp_3dd EXP_3DD [EXP_3DD ...]] [-sl]
               [-dlist DLIST [DLIST ...]] [-dlim DLIM] [-o O]
```

When the help option is invoked, a description of each argument is printed below the usage message (shown below). The description for each argument needs to be read carefully. Inputs for sequence, secondary structures, BMRB tables, and distance lists are to be supplied as plain text files.

```

optional arguments:
-h, --help          show this help message and exit
-i I                input sequence as a text file (REQUIRED)
-ss SS              secondary structures as a text file: 'a'- alpha helix,
                   'b'- beta sheet & 'c'- random coil; if unspecified,
                   will use BMRB averages: 'n'
-bt BT              BMRB tables as a text file in NMR Star format
-btc BTC BTC BTC   column numbers for residue number, atom name and
                   chemical shifts in the BMRB tables
-ls LS              labelling scheme. Default is Fully Labelled. Other
                   options: fw = Forward Labelled, rv = Reverse Labelled,
                   gl1= 1,3 Glycerol Labelling, gl2= 2 Glycerol
                   Labelling
-dl DL [DL ...]    list of 13C & 15N (for forward labelling) or 12C & 14N
                   (for reverse labelling) amino acids
-cl CL [CL ...]    list of 13C (for forward labelling) or 12C (for
                   reverse labelling) amino acids
-nl NL [NL ...]    list of 15N (for forward labelling) or 14N (for
                   reverse labelling) amino acids
-fd                fractionally deuterated, if you use this flag, it
                   would be implemented automatically
-exp_2d EXP_2D [EXP_2D ...]
                   list of 2D experiments: ['NH', 'HN', 'CH', 'HC', 'HH',
                   'DQSQ', 'CC_SPINDIFF_INTRA', 'CC_SPINDIFF_INTER',
                   'NCA', 'NCO', 'NCACX', 'NCACX_INTER', 'NCOCX',
                   'NCOCA_CB', 'CANH', 'CONH', 'CACONH', 'COCANH',
                   'NCAH']
-exp_2dd EXP_2DD [EXP_2DD ...]
                   list of distance encoded (distance list, -dl and limit
                   -dlm must be provided) 2D experiments: ['CC_SPINDIFF',
                   'HH', 'CHHC', 'NHHC', 'CHH', 'NHH', 'HHC', 'NCACX',
                   'NCOCX']
-exp_3d EXP_3D [EXP_3D ...]
                   list of 3D experiments: ['NCACX', 'NCACX_INTER',
                   'NCOCX', 'NCOCA_CB', 'SQSQSQ_INTER', 'DQSQSQ_INTRA',
                   'DQSQSQ_INTER', 'CANH', 'CONH', 'CACONH', 'COCANH',
                   'NCAH']
-exp_3dd EXP_3DD [EXP_3DD ...]
                   list of distance encoded (distance list, -dl and limit
                   -dlm must be provided) 3D experiments: ['CHH', 'NHH',
                   'NCACX', 'NCOCX']
-sl                automatically assign peak labels in the sparky file
-dlist DLIST [DLIST ...]
                   distance list (as a file) in angstroms as follows:
                   "resi_num,atm_nam,resi_num_2,atm_nam_2,dist".
                   EXAMPLE:"2,CA,4,CB,4.5"
-dlim DLIM         distance limit in angstroms, default: 5 Angstrom
-o 0                names for output, default: "fandas_output"

```

5. **Preparation of the input files:** To prepare the input files, we recommend creating a working directory to store all input text files such as sequence, secondary structure assignments, BMRB tables, and the distance lists. The files in the “**test_dataset.tar**” are used as an example below. The choice of text editor is not important, provided that the file created is a plain text file.
6. **Making peak predictions:** For easy use, it is recommended to create an alias for the path of the “**fandas.py**” script in the “**.bashrc**” or “**.bash_profile**” file so that the script is globally executable. To make the predictions, navigate to the working directory and type the following command (The description of each argument in the command input is given in Table 1):

```
$ python fandas.py -i lubq-seq.txt -ss lubq-ss.txt -bt lubq-bmrbs-tables.txt -btc 1 3 4 -dlist lubq-dist.txt -dlim 5 -exp_2d hn -exp_2dd chh -exp_3d ncacx -exp_3dd ncacx -sl
```

7. **Guide to using labeling schemes:** By default, the protein is assumed to be fully ^{13}C & ^{15}N labeled unless arguments defining the labeling schemes are specified. Alternative labeling schemes available are shown in the table below (Table 2).

Table 1
Description of all parameters in a sample FANDAS input

Flag + Argument(s)	Description
-i lubq-seq.txt	input file (or path) is “ lubq-seq.txt ”
-ss lubq-ss.txt	file (or path) containing secondary structure is “ lubq-ss.txt ”
-bt lubq-bmrbs-tables.txt	file (or path) containing the BMRB tables is “ lubq-bmrbs-tables.txt ”
-btc 1 3 4	column indices for residue number, atom name & the chemical shift in the BMRB tables are 1, 3 & 4
-dlist lubq-dist.txt	file (or path) containing the distance list is “ lubq-dist.txt ”
-dlim 5	distance limit between nuclei for them to be treated as neighbors for predicting peaks in distance-encoding NMR experiments
-exp_2d cc_spindiff_intra	predict peaks for a 2D intra residue CC spin diffusion experiment
-exp_2dd chh	predict peaks for a 2D distance-edited CHH experiment
-exp_3d ncacx	predict peaks for 3D N-Ca-Cx spectrum
-exp_3dd ncacx	predict peaks for 3D distance edited N-Ca-Cx spectrum
-sl	assign peak labels that would be visible on the spectrum in SPARKY

Table 2
Alternative and supplementary labeling schemes to the default—uniformly ^{13}C & ^{15}N default labeling scheme

Flag + Argument(s)	Description
-ls <labeling scheme>	labeling scheme to be used; when left empty, the protein would be assumed to be fully ^{13}C and ^{15}N labeled
-dl <residue list>	list of forward ^{13}C + ^{15}N labeled amino acids or reverse ^{12}C + ^{14}N (unlabeled) amino acids as per the chosen labeling scheme (-Is)
-cl <residue list>	list of forward ^{13}C labeled or reverse ^{12}C (unlabeled) amino acids as per the chosen labeling scheme (-Is)
-nl <residue list>	list of forward ^{15}N labeled or reverse ^{14}N (unlabeled) amino acids as per the chosen labeling scheme (-Is)
-fd	include fractional deuteration; this can be combined with other labeling schemes

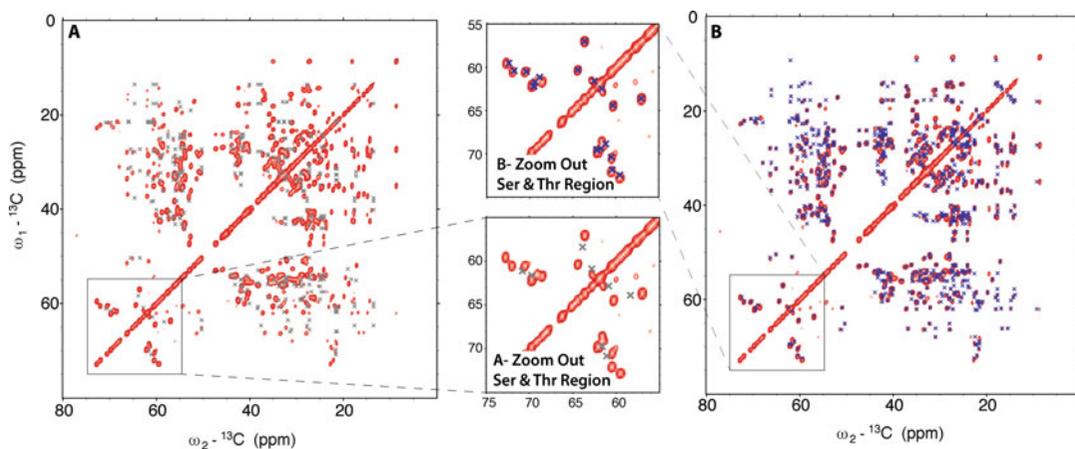


Fig. 12 FANDAS Predictions of CC (PARIS) experiment [25] when only secondary structures are supplied (Panel A with the *gray peaks*) and in an information-rich regime where chemical shifts of all atoms were predicted from the crystal structure using SHIFTX2 (Panel B with *blue peaks*). In the zoom-in of the serine and threonine region, one can clearly observe that the prediction accuracy is greatly improved when more information is supplied to FANDAS

4 Analysis and Case Study

As mentioned in the introduction section, FANDAS operates by integrating a variety of inputs to predict the peaks occurring in different experiments. Sample outputs for a CC Spin diffusion experiment produced by FANDAS in a data rich and data deprived cases are shown in Fig. 14. Thus, indicating that the quality of the predictions entirely relies on the quality of the input provided. Even when operating in a low information regime, it is possible to get a

substantial amount of preliminary information for a FANDAS-based spectral analysis (Fig. 12).

To highlight some of the areas where FANDAS could be useful besides spectral analysis, the following case studies are described.

1. **Case Study- Choosing Amino Acid Selective Labeling Schemes using FANDAS:** This is one of the most insightful features that FANDAS offers, particularly at the early stages of an NMR study. Tailored amino-acid selective labeling schemes can drastically reduce spectral crowding and allow the user to focus on selected protein regions. This section demonstrates how FANDAS can be used to rapidly assess the effect of changing labeling schemes. To illustrate the output generated for the different labeling schemes, two 2D experiments have been selected:
 - (a) **N-Ca intra-residue correlations** that probe protein residues that are both ^{13}C and ^{15}N labeled.
 - (b) **N-Co inter-residue correlations** that probe the polarization transfer between the backbone nitrogen atom of the (i)th residue to the carboxyl carbon atom of the (i-1)th residue. This requires the (i)th residue to be at least ^{15}N labeled and the (i-1)th residue to be at least ^{13}C labeled.

Case 1: Fully Labeled (Default): *The default option (Fig. 13a) assumes that all the residues are fully labeled. The output for such a labeling scheme would contain all peaks that could possibly occur in each experiment (Fig. 13b).*

Case 2: ^{13}C & ^{15}N Labeled: *If specific residue types are ^{13}C & ^{15}N labeled, FANDAS retains the ^{13}C & ^{15}N (forward labeled Fig. 14a) chemical shifts or removes them and retains the remaining amino acids (reverse labeled Fig. 14b). When a forward labeling scheme of this nature is used, the spectral crowding is vastly reduced in both N-Co and N-Ca spectra, and sequential correlations can be observed at specific sites as shown in Fig. 14c.*

Case 3: Using ^{13}C only labeled amino acids and ^{15}N only labeled amino acids: *There exist labeling schemes where either ^{13}C or ^{15}N amino acids are labeled in combination to probe site selective sequential correlations on proteins (see, refs. 3, 4). FANDAS treats inputs for this labeling scheme in a similar fashion to the previous case. If residues are ^{13}C or ^{15}N labeled, FANDAS either retains only the ^{13}C or ^{15}N (forward labeled Fig. 15a) chemical shifts or removes them and retains the remaining amino acids (reverse labeled Fig. 15b) as shown in Fig. 15c.*

A

>> AMINO ACID SELECTIVE LABELLING SCHEMES

Select a labelling scheme

- Reverse labelling scheme (would remove the amino acids entered below)
 - Forward labelling scheme (would label only the amino acids entered below)
 - Fully labelled (Default)
- ¹²C & ¹⁴N- List: ¹²C- List: ¹⁴N- List: Separate the amino acids by a space
- ¹³C & ¹⁵N- List: ¹³C- List: ¹⁵N- List: Separate the amino acids by a space

B

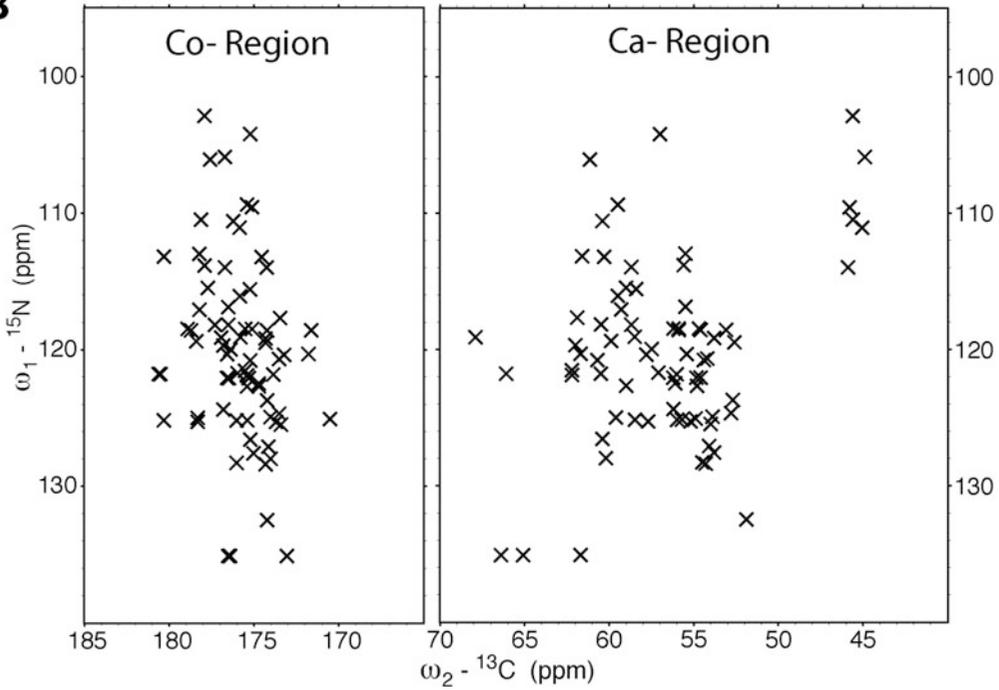


Fig. 13 The default fully labeled option and the resulting predictions for peaks in the N-Co and N-Ca spectra of Ubiquitin

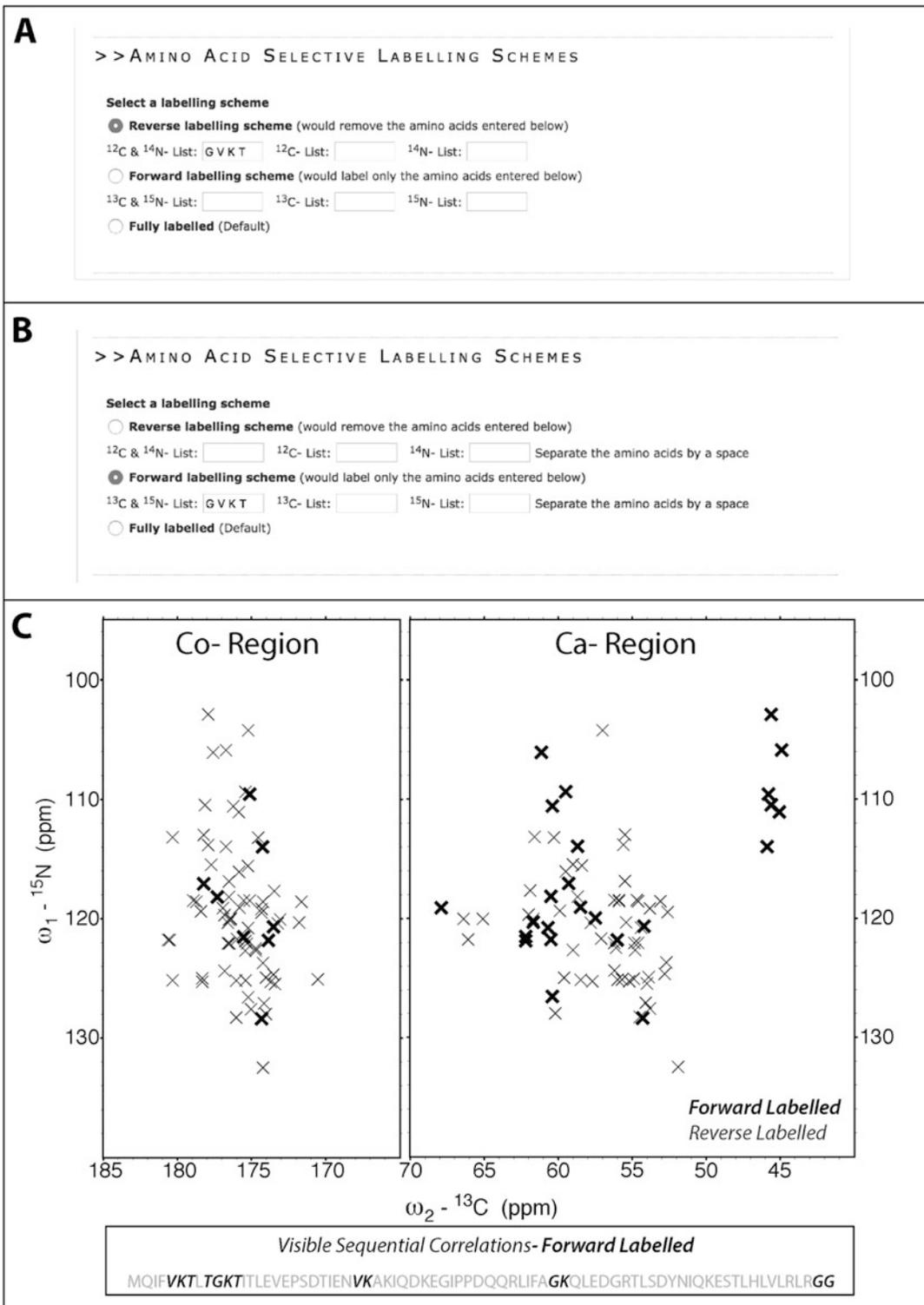


Fig. 14 Incorporating amino acid selective ¹³C + ¹⁵N labeling scheme in FANDAS for glycine, valine, lysine, and threonine residues. As can be seen in the peak prediction, the two labeling schemes complement each other

A

>> AMINO ACID SELECTIVE LABELLING SCHEMES

Select a labelling scheme

Reverse labelling scheme (would remove the amino acids entered below)

¹²C & ¹⁴N- List: ¹²C- List: RT ¹⁴N- List: LG Separate the amino acids by a space

Forward labelling scheme (would label only the amino acids entered below)

¹³C & ¹⁵N- List: ¹³C- List: ¹⁵N- List: Separate the amino acids by a space

Fully labelled (Default)

B

>> AMINO ACID SELECTIVE LABELLING SCHEMES

Select a labelling scheme

Reverse labelling scheme (would remove the amino acids entered below)

¹²C & ¹⁴N- List: ¹²C- List: ¹⁴N- List: Separate the amino acids by a space

Forward labelling scheme (would label only the amino acids entered below)

¹³C & ¹⁵N- List: ¹³C- List: RT ¹⁵N- List: LG Separate the amino acids by a space

Fully labelled (Default)

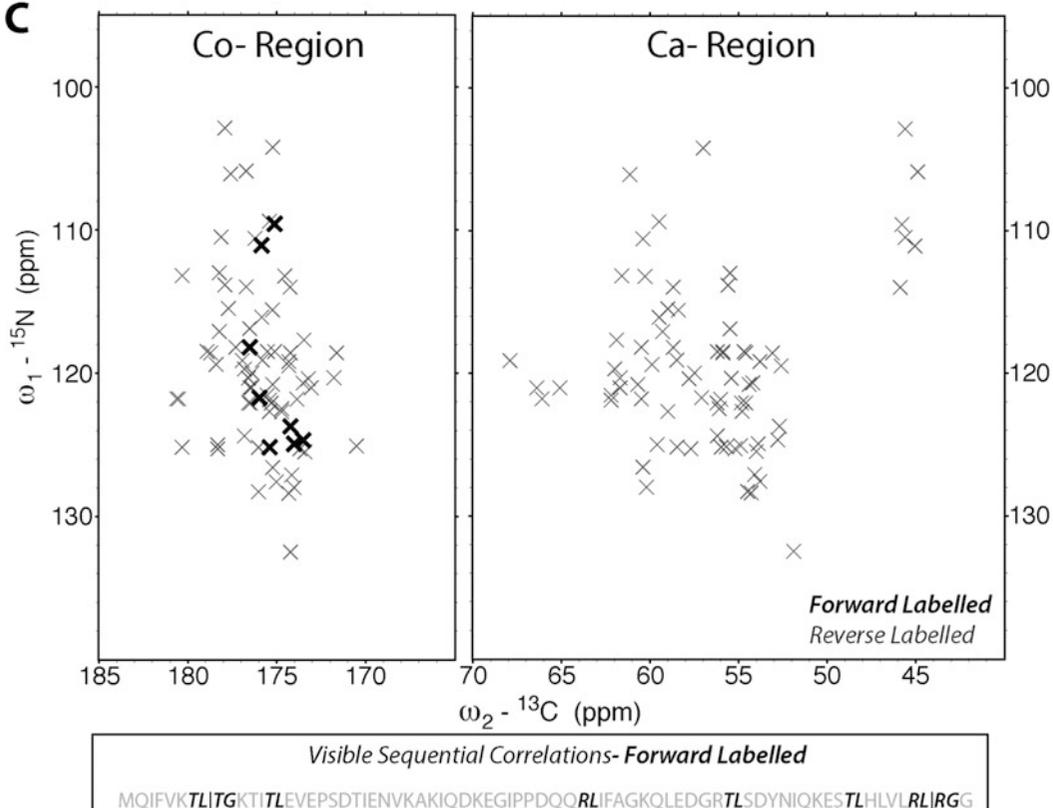
C

Fig. 15 Incorporating ¹³C labeled arginine, threonine and ¹⁵N labeled leucine, glycine amino acid selective labeling scheme in FANDAS. As can be seen in the peak prediction, sequential correlations can be observed for specific parts of the protein

Acknowledgments

This work was funded in part by the Netherlands Organization for Scientific Research (NWO) (grants 700.26.121 and 700.10.443 to M.B.). The development of the web portal was supported by a European H2020 e-Infrastructure grant West-Life (grant no. 675858 to A.B.). The authors would like to thank Panagiotis Koukos of the Computational Structural Biology Group for his humble assistance in hosting the webserver.

References

1. Renault M, Pawsey S, Bos MP, Koers EJ, Nand D, Tommassen-van Boxtel R, Rosay M, Tommassen J, Maas WE, Baldus M (2012) Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew Chem Int Ed Engl* 51(12):2998–3001. <https://doi.org/10.1002/anie.201105984>
2. Renault M, Tommassen-van Boxtel R, Bos MP, Post JA, Tommassen J, Baldus M (2012) Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc Natl Acad Sci USA* 109(13):4863–4868 [10.1073/pnas.1116478109](https://doi.org/10.1073/pnas.1116478109)
3. Kaplan M, Cukkemane A, van Zundert GC, Narasimhan S, Daniels M, Mance D, Waksman G, Bonvin AM, Fronzes R, Folkers GE, Baldus M (2015) Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat Methods* 12(7):649–652. <https://doi.org/10.1038/nmeth.3406>
4. Kaplan M, Narasimhan S, de Heus C, Mance D, van Doorn S, Houben K, Popov-Celeketic D, Damman R, Katrukha EA, Jain P, Geerts WJ, Heck AJ, Folkers GE, Kapitein LC, Lemeer S, van Bergen En Henegouwen PM, Baldus M (2016) EGFR dynamics change during activation in native membranes as revealed by NMR. *Cell* 167(5):1241–1251. [e1211. https://doi.org/10.1016/j.cell.2016.10.038](https://doi.org/10.1016/j.cell.2016.10.038)
5. Kaplan M, Pinto C, Houben K, Baldus M (2016) Nuclear magnetic resonance (NMR) applied to membrane-protein complexes. *Q Rev Biophys* 49:e15. <https://doi.org/10.1017/S003358351600010X>
6. Gradmann S, Ader C, Heinrich I, Nand D, Dittmann M, Cukkemane A, van Dijk M, Bonvin AM, Engelhard M, Baldus M (2012) Rapid prediction of multi-dimensional NMR data sets. *J Biomol NMR* 54(4):377–387. <https://doi.org/10.1007/s10858-012-9681-y>
7. Sinnige T, Weingarth M, Renault M, Baker L, Tommassen J, Baldus M (2014) Solid-state NMR studies of full-length BamA in lipid bilayers suggest limited overall POTRA mobility. *J Mol Biol* 426(9):2009–2021. <https://doi.org/10.1016/j.jmb.2014.02.007>
8. Sinnige T, Houben K, Pritisana I, Renault M, Boelens R, Baldus M (2015) Insight into the conformational stability of membrane-embedded BamA using a combined solution and solid-state NMR approach. *J Biomol NMR* 61(3–4):321–332. <https://doi.org/10.1007/s10858-014-9891-6>
9. Baker LA, Daniels M, van der Cruijssen EAW, Folkers GE, Baldus M (2015) Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J Biomol NMR* 62(2):199–208. <https://doi.org/10.1007/s10858-015-9936-5>
10. Renault M, Cukkemane A, Baldus M (2010) Solid-state NMR spectroscopy on complex biomolecules. *Angew Chem Int Ed Engl* 49(45):8346–8357. <https://doi.org/10.1002/anie.201002823>
11. Pauli J, Baldus M, van Rossum B, de Groot H, Oschkinat H (2001) Backbone and side-chain ^{13}C and ^{15}N signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 Tesla. *Chembiochem* 2(4):272–281
12. Sinnige T, Daniels M, Baldus M, Weingarth M (2014) Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J Am Chem Soc* 136(12):4452–4455. <https://doi.org/10.1021/ja412870m>
13. Mance D, Sinnige T, Kaplan M, Narasimhan S, Daniels M, Houben K, Baldus M, Weingarth M (2015) An Efficient labelling approach to harness backbone and side-chain protons in

- ¹H-detected solid-state NMR spectroscopy. *Angew Chem Int Ed Engl* 54(52):15799–15803 <https://doi.org/10.1002/anie.201509170>
14. Goddard TD, Kneller DG SPARKY 3. University of California, San Francisco
 15. Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* 11(4):852–861. <https://doi.org/10.1110/ps.3180102>
 16. Joosten RP, te Beek TA, Krieger E, Hekkelman ML, Hooft RW, Schneider R, Sander C, Vriend G (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res* 39(Database issue):D411–D419. <https://doi.org/10.1093/nar/gkq1105>
 17. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins* 23(4):566–579. <https://doi.org/10.1002/prot.340230412>
 18. Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43(W1):W389–W394. <https://doi.org/10.1093/nar/gkv332>
 19. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292(2):195–202. <https://doi.org/10.1006/jmbi.1999.3091>
 20. Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50(1):43–57. <https://doi.org/10.1007/s10858-011-9478-4>
 21. LeMaster DM, Kushlan DM (1996) Dynamical mapping of E. coli thioredoxin via ¹³C NMR relaxation analysis. *J Am Chem Soc* 118(39):9255–9264 doi:<https://doi.org/10.1021/ja960877r>
 22. Hong M, Jakes K (1999) Selective and extensive ¹³C labeling of a membrane protein for solid-state NMR investigations. *J Biomol NMR* 14(1):71–74
 23. 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(6911):98–102. <https://doi.org/10.1038/nature01070>
 24. Nand D, Cukkemane A, Becker S, Baldus M (2012) Fractional deuteration applied to biomolecular solid-state NMR spectroscopy. *J Biomol NMR* 52(2):91–101. <https://doi.org/10.1007/s10858-011-9585-2>
 25. Weingarth M, Demco DE, Bodenhausen G, Tekely P (2009) Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem Phys Lett* 469(4–6):342–348. <https://doi.org/10.1016/j.cplett.2008.12.084>

Strategies for Efficient Sample Preparation for Dynamic Nuclear Polarization Solid-State NMR of Biological Macromolecules

Boris Itin and Ivan V. Sergeyev

Abstract

Solid-state NMR (SSNMR) is a powerful tool for the elucidation of structure and dynamics in biological macromolecules. Over the years, SSNMR spectroscopists have developed an array of techniques enabling the measurement of internuclear correlations, distances, and torsional angles; these have been applied to the study of a number of biological systems that are difficult to study by X-ray crystallography and solution NMR, including key biological targets such as membrane proteins and amyloid fibrils. Applications of SSNMR to other topic areas, including materials science, pharmaceuticals, and small molecules, have also flourished in recent years. These studies, however, have always been hampered by the low inherent sensitivity of SSNMR, requiring large amounts of both sample and time for data collection. By taking advantage of unpaired electrons doped into a sample as a ready source of additional nuclear polarization, dynamic nuclear polarization (DNP) has brought about large improvements in SSNMR sensitivity. These, in turn, have enabled structural studies of previously inaccessible targets, such as large protein complexes, nucleic acids, viral capsids, and membrane proteins *in vivo*. Herein, we focus on sample preparation strategies and considerations for scientists interested in applying DNP to challenging systems.

Key words Dynamic nuclear polarization, Signal enhancement, Magnetic resonance, Solid-state NMR, Radicals

1 Introduction

Dynamic nuclear polarization (DNP) provides an opportunity to dramatically increase the sensitivity of solid-state NMR (SSNMR) experiments. By doping a sample with radicals, then transferring the much larger spin polarization of electrons to the protons ($\gamma_e/\gamma_H \approx 660$) or other NMR-active nuclei of the sample, large enhancements to the NMR signal are possible, on the order of several hundred- to several thousand-fold depending on the details of the polarization transfer pathway [1–3] (Fig. 1).

The resulting boost in signal-to-noise ratio provides an opportunity to study the structure of large proteins and protein assemblies

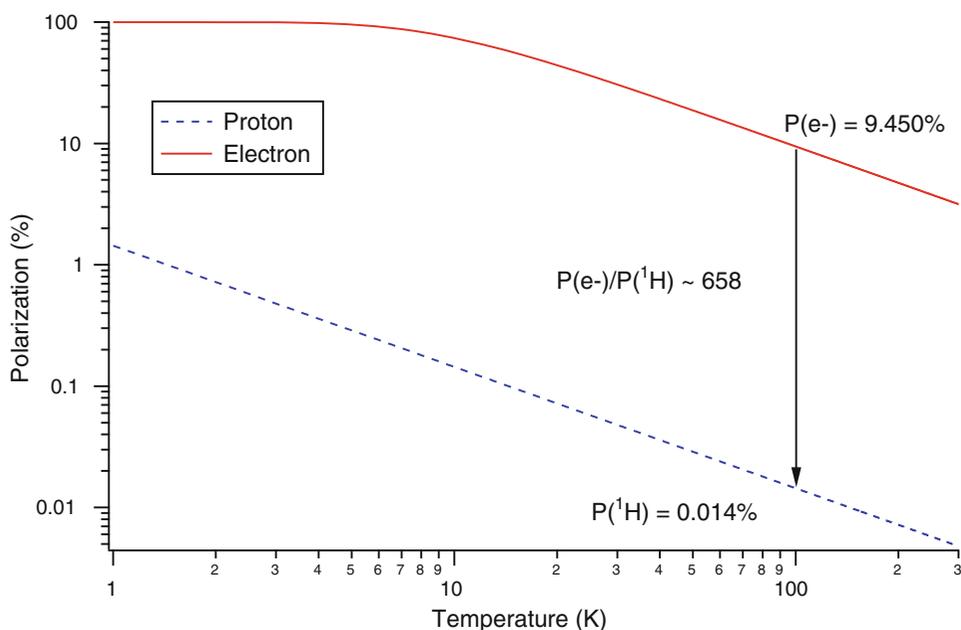


Fig. 1 Polarization of electron and ^1H spin baths at 14.1 T (600 MHz) as a function of sample temperature. Around 100 K, the theoretical maximum for ^1H enhancement approaches $\varepsilon = 658$

in vitro and even in vivo. To be successful, however, DNP studies require sophisticated hardware, methodology, and specific procedures for sample preparation. While we will touch on the first two topics briefly, this chapter focuses largely on the latter.

For readers interested in learning more about DNP, excellent reviews are available to cover recent advances, including applications of DNP at high fields [4, 5], for membrane proteins in lipid bilayers and in cells [6–8], matrix-free methodology [9], ultra-low temperature [10], sources and mechanisms of polarization [11], and others [12, 13]. Over the past 2 years, many more DNP studies have explored highly challenging biological systems and processes such as the cis-trans isomerization of retinal in Channelrhodopsin [14], ^{15}N PISEMA spectroscopy of oriented PGLa [15], membrane-anchored cytochrome-b5 in native *E. coli* cells [16], conformation, proton conduction and hydration in influenza B virus [17], and prion structure in a deuterated cellular lysate environment [18].

1.1 Goals

DNP provides significant signal enhancements and, consequently, tremendous savings in NMR acquisition time. Unfortunately, this often comes together with some loss in spectral resolution. Enhancement and resolution in DNP depend on many interconnected factors, often temperature- and sample-specific; some of these are well characterized, while others are only partially understood. The polarizing agent used, alongside its solubility/penetration into the material to be studied, is the single most

important factor determining DNP enhancement. Poor penetration of radicals may often lead to long DNP buildup times, which mean that maximum enhancement cannot be reached with reasonable values of the recycle delay. Line broadening, on the other hand, is a more complex phenomenon that appears to be primarily driven by sample inhomogeneity and/or intrinsic dynamics [5, 19]. As a result, DNP linewidths are difficult to predict or control, though care should be taken to ensure homogeneity of the sample to avoid the presence of a distribution of distinct molecular environments.

The primary goals in preparing a DNP sample, then, are to achieve the maximum possible enhancement to maximize signal-to-noise ratio (SNR), while keeping the DNP buildup time (T_B) short to avoid long recycle delays, and while also maintaining as narrow linewidths as possible. A secondary, but nonetheless important, objective for signal-to-noise considerations is to fit as much of the system of interest as possible into the DNP rotor, which implies keeping other components, such as any DNP solvent, to a minimum while still satisfying the conditions mentioned above. All of this must be done while maintaining compatibility and solubility between the biomolecule(s) of interest, the radicals, and the DNP solvent system. In essence, DNP sample preparation presents an optimization problem with a large number of opposing variables; herein, we attempt to merely suggest convenient starting points.

1.2 Hardware and Principles of Operation

In simplest terms, a modern, conventional DNP SSNMR spectrometer is a combination of five key components:

- NMR magnet, typically wide-bore, and often including an additional superconducting sweep coil with a range of several MHz.
- NMR console, typically optimized for SSNMR.
- Low-temperature magic-angle spinning (LTMAS) cabinet containing all equipment needed to generate and regulate spinning (bearing + drive) and variable temperature (VT) gases at approximately 100 K or below.
- SSNMR probe tunable to the nuclei of interest, with a microwave waveguide to direct microwaves toward the rotor.
- Microwave source at the corresponding EPR frequency, with waveguide to carry microwaves efficiently into the probe.

Microwave sources for DNP applications have typically been first- or second-harmonic gyrotrons, capable of output power 10–100 W at the EPR frequency corresponding to the magnetic field of the NMR magnet. Second-harmonic gyrotrons have gained favor in recent years, as they require a smaller magnetic field to operate and can therefore be placed closer to the NMR magnet, reducing microwave transfer losses. Recent developments have seen

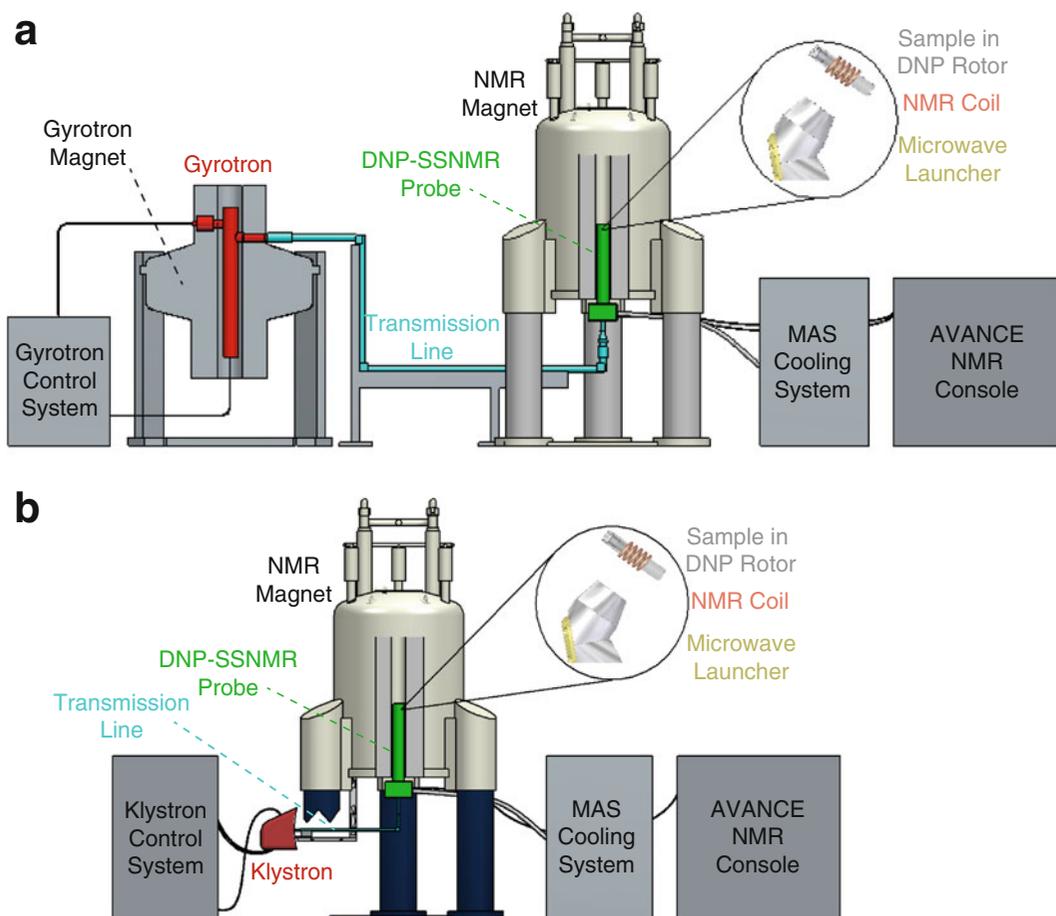


Fig. 2 Schematic of commercial DNP-SSNMR systems with (a) gyrotron and (b) klystron microwave sources, showing interconnection of the various components

the introduction of lower-power klystron microwave sources at 263 GHz, reducing cost and space requirements (Fig. 2). Commercial DNP hardware from Bruker Biospin is currently available at 400 MHz/263 GHz, 600 MHz/395 GHz, and 800 MHz/527 GHz.

Though there has been recent progress in this area [20], due to limitations in the tuneability of microwave sources at the frequencies required for DNP, the EPR frequency is generally kept constant while the NMR field is swept to move around the EPR spectrum (e.g., to match the EPR absorption frequencies of different radicals). As a result, the DNP spectrometer can be viewed as a mash-up of NMR and EPR spectrometers, with the EPR field sweeping accomplished via the NMR sweep coil, and NMR being performed in a standard fashion. Further, due to the use of linear-beam vacuum tubes such as gyrotrons or klystrons to generate microwaves,

conventional DNP operates in the continuous wave (CW) regime; that is, microwaves cannot be turned on and off on the NMR time-scale. For DNP-enhanced experiments then, the microwaves are always on.

In light of the above considerations, the typical DNP experiment proceeds very much like a conventional SSNMR experiment with several added steps. Once a DNP rotor has been inserted into the probe and spun up to the desired MAS frequency, a reference “microwave off” spectrum is typically taken prior to turning on the microwave source. With microwaves on, the experiment is then repeated and results overlaid with the microwave off spectrum to gauge the enhancement factor. Subsequently, the microwave source is left on continuously as the desired set of SSNMR experiments is conducted.

1.3 Key Considerations

1.3.1 Choice of Radical

The so-called cross-effect (CE) is the most commonly utilized DNP mechanism to date, and is so named because it arises from level anti-crossings in paired-radical systems. In its most common application, the CE is exploited via the addition of covalently-bonded biradicals; these are typically derivatives of the nitroxide radical 2,2,6,6-tetramethylpiperidine-N-oxyl (TEMPO) [21], long used for EPR studies. In general terms, a perfect radical for biological systems should combine the following properties: it should be stable, water soluble, able to access the molecules of interest in the sample, and have a large polarization transfer efficiency, preferably even at higher temperatures and magnetic fields. For commercial availability, it should also be reasonably inexpensive and easy to synthesize.

In recent years, the radicals TOTAPOL [22] and AMUPol [2] have gained popularity due to their large enhancements, with both now being available commercially. AMUPol in particular meets many of the criteria laid out above: it is very stable, scales reasonably well to higher fields, and provides good enhancement at temperatures significantly above 100 K. Enhancement factors as high as 250 have been reported on biological systems with AMUPol [23]. Other, more recent, biradicals, such as TEMTriPol [24], have been introduced to give more efficient CE enhancements at high magnetic field, ≥ 800 MHz, but are not yet commercially available.

In the vast majority of biological cases, AMUPol gives the largest enhancements and should be the biradical of choice for DNP studies. In especially hydrophobic systems, such as some materials (e.g., mesoporous materials, zeolites, long-chain polymers) and particularly hydrophobic biological environments, the bulky hydrophobic radicals of the TEKPol series [25] can be considered alternatives to AMUPol. It should be noted that these radicals are soluble only in organic solvents, and are typically impregnated into the sample (e.g., sample preparation Subheading

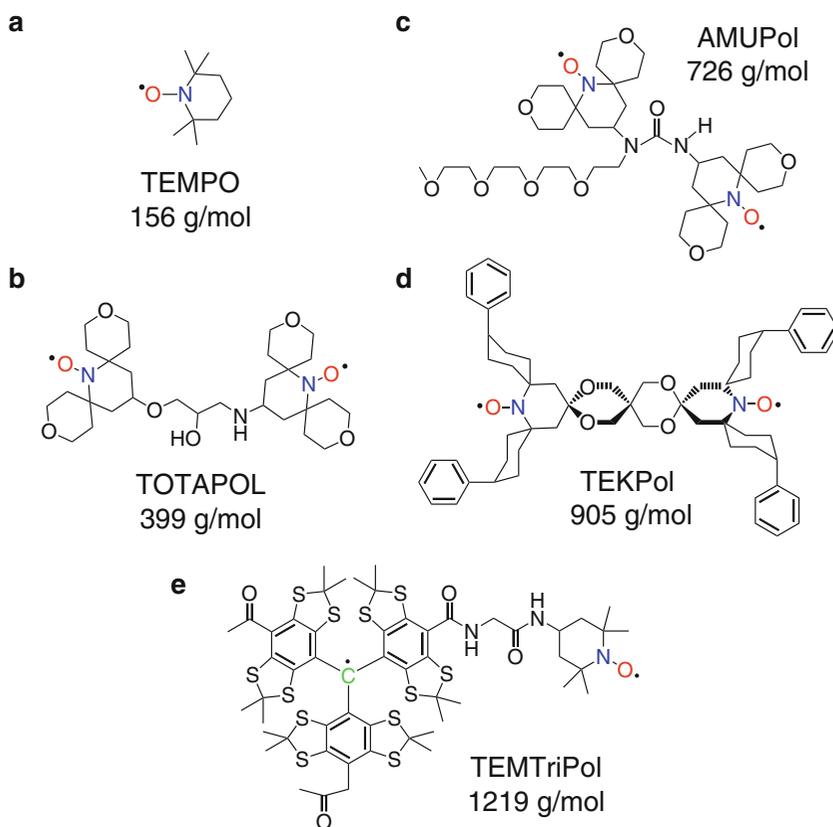


Fig. 3 Structures of (a) TEMPO, a nitroxide monoradical that is a precursor to most DNP biradicals, as well as the common DNP biradicals (b) TOTAPOL [22], (c) AMUPol [2], (d) TEKPol [25], and (e) the heterogeneous trityl-nitroxide biradical TEMTriPol [24]

3.4, below) in 1,1,2,2-tetrachloroethane (TCE). AMUPol and TEKPol are both soluble in toluene; toluene can be used as a glassy DNP solvent as well as an impregnation medium in cases where TCE is not well tolerated.

For high-field applications >800 MHz, some consideration should be given to acquiring biradicals designed for this purpose (e.g., TEMTriPol [24]). AMUPol is capable of providing solid enhancement factors $\epsilon \geq 40$ at 800 MHz; however, these drop off sharply with increasing magnetic field above this point (Fig. 3).

1.3.2 Choice of Solvent System

The vast majority of DNP studies have been carried out in glassy matrices [26, 27]. Glass-forming solvents serve several purposes, such as ensuring uniform distribution of radicals by preventing the formation of large solvent crystals, which can in turn lead to radical aggregation. Glassy solvents also largely prevent absorption of microwaves by the solvent, allowing them to more efficiently excite the radicals while also reducing heating of the sample. In addition,

many glass-forming mixtures contain either glycerol or dimethyl sulfoxide (DMSO), which concomitantly provide cryoprotection for the sample even at relatively low mole fractions, albeit by different mechanisms [28].

Biological DNP studies generally use the standard “DNP juice” recipe, consisting of d_8 -glycerol, D_2O , and H_2O , typically in a 60:30:10 ratio [2, 22] but with some variation in the ratios [18]. DMSO- d_6 has also been routinely used in mixtures with D_2O/H_2O , in cases of solubility problems with glycerol or where the high density and viscosity of glycerol pose problems; standard published recipes are 60/40 or as high as 77/23 DMSO:water mixtures [29]. For proton spin diffusion to spread the polarization throughout the sample, ~10% protonation is required. Hence, these ratios are more accurately written as 60:30:10 and 77:13:10 DMSO- d_6 : D_2O : H_2O . Lower concentrations of glycerol or DMSO are acceptable in terms of both cryoprotection and glass-forming potential; in the authors’ hands, samples prepared with 30% v/v glycerol- d_8 and 30% v/v DMSO- d_6 have shown promising DNP results with no deleterious effects relative to higher concentrations. In practical terms, 30% glycerol or DMSO allows for a larger filling factor for the sample of interest, typically affording superior signal-to-noise.

For highly hydrophobic systems that are not soluble in DMSO, one may consider alternate solvent combinations. Meyer [30] offers a wide variety of organic glass-forming solvent mixtures. Toluene is a particularly attractive solvent in that it is relatively nontoxic and neat toluene is a glass-former. However, due to its high vapor pressure, care should be taken in sealing samples in the rotor; toluene DNP samples are not likely to be viable for extended periods even when refrigerated.

Quite often, as in the case of membrane proteins embedded in lipid bilayers, fibril aggregates, virus capsids, and other complex biological samples, the mixture resulting from addition of DNP solvents and radicals is heterogeneous. Mechanical stirring and/or gentle sonication are the recommended means of homogenization. Ideally, the sample should be homogenized immediately before insertion into the DNP spectrometer, or should be immediately frozen to prevent phase separation.

If a sample already contains H_2O and D_2O , instead of adding DNP juice, the optimal protocol is to add pure d_8 -glycerol. For instance, Frederick et al. [18] resuspended cell lysate samples in 100% ^{13}C -depleted glycerol- d_8 with 20 mM of TOTAPOL. As an alternative, an incubation step with excess solvent can be added; for instance, Gupta et al. [31] incubated the protein preparation in 50/35/15 d_8 -glycerol/ D_2O / H_2O solution overnight and removed the excess supernatant before packing the sample.

When the protein/biomolecule of interest cannot be conveniently grown/synthesized with substantial deuteration, the protonation of the protein is generally sufficient to achieve adequate protonation of the overall sample. In such instances, we recommend that the added H₂O fraction be reduced to 5% or excluded altogether in favor of D₂O, depending on the biomolecule concentration in the sample.

1.3.3 Choice of Magnetic Field and MAS Frequency

Line broadening at lower temperatures is caused by a number of factors including inhomogeneous sample broadening, conformational exchange dynamics, and dipolar coupling. In essence, at the cryogenic temperatures most typically used in DNP (typically 85–115 K) numerous dynamical modes are “frozen out,” and the sample begins to resemble a complex mixture of various conformers, with particular heterogeneity at protein sidechains and similar small, generally mobile, domains. Naturally, such modes depend heavily on the intrinsic dynamics of the sample [7]. At the present time, no particular combination of sample conditions has been found to dramatically reduce inhomogeneity; samples that are well ordered at room temperature tend to also give sharper DNP linewidths [5, 19]. As an example, Fricke et al. [19] studied type 3 secretion system needles (T3SS) where the low flexibility of the complex decreased any inhomogeneous broadening contribution, improving resolution significantly.

Faster magic-angle spinning may alleviate line broadening contributions due to dipolar coupling, improving resolution. Unfortunately, fast spinning at the lower temperatures needed for DNP is challenging mechanically; lower gas viscosity requires higher nitrogen gas flows. Maximal MAS frequency in the 100–110 K temperature range is 30–50% lower for a corresponding rotor size than it is at room temperature. As a result, the overwhelming majority of DNP research is performed in 3.2 mm rotors with MAS in the 8–12 kHz spinning frequency range. Interestingly, an amyloid fibril sample spinning at 25 kHz in a 1.9 mm rotor demonstrated 25–30% line narrowing for a large number of (but not all) ¹³C sites [32]. Further, in studies of intact Pfl bacteriophage at 25 kHz, the authors observed significant linewidth reductions: ¹³C linewidths dropped to 1.0 ppm from 1.5 ppm at 12 kHz, while ¹⁵N linewidths decreased to 3.5 ppm from 4.9 ppm (Fig. 4). Chaudhary et al. [33] studied DNP effects at 40 kHz MAS using the 1.3 mm rotor system. They did not, however, observe any resolution improvement in a frozen solution proline sample, where inhomogeneity is likely to dominate, suggesting that narrower DNP linewidths due to faster MAS may not generalize to all circumstances.

When lineshapes are not dominated by inhomogeneous line broadening, higher magnetic fields improve resolution. Availability of 600 MHz and 800 MHz DNP instruments has made it possible

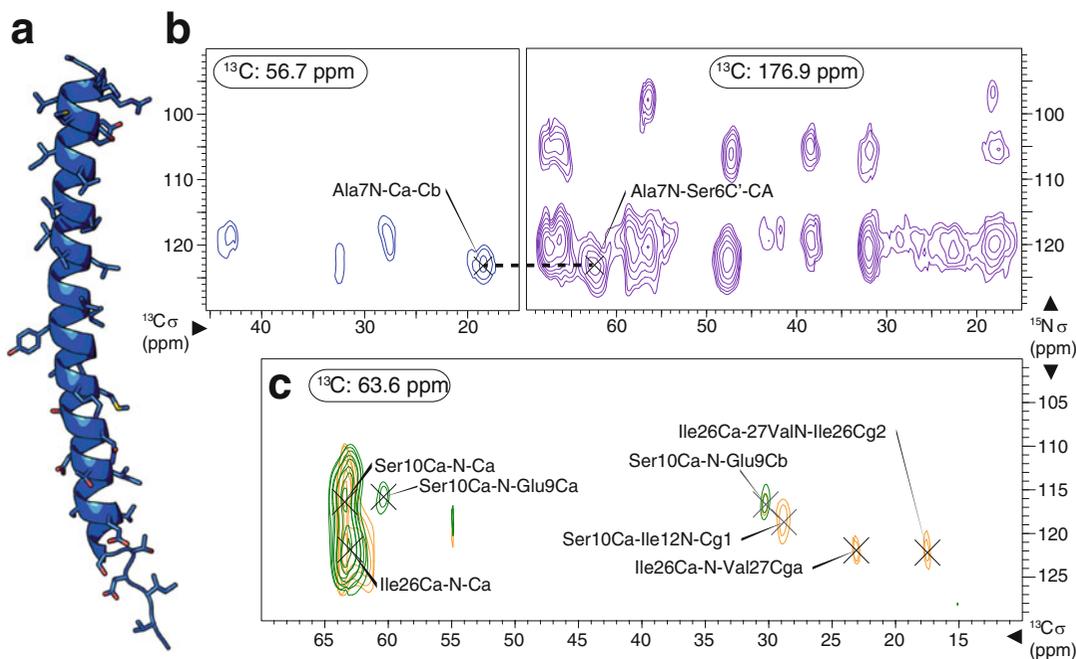


Fig. 4 An illustration of the resolution possible in 3-dimensional NCC/CNC correlations of intact U- ^{13}C , ^{15}N Pf1 bacteriophage at 25 kHz MAS using a 1.9 mm rotor. With the benefit of DNP enhancement, short-range ^{15}N - ^{13}C - ^{13}C correlations [NCACX, *blue* & NCOCA, *purple*] (b) alongside longer-range inter-residue ^{13}C - ^{15}N - ^{13}C correlations [$i \rightarrow i-1$, *green* & $i \rightarrow i+1$, *orange*] (c) enabled the full resonance assignment of the 46-residue Pf1 major coat protein (a) in a time-efficient manner

to conduct SSNMR structural studies on highly challenging systems including large protein complexes *in vivo*. Kaplan et al. [34] studied 1 MDa bacterial type IV secretion system core complex T4SScc in a cellular setting. Enhancement factors of 60 at 400 MHz and 15 at 800 MHz enabled 2D and 3D NC α CX and NCOCX experiments to be performed. Gupta et al. [31] were able to detect side-chain conformers and conformationally disordered states, as well as to measure intermolecular correlations between CA monomers in HIV1 capsids.

2 Materials

1. D_2O , glycerol- d_8 , U- ^{13}C , ^{15}N -Proline, AMUPol, KBr, adamantane.
 - (a) DMSO- d_6 , Toluene- d_8 , 1,1,2,2-tetrachloroethane (TCE) may be necessary solvents for some samples.
 - (b) DOPE/DOPS or VM+, octyl- β -D-glucoside (OG) or decyl- β -maltopyranoside (DM) may be necessary for some membrane protein samples.
 - (c) Trehalose may be necessary for matrix-free preparations.

2. Bruker DNP rotor and packing set: 3.2 mm sapphire DNP rotor, soft plug kit, zirconium oxide drive cap, powder funnel, packing tool, depth gauge tool, soft plug extraction tool.
3. 1.5–2.0 mL Eppendorf tubes, 5–100 μL pipettes, flat needles in various gauges, mortar and pestle grinding set.
4. Benchtop centrifuge (capable of up to $\sim 15,000 \times g$), vortexer, sonicator.
 - (a) Dessicator or lyophilizer may be necessary for some samples.
 - (b) Ultracentrifuge may be necessary for some membrane protein samples.

3 Methods

Below, we make an attempt to summarize standard protocols for the preparation of several common types of DNP samples. It is important to note here that there are infinite variations possible upon the general theme, many of them likely leading to successful outcomes. We hope that these protocols will serve as general guidelines for readers starting DNP studies, but advise that they be adapted as needed to the challenges of the system at hand. Where available, we have cited recent work to highlight similarities and differences with the proposed protocols.

3.1 *Sample Preparation—Proline/AMUPol Standard*

To assess the performance of a DNP system, a proline/AMUPol standard sample is recommended. To prepare a standard solution consisting of 0.25 M $\text{U-}^{13}\text{C}$, ^{15}N -Proline and 10 mM AMUPol in 60:30:10 glycerol- $\text{d}_8/\text{D}_2\text{O}/\text{H}_2\text{O}$ mixture:

1. To a 1.5–2.0 mL Eppendorf tube, add 15.6 mg $\text{U-}^{13}\text{C}$, ^{15}N -Proline (MW: 125.1 g/mol).
2. Add 3.6 mg AMUPol (MW: 726.0 g/mol).
3. Add 150 μL D_2O and 50 μL H_2O .
4. Vortex thoroughly until AMUPol is fully dissolved and a clear yellow solution is observed.
5. Separately, weigh out 411 mg of glycerol- d_8 ($d = 1.37 \text{ g/mL @ } 25 \text{ }^\circ\text{C}$), dropwise by mass, using a pipette tip that has been cut to enlarge the opening.
6. Mix the AMUPol/water solution into the glycerol.
7. Vortex thoroughly until a clear homogeneous solution is achieved.
8. Using a pipette, transfer 28 μL of the solution to the bottom of a 3.2 mm sapphire DNP rotor, being careful to avoid bubbles.

9. Insert soft plug into the rotor, using the depth gauge tool to push down to the appropriate level. Remove capillary tube once excess sample has been released (*see* Fig. 5o–s).
10. Seal rotor with drive zirconium oxide DNP cap (*see* Fig. 5t–v).

3.2 Sample Preparation—Water Soluble Proteins or Protein Assemblies

System to be studied: A large, water-soluble protein or protein assembly (e.g., globular proteins, amyloid fibrils, viral particles, or similar), uniformly or sparsely ^{13}C , ^{15}N -labeled in the protein(s) of interest. Use of a 3.2 mm sapphire DNP rotor is assumed.

1. If not already in one, sample should be reconstituted into fully deuterated buffer at a pH appropriate for the system of interest to remove excess protonation. The deuteration level of the buffer should be kept as high as feasible.
2. Sediment, precipitate, or concentrate protein and/or assembly out of solution as appropriate, forming a thick hydrogel.
3. For precipitates, spin sample repeatedly at approximately $15,000 \times g$ in a benchtop centrifuge to bring excess water to the surface. Remove any supernatant via pipette, being careful not to disturb the pellet. Repeat until no further supernatant forms above the pellet.
4. Estimate the volume of the sample. If $\leq 15 \mu\text{L}$ skip to **step 6**.
5. When the sample is in a small volume (approximately 1 mL) of D_2O , transfer just enough of the solution to a 0.5 mL Eppendorf tube to achieve a sample volume of approximately $15 \mu\text{L}$ upon re-precipitation or re-sedimentation. Once again, precipitate or sediment the sample in the 0.5 mL Eppendorf tube, spin down at $\sim 15,000 \times g$ several times, and remove any supernatant until no further supernatant results.
6. Separately, generate a stock solution containing 20 mM AMU-Pol in a mixture of 60% glycerol- d_8 , 40% D_2O .
7. Pipette a volume of the stock solution equal to that of the sample (generally $15 \mu\text{L}$), over the sample.
8. Stir the resulting mixture using a thin wire or similar tool until a uniform appearance is achieved. Sonicate briefly and gently to homogenize (*see* **Note 1**).
9. Spin the sample down gently (e.g., $2000 \times g$ for 30 s, longer if necessary depending on the viscosity of the sample) to compact the sample pellet at the bottom of the Eppendorf tube.
10. Cut off the bottom part of the 0.5 mL Eppendorf tube containing the homogenized sample with a razor blade or scalpel, cutting well above the pellet itself (*see* Fig. 5a, b).
11. Invert the lower portion of the Eppendorf tube containing the sample into a sample-packing funnel, into which a DNP rotor

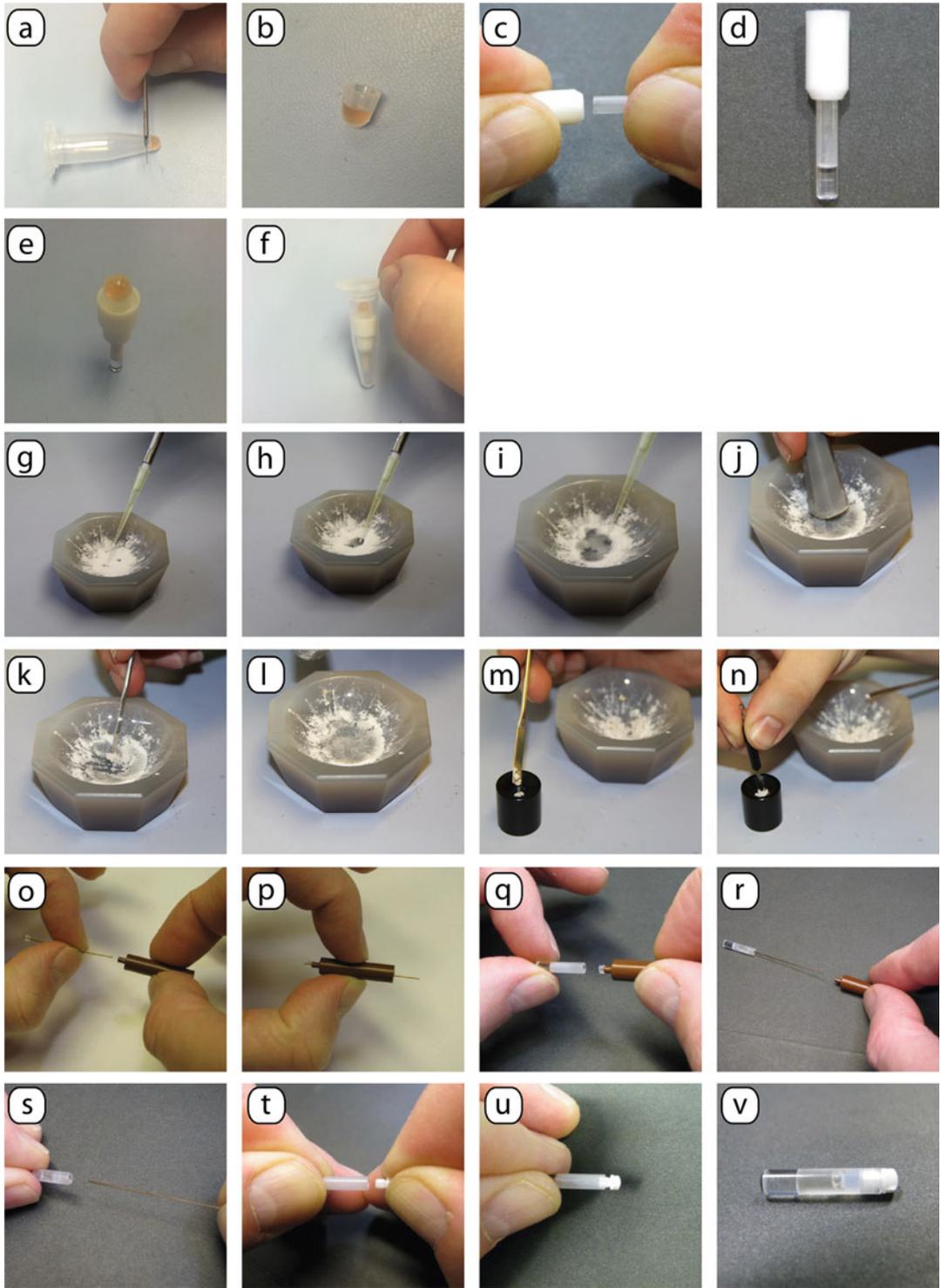


Fig. 5 Illustration of indicated steps from sample preparation protocols, numbered accordingly

has already been inserted. Insert the whole assembly into a 2 mL Eppendorf tube (*see* Fig. 5c–f).

12. Gently centrifuge the sample pellet (e.g., $2000 \times g$ for 30 s) into the rotor.
13. Some samples may be too viscous to be pipetted or centrifuged into the rotor. In such cases, it is typically possible to mechanically transfer the sample: Using a thick needle, pick up 1–2 mg of the sample and deposit on the rotor's inside surface near the opening. Centrifuge the rotor at $6000\text{--}8000 \times g$ in a swinging-bucket centrifuge for 2–3 min. Repeat **steps 11–12** to fill the rotor to the correct level, leaving space for the soft plug and drive cap. It is important not to overfill the rotor in such cases, as the excess sample can be very difficult to remove without losses.
14. If any separation of the sample pellet is visible, stir the sample inside the rotor using a thin needle.
15. Insert a soft plug into the rotor, using the depth gauge tool to push down to the appropriate level. Remove the capillary tube once excess sample has been released (*see* Fig. 5o–s).
16. Seal the rotor with a zirconium oxide DNP drive cap (*see* Fig. 5t–v).
17. (Optional step) Several insert-eject cycles can be performed in situ to remove the bulk of absorbed oxygen in the sample, which can boost enhancement [35–37]. Insert the sample into a DNP probe cooled down to ~ 100 K and allow it to spin at moderate MAS (~ 5 kHz) for 2 min. Stop sample spinning and eject the sample. Allow the sample to sit in the ejection cup under nitrogen flow for 2 min. Repeat these steps thrice to deoxygenate the sample.

3.3 Sample Preparation— Membrane Proteins

System to be studied: Membrane protein in native or semi-native lipid bilayer. *See* Liao et al. [38] for additional information.

1. Reconstitute the membrane protein of interest in DOPE/DOPS, VM+ [39], or similar lipid mixture as appropriate. Typically, the protein is purified separately and suspended in octyl- β -D-glucoside (OG), decyl- β -maltopyranoside (DM), or similar detergents. The lipid mixture is prepared in organic solvents and subsequently vacuum dried under nitrogen; it is then suspended in an appropriate aqueous buffer. Mix the two solutions and incubate for several hours to ensure proper incorporation of the protein, followed by a dialysis step to remove OG/DM. Perform numerous (e.g., 5–10) freeze-thaw cycles, between room and liquid nitrogen temperature, to homogenize vesicle size and remove excess water; this can be done

before or after protein incorporation, depending on protein stability.

2. Ultracentrifuge dialyzed proteoliposomes overnight to form a dense membrane pellet.
3. Separately, generate a stock solution of ~20 mM AMUPol in 60:30:10 glycerol- d_8 : D_2O : H_2O ; titrate the solution into the proteoliposome pellet to achieve an overall AMUPol concentration ~10 mM (*see Note 2*).
4. Vortex and/or sonicate the pellets gently to ensure homogeneous distribution of radicals.
5. Add D_2O as necessary to increase the deuteration level of the sample; the final D_2O : H_2O ratio should approach 3:1.
6. Remove excess water by lyophilization down to a hydration level of ~40% by mass, or as appropriate to the system of interest.
7. Using a razor blade, cut off the bottom part of the Eppendorf tube containing the homogenized sample, cutting well above the pellet itself (*see Fig. 5a, b*).
8. Invert the lower portion of the Eppendorf tube containing the sample into a sample-packing funnel, into which a DNP rotor has already been inserted. Insert the whole assembly into a 2 mL Eppendorf tube (*see Fig. 5c–f*).
9. Gently centrifuge the sample pellet (e.g., $2000 \times g$ for 30 s) into the rotor. Some samples may be too viscous to be pipetted or centrifuged into the rotor. In such cases, it is typically possible to mechanically transfer the sample: Using a thick needle, pick up 1–2 mg of the sample and deposit on the rotor's inside surface near the opening. Centrifuge the rotor at 6000 – $8000 \times g$ in a swinging-bucket centrifuge for 2–3 min. Repeat these steps to fill the rotor to the correct level, leaving space for the soft plug and drive cap. It is important not to overfill the rotor in such cases, as excess sample can be very difficult to remove. If any separation of the sample pellet is visible, stir the sample inside the rotor using a thin wire or similar tool.
10. Insert a soft plug into the rotor, using the depth gauge tool to push down to the appropriate level. Remove the capillary tube once excess sample has been released (*see Fig. 5o–s*).
11. Seal the rotor with a zirconium oxide DNP drive cap (*see Fig. 5t–v*).

3.4 Sample Preparation— Materials

System to be studied: porous materials (e.g., silica-based frameworks), crystalline small molecules, or crystalline/lyophilized proteins stable in organic solvents. The “impregnation” technique

developed by Emsley and coworkers [40] will be used to introduce radicals into the sample in a virtually solvent-free manner.

1. Crush and then finely grind approximately 30–40 mg (rough estimate for 3.2 mm rotors; *see* **Note 3**) of the compound of interest with a mortar and pestle.
2. Separately, generate a standard solution of biradical in an organic solvent. The biradical should be chosen based on the hydrophobicity of the compound of interest. For largely hydrophobic molecules, TEKPOL [41] or similar biradicals should be used; for more hydrophilic molecules, AMUPol remains the biradical of choice. Dissolve the biradical in sufficient 1,1,2,2-tetrachloroethane (TCE) to make a 10–20 mM solution (*see* **Note 4**). Toluene may be used in lieu of TCE in cases where TCE is incompatible with the compound of interest; as toluene evaporates slower than TCE, the impregnated compound should be left to sit longer prior to packing.
3. Pipette approximately 30 μL (*see* **Note 4**) of the radical-containing solution over the ground compound in the mortar. Typically, this will result in a “wet spot” at the site of contact, with the rest of the compound remaining dry. Vigorously grind the compound to spread the radical-containing solution evenly (*see* Fig. 5g–l). If portions of the ground compound remain visibly dry, add more radical-containing solution to wet these portions, keeping track of the total volume added.
4. Allow the wetted compound to stand and dry over approximately 10 min.
5. Using a powder funnel, gradually pack the impregnated compound into the rotor, compact periodically to maximize sample volume (*see* Fig. 5m, n).
6. With powdered samples, there is typically no need to use a soft plug to seal the top of the rotor; this step is therefore optional. A Teflon spacer, which is simpler to remove, may also be used in lieu of a soft plug, if desired.
7. Seal the rotor with a zirconium oxide DNP drive cap (*see* Fig. 5t–v).
8. (Optional step) Several insert-eject cycles can be performed in situ to remove the bulk of absorbed oxygen in the sample, if this is a concern [36]. Insert the sample into a DNP probe cooled down to ~ 100 K and allow it to spin at moderate MAS (~ 5 kHz) for 2 min. Stop sample spinning and eject the sample; allow the sample to sit in the ejection cup under nitrogen flow for 2 min. Repeat these steps thrice for proper deoxygenation.

**3.5 Sample
Preparation—
Precipitated Protein**

System to be studied: protein or protein mixture (generally, one which can be easily precipitated from the solution) in a “matrix-free” preparation.

1. Prepare a dilute (1–10 mM) solution of the protein of interest in D₂O with an appropriate buffering agent.
2. Separately, generate a solution buffered to the appropriate pH and containing 10–100 mM trehalose and 20 mM AMUPol in D₂O. This solution should be capable of precipitating the protein of interest in the 1–10 mM concentration range. For many proteins, this will entail either high-salt (e.g., 2 M NaCl), polyethylene glycol (e.g., 2% PEG-8000 w/v), or other additives such as isopropanol.
3. To 2–5 mL of the protein solution (a larger volume should be used in the case of low protein concentrations), add an equal volume of the trehalose-AMUPol-precipitant solution and thoroughly vortex.
4. Partially evaporate the solution to achieve precipitation (*see Note 5*).
5. Once a significant amount of precipitate has formed, transfer the remaining solution to 1.5 mL Eppendorf tubes, and collect precipitate by centrifugation (e.g., 8000 × *g* for 15 min) in a benchtop centrifuge. For each of the tubes, remove any supernatant, first via pipette, and then via gentle dabbing with laboratory tissue or similar material.
6. Using a razor blade or scalpel, cut off the bottom part of the Eppendorf tubes containing the precipitated samples, in each case cutting well above the pellet itself (*see Fig. 5a, b*).
7. Invert the lower portion of an Eppendorf tube containing the sample into a sample-packing funnel, into which a DNP rotor has already been inserted. Insert the whole assembly into a 2 mL Eppendorf tube (*see Fig. 5c–f*).
8. Gently centrifuge the sample pellet (e.g., 2000 × *g* for 30 s) into the rotor. Some samples may be too viscous to be pipetted or centrifuged into the rotor. In such cases, it is typically possible to mechanically transfer the sample: Using a thick needle, pick up 1–2 mg of the sample and deposit on the rotor’s inside surface near the opening. Centrifuge the rotor at 6000–8000 × *g* in a swinging-bucket centrifuge for 2–3 min. Repeat these steps to fill the rotor to the correct level, leaving space for the soft plug and drive cap.
9. Repeat **steps 7–8** for the contents of the other Eppendorf tubes or until the rotor is filled to capacity, leaving space for the soft plug and drive cap.

10. Insert a soft plug into the rotor, using the depth gauge tool to push down to the appropriate level. Remove the capillary tube once excess sample has been released (*see* Fig. 5o–s).
11. Seal the rotor with a zirconium oxide DNP drive cap (*see* Fig. 5t–v).

4 Notes

1. The resulting sample is approximately 30% glycerol- d_8 , 70% D_2O by volume, with approximately 10 mM AMUPol. While higher percentages of glycerol are used in standard “DNP Juice” recipes, we have found that 30% provides sufficient cryoprotection and a sufficiently glassy matrix at ~ 100 K; increasing the glycerol concentration did not benefit DNP parameters. Residual protonation of the sample is sufficient to provide the 10% 1H of conventional 60:30:10 “DNP Juice.” If, however, the biomolecule of interest is heavily deuterated, it is advisable to add 10% H_2O to the solvent mixture to achieve satisfactory 1H spin diffusion.
2. To reduce swelling of proteoliposome samples, some groups advocate using a much lower glycerol fraction [42]. As in the first example (Subheading 3.2), we believe that 20–30% glycerol is adequate for cryoprotection; therefore, the glycerol:water ratio may be safely lowered if swelling is a concern, or simply to fit more protein into the rotor.
3. The amount of compound loaded should be slightly larger than the amount required to fill a rotor, while leaving space for the drive cap and optional spacers. If the packing density of the compound is not known, it is advisable to first fill a rotor with the ground non-impregnated material and measure the mass of compound that fits. At a later stage, this information will also help estimate the sample volume and radical concentration.
4. It is advisable to aim for a total radical concentration of approximately 10–20 mM in the rotor. A simple method of doing so is to approximate the volume of the sample based on the mass of compound loaded into the mortar, and then add a sufficient amount of radical-containing solution to achieve the desired radical concentration, assuming that all solvents will evaporate and the sample will not swell.
5. Protein solutions can be evaporated in numerous ways, including:
 - (a) In a dessicator under vacuum for an appropriate amount of time.

- (b) In an airflow chamber, where dry air or nitrogen is passed over the sample.
- (c) Via a lyophilizer, using a quick run where target water concentration should be roughly 40–50%. The sample should be partially dessicated, not lyophilized. If over-dessication is a concern, the mass of the sample can be tracked periodically to ensure that sufficient water remains.
6. Recommended setup sequences for DNP-SSNMR experiments are described in Tables 1 and 2.
7. Regarding enhancement factors (ϵ) and absolute sensitivity ratio (ASR):
Historically, the enhancement in DNP experiments has been measured as an intensity ratio, with microwaves on relative to microwaves off, for a given peak in otherwise identical spectra of the same sample. This is a simple and quick approach; however, it has some drawbacks. One important consideration is the choice of recycle delay—using recycle delays that are too

Table 1
Room temperature setup/calibration experiments (before cooling LTMAS probe)

Sample	MAS (kHz)	Expt.	μ waves	Purpose
KBr	5	1-pulse	Off	Set magic angle
Adamantane	8–12	CP	Off	Reference downfield ^{13}C peak to DSS @ 40.48 ppm [43]
Adamantane	8–12	1-pulse with decoupling during acq.	Off	(optional) optimize ^1H homonuclear decoupling, if desired, by optimizing splitting due to ^1H - ^{13}C ^1J -couplings to be ~ 75 Hz.

Table 2
Low temperature setup/calibration experiments (after cooling LTMAS probe)

Sample	MAS (kHz)	Expt.	μ waves	Purpose
KBr	5	1-pulse	Off	Set magic angle
Proline/AMUPol standard	8	CP	Off, On	Check enhancement
Proline/AMUPol standard	≥ 8	CP-90	On	Check/optimize power levels, CP conditions, decoupling
Proline/AMUPol standard	≥ 8	Double CP	On	Setup of 2D, 3D experiments (e.g., using TopSolids in topspin), if necessary

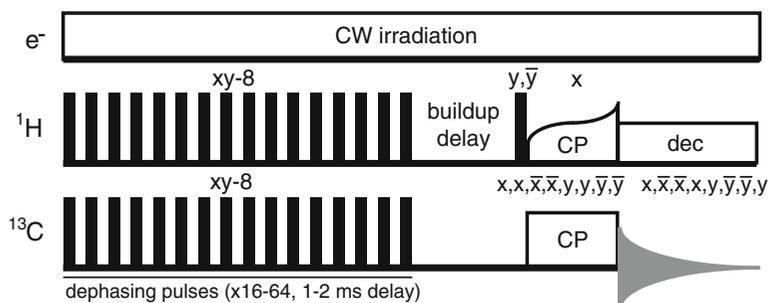


Fig. 6 Pulse sequence for measurement of the DNP buildup time (T_B) via ^1H - ^{13}C cross-polarization following saturation-recovery. The experiment is acquired in a pseudo-2D manner where the buildup delay is incremented between successive slices

short will not permit the DNP enhancement to build up to its maximal value, while delays that are too long will unnecessarily waste instrument time. The authors recommend using a standard recycle delay across all samples for enhancement values obtained in this manner, typically 10 s.

Further, the DNP buildup time [44, 45] of the sample should be assessed and reported using a CP-based saturation recovery pulse sequence such as shown below. It is advisable to include a train of saturation pulses on all nuclei used in the experiment. Both the ^1H longitudinal relaxation and DNP mechanisms will contribute to the buildup profile; typically the ^1H T_1 and T_B are closely linked. Peak integrals from saturation recovery experiments are typically fit with the formula $I(t) = I(0) * (1 - e^{-t/T_B})$ (Fig. 6).

Another key consideration is the issue of depolarization [46, 47]. Energy level anti-crossings typical of the cross-effect, under MAS conditions but without the presence of microwaves, can have the effect of artificially relaxing the nuclear polarization of the sample, thereby resulting in a reduced signal with microwaves off. This would skew enhancements reported using the simple method above toward larger values. As a result, Takahashi and coworkers [47] suggest the use of a more complex metric, the absolute sensitivity ratio (ASR). ASR involves first measuring a sample of the same protein/biomolecule close to ambient temperature and without radicals, then comparing the signal intensities to those under DNP to assess the true gain in signal-to-noise ratio. Needless to say, this can be difficult and very time-consuming, and is not possible for some systems that are not amenable to room temperature spectroscopy.

While significant depolarization has been observed when using AMUPol, the signals examined were those of the solvent (glycerol), which is by definition in direct contact with the radicals in the sample. Whether this effect is generalizable to solutes

remains an open question. In the authors' hands, depolarization effects on proteins at 10 mM AMUPol concentrations have been insignificant. As a result, we suggest a compromise: the extent of depolarization may be assessed by comparing CP intensity at several different MAS frequencies. If the intensity does not dramatically decrease with increasing MAS frequency in the range of 2–12 kHz, it is reasonable to infer that depolarization is not a major concern. In such cases, the simple on/off enhancement measurement can safely be reported in lieu of ASR.

Acknowledgment

The authors would like to thank Dr. James Kempf, Dr. Shane Pawsey, Bryce Smith, and Christopher Hickey of Bruker Biospin for many helpful discussions, figures, and photographs.

References

- Hirsch ML, Kalechofsky N, Belzer A et al (2015) Brute-force hyperpolarization for NMR and MRI. *J Am Chem Soc* 137 (26):8428–8434. doi:[10.1021/jacs.5b01252](https://doi.org/10.1021/jacs.5b01252)
- Sauvée C, Rosay M, Casano G et al (2013) Highly efficient, water-soluble polarizing agents for dynamic nuclear polarization at high frequency. *Angew Chem Int Ed* 52:10858–10861. doi:[10.1002/anie.201304657](https://doi.org/10.1002/anie.201304657)
- Smith AN, Caporini MA, Fanucci GE, Long JR (2015) A method for dynamic nuclear polarization enhancement of membrane proteins. *Angew Chem Int Ed* 54:1542–1546. doi:[10.1002/anie.201410249](https://doi.org/10.1002/anie.201410249)
- Ni QZ, Daviso E, Can TV et al (2013) High frequency dynamic nuclear polarization. *Acc Chem Res* 46:1933–1941. doi:[10.1021/ar300348n](https://doi.org/10.1021/ar300348n)
- Su Y, Andreas L, Griffin RG (2015) Magic angle spinning NMR of proteins: high-frequency dynamic nuclear polarization and $(1)H$ detection. *Annu Rev Biochem* 84:465–497. doi:[10.1146/annurev-biochem-060614-034206](https://doi.org/10.1146/annurev-biochem-060614-034206)
- Cheng C-Y, Han S (2013) Dynamic nuclear polarization methods in solids and solutions to explore membrane proteins and membrane systems. *Annu Rev Phys Chem* 64:507–532. doi:[10.1146/annurev-physchem-040412-110028](https://doi.org/10.1146/annurev-physchem-040412-110028)
- Kaplan M, Pinto C, Houben K, Baldus M (2016) Nuclear magnetic resonance (NMR) applied to membrane-protein complexes. *Q Rev Biophys* 49:e15. doi:[10.1017/S003358351600010X](https://doi.org/10.1017/S003358351600010X)
- Warnet XL, Arnold AA, Marcotte I, Warschawski DE (2015) In-cell solid-state NMR: an emerging technique for the study of biological membranes. *Biophys J* 109: 2461–2466. doi:[10.1016/j.bpj.2015.10.041](https://doi.org/10.1016/j.bpj.2015.10.041)
- Lee D, Hediger S, De Paëpe G (2015) Is solid-state NMR enhanced by dynamic nuclear polarization? *Solid State Nucl Magn Reson* 66:6–20. doi:[10.1016/j.ssnmr.2015.01.003](https://doi.org/10.1016/j.ssnmr.2015.01.003)
- Lee D, Bouleau E, Saint-Bonnet P et al (2016) Ultra-low temperature MAS-DNP. *J Magn Reson* 264:116–124. doi:[10.1016/j.jmr.2015.12.010](https://doi.org/10.1016/j.jmr.2015.12.010)
- Hu KN (2011) Polarizing agents and mechanisms for high-field dynamic nuclear polarization of frozen dielectric solids. *Solid State Nucl Magn Reson* 40:31–41. doi:[10.1016/j.ssnmr.2011.08.001](https://doi.org/10.1016/j.ssnmr.2011.08.001)
- Smith AN, Long JR (2016) Dynamic nuclear polarization as an enabling Technology for Solid State Nuclear Magnetic Resonance Spectroscopy. *Anal Chem* 88:122–132. doi:[10.1021/acs.analchem.5b04376](https://doi.org/10.1021/acs.analchem.5b04376)
- Akbey Ü, Oschkinat H (2016) Structural biology applications of solid state MAS DNP NMR. *J Magn Reson* 269:213–224. doi:[10.1016/j.jmr.2016.04.003](https://doi.org/10.1016/j.jmr.2016.04.003)
- Bruun S, Stoeppler D, Keidel A et al (2015) Light-dark adaptation of channelrhodopsin involves photoconversion between the all-trans and 13-cis retinal isomers. *Biochemistry* 54:5389–5400. doi:[10.1021/acs.biochem.5b00597](https://doi.org/10.1021/acs.biochem.5b00597)

15. Salnikov ES, Aisenbrey C, Aussenac F et al (2016) Membrane topologies of the PGLa antimicrobial peptide and a transmembrane anchor sequence by dynamic nuclear polarization/solid-state NMR spectroscopy. *Sci Rep* 6:20895. doi:[10.1038/srep20895](https://doi.org/10.1038/srep20895)
16. Yamamoto K, Caporini MA, Im S-C et al (2015) Cellular solid-state NMR investigation of a membrane protein using dynamic nuclear polarization. *Biochim Biophys Acta Biomembr* 1848:342–349. doi:[10.1016/j.bbmem.2014.07.008](https://doi.org/10.1016/j.bbmem.2014.07.008)
17. Williams JK, Tietze D, Lee M et al (2016) Solid-state NMR investigation of the conformation, proton conduction, and hydration of the influenza B virus M2 transmembrane proton channel. *J Am Chem Soc* 138:8143–8155. doi:[10.1021/jacs.6b03142](https://doi.org/10.1021/jacs.6b03142)
18. Frederick KK, Michaelis VK, Corzilius B et al (2015) Sensitivity-enhanced NMR reveals alterations in protein structure by cellular milieu. *Cell* 163:620–628. doi:[10.1016/j.cell.2015.09.024](https://doi.org/10.1016/j.cell.2015.09.024)
19. Fricke P, Mance D, Chevelkov V et al (2016) High resolution observed in 800 MHz DNP spectra of extremely rigid type III secretion needles. *J Biomol NMR* 65:121–126. doi:[10.1007/s10858-016-0044-y](https://doi.org/10.1007/s10858-016-0044-y)
20. Torrezan AC, Shapiro MA, Sirigiri JR et al (2011) Operation of a continuously frequency-tunable second-harmonic CW 330-GHz Gyrotron for dynamic nuclear polarization. *IEEE Trans Electron Devices* 58:2777–2783. doi:[10.1109/TED.2011.2148721](https://doi.org/10.1109/TED.2011.2148721)
21. Wollan DS (1976) Dynamic nuclear polarization with an inhomogeneously broadened ESR line. II. Experiment. *Phys Rev B* 13:3686–3696
22. Song C, Hu K-NN, Joo C-GG et al (2006) TOTAPOL: a biradical polarizing agent for dynamic nuclear polarization experiments in aqueous media. *J Am Chem Soc* 128:11385–11390. doi:[10.1021/ja061284b](https://doi.org/10.1021/ja061284b)
23. Wenk P, Kaushik M, Richter D et al (2015) Dynamic nuclear polarization of nucleic acid with endogenously bound manganese. *J Biomol NMR* 63:97–109. doi:[10.1007/s10858-015-9972-1](https://doi.org/10.1007/s10858-015-9972-1)
24. Mathies G, Caporini MA, Michaelis VK et al (2015) Efficient dynamic nuclear polarization at 800 MHz/527 GHz with Trityl-Nitroxide Biradicals. *Angew Chem Int Ed* 54:11770–11774. doi:[10.1002/anie.201504292](https://doi.org/10.1002/anie.201504292)
25. Zagdoun A, Rossini AJ, Conley MP et al (2013) Improved dynamic nuclear polarization surface-enhanced NMR spectroscopy through controlled incorporation of deuterated functional groups. *Angew Chem Int Ed* 52:1222–1225. doi:[10.1002/anie.201208699](https://doi.org/10.1002/anie.201208699)
26. Barnes AB, De Paëpe G, van der Wel PCA et al (2008) High-field dynamic nuclear polarization for solid and solution biological NMR. *Appl Magn Reson* 34:237–263. doi:[10.1007/s00723-008-0129-1](https://doi.org/10.1007/s00723-008-0129-1)
27. Elisei E, Filibian M, Carretta P et al (2015) Dynamic nuclear polarization of a glassy matrix prepared by solid state mechanochemical amorphization of crystalline substances. *Chem Commun* 51:2080–2083. doi:[10.1039/C4CC08348B](https://doi.org/10.1039/C4CC08348B)
28. Schrader AM, Cheng CY, Israellachvili JN, Han S (2016) Communication: contrasting effects of glycerol and DMSO on lipid membrane surface hydration dynamics and forces. *J Chem Phys* 145(4):041101. doi:[10.1063/1.4959904](https://doi.org/10.1063/1.4959904)
29. Matsuki Y, Takahashi H, Ueda K et al (2010) Dynamic nuclear polarization experiments at 14.1 T for solid-state NMR. *Phys Chem Chem Phys* 12:5799–5803. doi:[10.1039/C002268C](https://doi.org/10.1039/C002268C)
30. Meyer B (1971) *Low temperature spectroscopy*. American Elsevier Publ. Comp, New York
31. Gupta R, Lu M, Hou G et al (2016) Dynamic nuclear polarization enhanced MAS NMR spectroscopy for structural analysis of HIV-1 protein assemblies. *J Phys Chem B* 120:329–339. doi:[10.1021/acs.jpcc.5b12134](https://doi.org/10.1021/acs.jpcc.5b12134)
32. Sergeev IV, Itin B, Rogawski R, Day LA, McDermott AE (2017) Efficient assignment and NMR analysis of an intact virus using sequential side-chain correlations and DNP sensitization. *Proc Natl Acad Sci USA* 114 (20):5171–5176. doi:[10.1073/pnas.1701484114](https://doi.org/10.1073/pnas.1701484114)
33. Chaudhari S, Berruyer P, Gajan D et al (2016) Dynamic nuclear polarization at 40 kHz magic angle spinning. *Phys Chem Chem Phys* 18:10616–10622. doi:[10.1039/C6CP00839A](https://doi.org/10.1039/C6CP00839A)
34. Kaplan M, Cukkemane A, van Zundert GCP et al (2015) Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat Methods* 12:5–9. doi:[10.1038/nmeth.3406](https://doi.org/10.1038/nmeth.3406)
35. Kubicki DJ, Casano G, Schwarzwälder M et al (2016) Rational design of dinitroxide biradicals for efficient cross-effect dynamic nuclear polarization. *Chem Sci* 7:550–558. doi:[10.1039/C5SC02921J](https://doi.org/10.1039/C5SC02921J)
36. Le D, Ziarelli F, Phan TNT et al (2015) Up to 100% improvement in dynamic nuclear polarization solid-state NMR sensitivity enhancement of polymers by removing oxygen. *Macromol Rapid Commun* 36:1416–1421. doi:[10.1002/marc.201500133](https://doi.org/10.1002/marc.201500133)
37. Kubicki DJ, Rossini AJ, Porea A et al (2014) Amplifying dynamic nuclear polarization of frozen solutions by incorporating dielectric particles. *J Am Chem Soc* 136:15711–15718. doi:[10.1021/ja5088453](https://doi.org/10.1021/ja5088453)

38. Liao SY, Lee M, Wang T et al (2016) Efficient DNP NMR of membrane proteins : sample preparation protocols , sensitivity , and radical location. *J Biomol NMR* 64(3):223–237. doi:[10.1007/s10858-016-0023-3](https://doi.org/10.1007/s10858-016-0023-3)
39. Cady S, Wang T, Hong M (2011) Membrane-dependent effects of a cytoplasmic helix on the structure and drug binding of the influenza virus M2 protein. *J Am Chem Soc* 133:11572–11579. doi:[10.1021/ja202051n](https://doi.org/10.1021/ja202051n)
40. Lesage A, Lelli M, Gajan D et al (2010) Surface enhanced NMR spectroscopy by dynamic nuclear polarization. *J Am Chem Soc* 132:15459–15461. doi:[10.1021/ja104771z](https://doi.org/10.1021/ja104771z)
41. Lelli M, Chaudhari SR, Gajan D et al (2015) Solid-state dynamic nuclear polarization at 9.4 and 18.8 T from 100 K to room temperature. *J Am Chem Soc* 137:14558–14561. doi:[10.1021/jacs.5b08423](https://doi.org/10.1021/jacs.5b08423)
42. Koers EJ, López-Deber MP, Weingarth M et al (2013) Dynamic nuclear polarization NMR spectroscopy: revealing multiple conformations in lipid-anchored peptide vaccines. *Angew Chem Int Ed* 52:10905–10908. doi:[10.1002/anie.201303374](https://doi.org/10.1002/anie.201303374)
43. Morcombe CR, Zilm KW (2003) Chemical shift referencing in MAS solid state NMR. *J Magn Reson* 162:479–486. doi:[10.1016/S1090-7807\(03\)00082-X](https://doi.org/10.1016/S1090-7807(03)00082-X)
44. Smith AA, Corzilius B, Barnes AB et al (2012) Solid effect dynamic nuclear polarization and polarization pathways. *J Chem Phys* 136(1):015101. doi:[10.1063/1.3670019](https://doi.org/10.1063/1.3670019)
45. Corzilius B, Andreas LB, Smith AA et al (2014) Paramagnet induced signal quenching in MAS-DNP experiments in frozen homogeneous solutions. *J Magn Reson* 240:113–123. doi:[10.1016/j.jmr.2013.11.013](https://doi.org/10.1016/j.jmr.2013.11.013)
46. Mentink-Vigier F, Paul S, Lee D et al (2015) Nuclear depolarization and absolute sensitivity in magic-angle spinning cross-effect dynamic nuclear polarization. *Phys Chem Chem Phys* 17(34):21824–21836. doi:[10.1039/C5CP03457D](https://doi.org/10.1039/C5CP03457D)
47. Takahashi H, Fernández-De-Alba C, Lee D et al (2014) Optimization of an absolute sensitivity in a glassy matrix during DNP-enhanced multidimensional solid-state NMR experiments. *J Magn Reson* 239:91–99. doi:[10.1016/j.jmr.2013.12.005](https://doi.org/10.1016/j.jmr.2013.12.005)

In-Vitro Dissolution Dynamic Nuclear Polarization for Sensitivity Enhancement of NMR with Biological Molecules

Yaewon Kim, Yunyi Wang, Hsueh-Ying Chen, and Christian Hilty

Abstract

Dissolution dynamic nuclear polarization (D-DNP) is a technique to prepare hyperpolarized nuclear spin states, yielding a signal enhancement of several orders of magnitude for liquid-state NMR. Here, we describe experimental procedures for the application of D-DNP in high-resolution NMR of biochemical compounds, to determine the time evolution of biochemical processes and intermolecular interactions.

Key words Hyperpolarization, Real-time NMR, Kinetics, Spin correlations, Spin relaxation

1 Introduction

Because of a low equilibrium population difference of nuclear spin Zeeman energy levels, the fraction of molecules in a sample that contributes to the net NMR signal is typically only on the order of 10^{-5} . The possible applications of NMR spectroscopy that can be opened up by increasing this fraction to become closer to unity, thereby dramatically increasing NMR signal, are many-fold. Several methods for creating non-equilibrium, hyperpolarized spin states with increased population difference are known. Such techniques include the polarization of noble gases by optical pumping [1], para-hydrogen induced polarization [2], chemically induced dynamic nuclear polarization (CIDNP) [3], dynamic nuclear polarization (DNP) [4], and optical polarization of nitrogen vacancy centers in diamond [5]. None of these hyperpolarization techniques is suited for general use because of specific requirements on the types and properties of samples. However, when applicable, they can enable NMR spectroscopy under conditions that would otherwise yield no observable signal. For work in the liquid state, in particular with dilute samples, dissolution dynamic nuclear polarization (D-DNP) has emerged as a versatile technique capable of generating high spin polarization. D-DNP was originally

demonstrated by Ardenkjaer-Larsen et al. [6] with ^{13}C sensitivity enhancements of a factor of over 10^4 in a single scan compared to a conventional NMR spectrum. It has since then been applied for metabolic imaging [7], as well as for high-resolution NMR spectroscopy. Briefly, D-DNP involves the generation of nuclear spin hyperpolarization, a state of increased relative Zeeman population difference, in the solid-state at a temperature on the order of 1 K. Microwave saturation of an electron spin transition of a free radical in effect causes polarization transfer from the electron spins to nuclear spins. This hyperpolarization step is followed by dissolution of the sample into a heated solvent and injection into an NMR spectrometer for signal acquisition. Multiple nuclei such as ^{13}C , ^1H , ^{19}F , and others can be hyperpolarized, with the primary limitation that the lifetime of the hyperpolarization is governed by the spin-lattice relaxation time (T_1) of the nuclei in question, typically on the order of seconds in the liquid state (Fig. 1). The availability of a high signal during this time, however, lends itself in particular to the study of time-dependent chemical or biochemical processes, as well as of spin dynamics. D-DNP is compatible with biochemically relevant conditions, including molecules at final concentrations in the micromolar or sub-micromolar range, in aqueous buffered solutions. It has been applied to problems such as cellular metabolism, the kinetics and mechanisms of catalyzed or uncatalyzed reactions, protein-ligand interactions, and protein folding. Here, we summarize experimental protocols for the application of high-resolution NMR with D-DNP to these and other problems in the biochemical sciences.

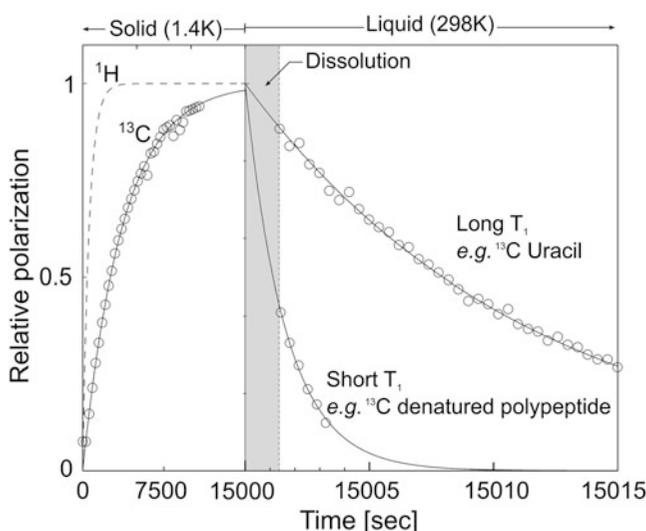


Fig. 1 A composite diagram showing the build-up curves of ^1H and ^{13}C polarization in the solid-state at 1.4 K (0–15,000 s) and hyperpolarization lifetime of two molecules of different T_1 relaxation times in the liquid state (15,000–15,015 s). It is noted that the signal in the solid and liquid states cannot be compared directly. The gray strip indicates the dead time of the NMR measurement, required for dissolution and delivery of hyperpolarized samples

2 Materials

2.1 Stock Solutions

Prepare molecules of interest with a polarizing agent in a glassing matrix, which causes the sample to vitrify when frozen, as described below.

2.1.1 Molecules of Interest

The DNP techniques can be applied to hyperpolarizing various classes of molecules, from low molecular weight compounds to macromolecules, including polypeptides [8, 9].

2.1.2 Polarizing Agents

Typical polarizing agents include nitroxides (2,2,6,6-tetramethylpiperidine-1-oxyl: TEMPO and 4-hydroxy-2,2,6,6-tetramethylpiperidine-1-oxyl: TEMPOL [10]) for ^1H and ^{19}F polarization, trityl radicals (tris(8-carboxy-2,2,6,6-tetra(hydroxyethyl)benzo[1,2-d;4,5-d']-bis-(1,3)-dithiol-4-yl)methyl sodium salt: OXO63 [11] and tris(8-carboxy-2,2,6,6-tetramethylbenzo[1,2-d;4,5-d']-bis-(1,3)-dithiol-4-yl)methyl sodium salt: Finland), 1,3-bisdiphenylene-2-phenylallyl (BDPA) [12] and water-soluble sulfonated BDPA (SA-BDPA) [13] for nuclei such as ^{13}C , ^{15}N , or ^{31}P . Radical concentration is often chosen between 15 ~ 30 mM for optimal polarization enhancement [14]. A small amount of lanthanide ions such as Gd^{3+} can be used to dope the samples to enhance the polarization level (*see Note 1*). The choice of the polarizing agent depends both on the solubility of the substrate and the identity of the nucleus to be polarized (*see Note 2*) (Fig. 2).

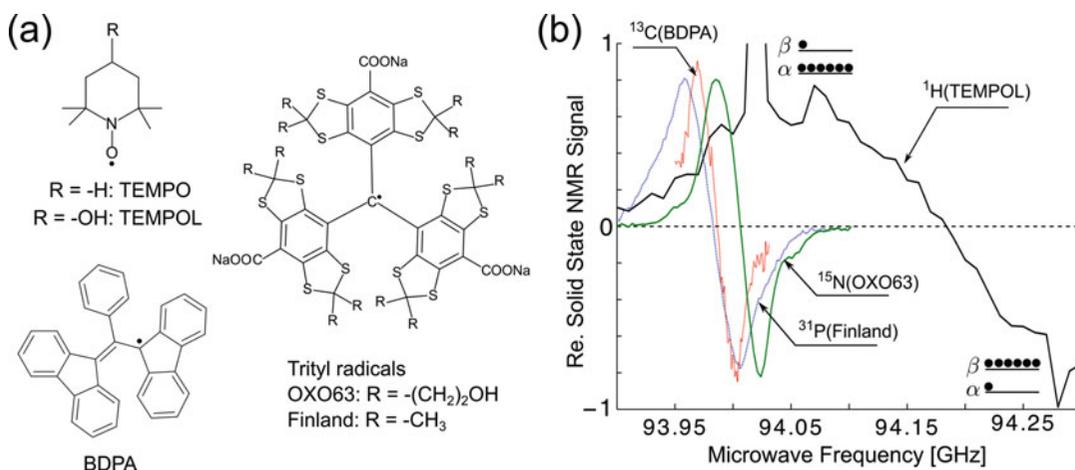


Fig. 2 (a) Typical radicals used for hyperpolarization. (b) Microwave frequency dependence (“sweep”) of selected nuclei with different radicals. The spin population can be manipulated by microwave frequency. The microwave frequency at maximum (or minimum) NMR signal can be selected to obtain positive (or negative) signal enhancement as indicated by the population difference shown in the two energy levels (α and β)

2.1.3 Glassing Matrix

Successful hyperpolarization of the target nuclei requires that the sample freezes to form a glass. Depending on the solubility of the molecule of interest, a mixture of two or more solvents can be used. Mixtures of glycerol/water (v/v 60:40), ethylene glycol/water (v/v 60:40), water/ethanol (v/v 60:40), and water/dimethyl sulfoxide (DMSO) (v/v 50:50, 60:40) are possible glassing matrixes for water-soluble samples. When using self-glassing liquids, such as neat pyruvic acid [7] or 1-butanol [15], no additional agent is needed. A list of self-glassing liquids and glass-forming mixtures can be found in [16].

2.1.4 Isotope Enrichment

For polarization of carbon or other nuclei with low natural abundance of the NMR active isotopes (such as ^{13}C or ^{15}N), enrichment of this isotope is not required but can substantially increase the signal. For ^1H polarization, use deuterated glassing matrix to avoid large ^1H solvent signals which are also enhanced by hyperpolarization. For example, mixtures of d_6 -DMSO and D_2O or d_6 -ethylene glycol and D_2O can be chosen as the glassing solvent.

2.1.5 Special Sample Preparations

1. Photoinduced non-persistent radicals, which are only stable at low temperature during polarization and disappear when dissolved in a hot solvent, can be used so that the effect of accelerated spin relaxation due to the presence of radicals can be eliminated during the dissolution and transfer processes [17, 18].
2. DNP samples may also be prepared in a solid polarizing matrix. In such a design, radicals are immobilized on solid materials, which can be wetted with a solution of DNP substrates for hyperpolarization and easily separated by filtration afterward [19–21].
3. Frozen beads containing radical scavengers (e.g., ascorbic acid) can be added to the stock solution to remove the polarizing agent (e.g., TEMPOL) during the dissolution, and to extend the lifetime of hyperpolarization on the molecule of interest [22].

2.2 Dissolution Solvent

Degas the dissolution solvent before use (*see Note 3*). Besides dissolving hyperpolarized sample, the dissolution solvent can also be chosen to provide other functions:

1. Use a buffer solution to adjust the pH of the final mixture.
2. Select the sample conditions based on the choice of dissolution solvent, e.g., a denaturing agent can be added to the dissolution solvent to denature a hyperpolarized protein [8].
3. Use the dissolution solvent as a reactant or ensure that it provides a proper environment to allow a reaction to proceed, which can then be observed by NMR.

- Use the buffer solution to eliminate radicals in the hyperpolarized sample. Chelating agents such as ethylenediaminetetraacetic acid (EDTA) can be added to the buffer for binding to paramagnetic ions. Two-phase systems including solvents that are miscible and immiscible with the sample can be used to remove the radicals and to concentrate the hyperpolarized sample into a smaller volume of dissolution solvent [23].

2.3 Equipment

As shown in Fig. 3, components needed for D-DNP NMR experiments include a DNP polarizer, a sample injector, and an NMR spectrometer. Samples are hyperpolarized in the DNP polarizer and transferred to the NMR spectrometer for NMR measurement immediately after dissolution, by vapor pressure of the heated dissolution solvent or using an additional sample injector for more rapid injection.

2.3.1 DNP Polarizer

The commercially available HyperSense DNP polarizer (Oxford Instruments, Abingdon, UK) contains a super-conducting magnet (3.35 T), a cryostat, a DNP insert, a vacuum pump to maintain low helium vapor pressure, a microwave source, and a dissolution system. DNP polarizers that operate at higher magnetic fields (4.64 T [25] and 6.7 T [26]), as well as combined NMR and DNP magnets [27], have also been developed.

2.3.2 Sample Injector

Unassisted transfer time for the hyperpolarized sample from the polarizer to the NMR spectrometer for measurement may require up to 6 s, during which there is a significant polarization loss caused by T_1 relaxation. The transfer time can be reduced to 1 ~ 2 s by means of an additional sample injector driven by pressurized N_2 or Ar gas [28–30] or liquid [24]. During injection, the sample traverses a region of lower magnetic field outside the magnet (Fig. 4).

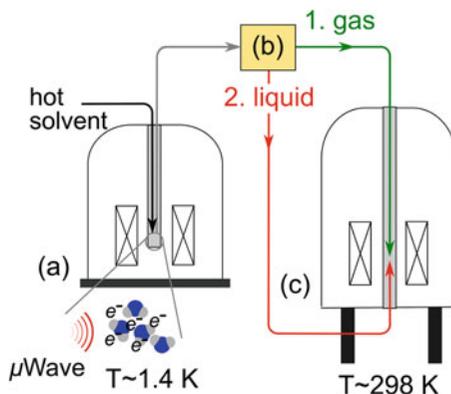


Fig. 3 Schematic of a DNP polarizer with a sample transfer line to an NMR spectrometer. (a) DNP polarizer. (b) Sample injector (gas or liquid driven). (c) NMR spectrometer (re-printed with modification, with permission from ref. 24)

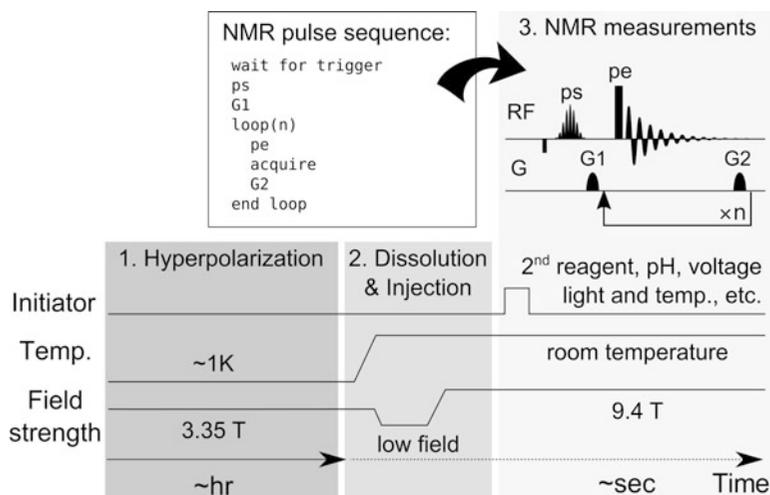


Fig. 4 The experimental procedure for D-DNP NMR includes three steps: 1. Hyperpolarization; 2. Dissolution and Injection; 3. NMR measurements. An initiator can trigger a biochemical process that can be monitored in real time. Field strengths are based on a 3.35 T DNP polarizer and a 400 MHz NMR spectrometer. Radio frequency (RF) pulses are denoted as “ps” and “pe.” The “ps” can be applied for the suppression of a resonance from solvent or glassy matrix, or inversion of hyperpolarized signals as desired, followed by a pulsed field gradient (G1). A series of NMR spectra are acquired multiple times after an excitation pulse “pe” of a small flip angle is applied. The following pulsed field gradient (G2) is used to remove unwanted coherence prior to the next acquisition

For preserving the hyperpolarization, it is important that the field strength does not drop to zero. A magnetic tunnel has been proposed to transfer hyperpolarized samples to minimize losses when samples are transferred over long distances [31].

2.3.3 NMR Spectrometer

A high-resolution NMR spectrometer may be used, equipped with a standard 5 mm or 10 mm probe. Alternately, a flow-NMR probe may be used together with liquid-driven injection.

3 Methods

A dissolution DNP experiment generally consists of three steps (Fig. 4). First, the sample is irradiated with microwaves in the DNP polarizer at a temperature of 1 ~ 2 K. Then, the hyperpolarized sample is dissolved with preheated solvent and injected to the NMR magnet. NMR measurements are performed, once the injection has finished. Due to the rapid nature of D-DNP experiments, all parameters for sample injection and NMR measurements, including NMR shims, need to be preset.

3.1 Preparing for DNP Experiments

1. Connect the NMR tube to the end of the transfer line. If using a flow-NMR system, connect the line to the flow cell and check for leaks while the liquid is flowing.

2. Load 4 ~ 5 mL of the degassed dissolution solvent into the D-DNP polarizer. Perform test injections to set injection time, pressure, and other parameters such that the desired sample volume is delivered to the spectrometer and the sample condition is stable (e.g., no air bubbles). Detailed procedures for using an additional gas-driven and liquid-driven injector can be found in [24, 28].
3. If the hyperpolarized sample needs to be mixed with a second reagent, which initiates a reaction in-situ in the NMR spectrometer, preload it into the NMR tube (or to a sample loop for the flow-NMR system) and perform a test injection. Mixing efficiency can be checked by acquiring images with a pulsed field gradient spin echo (PFGSE) sequence [32], or performing known test reactions similar to the setup of a stopped-flow experiment [33]. Using dye as a substitute for the second reagent can be convenient to visually check homogeneous mixing.
4. Tune and shim the NMR spectrometer with a sample from a test injection, and set the sample target temperature to that of the injected solution of the hyperpolarized sample. (For the procedure to measure and adjust the temperature of injected solution from the polarizer, *see* **Note 4**).
5. Conduct an NMR pulse length calibration using the same sample.
6. The amount of originally loaded sample that is recovered for the NMR experiment and final dilution factors for the sample from the polarizer and the second reagent can be determined by using reference standards added to the samples. For calibration purposes, the measurement of the concentration of reference standards in the final sample can be done either by NMR spectroscopy or other analytical techniques, such as UV-VIS spectroscopy or high-performance liquid chromatography [28].

3.2 Performing D-DNP Experiments

3.2.1 Hyperpolarization

1. Cool down the DNP polarizer, to reach the operating temperature close to 1 K.
2. Load the properly made sample solution to the cup-shaped sample holder and insert the sample cup into the DNP instrument (*see* **Note 5**). The sample volume is typically in the range of 1 ~ 100 μL , but different sizes of sample cups can be used to accommodate even larger sample volumes [34].
3. Start polarization by turning on the microwave irradiation. The microwave frequency should be chosen depending on the radical type and nucleus to be polarized (Fig. 2).
4. Continue polarizing the sample. The polarization time required to approach the maximum polarization level depends on microwave power, type of nucleus, type and concentration

of radical, glassing matrix [35, 36], and other experimental conditions. For example, ^1H and ^{19}F may be polarized in about 10 min using TEMPOL radical, ^{13}C may be polarized for 3 h using trityl radical. The optimal polarization time can be determined by measuring a solid-state polarization build-up curve (*see Note 6*). The efficiency of polarization can be improved by polarizing multiple samples [37, 38] or applying Hartmann-Hahn cross polarization (CP) to D-DNP for low- γ nuclei (e.g. ^{13}C) polarization using specially made DNP polarizers [39–41] (*see Note 7*).

3.2.2 Sample Dissolution and Injection

1. Load about 4 ~ 5 mL of degassed solvent for dissolution.
2. Connect the transfer line to the NMR tube (or flow-cell) the same way as for the test injections. Adjust the position of the NMR tube according to the sample volume from the test injection. Then, place it into the NMR spectrometer (*see Note 5* when using high-pressure gas for rapid sample injection).
3. Set a target vapor pressure and start heating the dissolution solvent (*see Note 4*). When the target vapor pressure is achieved, the automated dissolution process is triggered. A typical target vapor pressure is 5 ~ 10 bar.

3.2.3 NMR Measurements

1. Before dissolution, start the desired NMR pulse program (Fig. 4) that includes a command for waiting for the trigger from the dissolution. This can be done while the dissolution solvent is being heated. The measurement will be initiated by the trigger.
2. After the DNP measurement is finished, acquire an NMR spectrum under thermal polarization for the observation of reference standards in the sample, for the calculation of spin polarization, or for other characterizations of the final sample conditions.
3. If subsequent experiments are to be performed, the injected solution can be used for re-adjusting shims for the next experiment.

3.3 Type of NMR Experiments

Various types of D-DNP NMR experiments are available for characterization of biological systems in-vitro. A (necessarily incomplete) summary of some of these experimental techniques is given here.

3.3.1 Characterization of Non-equilibrium Chemical Reactions or Processes

1. For the purpose of investigating time-dependent signal changes in a D-DNP experiment, perform a series of scans, each containing a small flip angle excitation pulse. The pulse converts a fraction of available spin polarization into observable coherence in each scan, while preserving the remaining polarization for later scans. The flip angle is adjusted based on the

number of scans and T_1 relaxation times of the spins [33]. The time resolution is determined by the sampling requirement to obtain sufficient spectral resolution (≥ 50 ms), and the maximum observable time by the spin-lattice relaxation.

2. Selective excitation pulses followed by pulsed field gradients can be applied to suppress unwanted signals that would saturate the receiver. Selective inversion pulses can be used at the beginning of the experiment to encode a resonance of interest and track it during the measurement time [33]. The selectively inverted spin population leads to an inverted product signal, which allows the correlation of reaction product chemical shifts to the corresponding chemical shifts in the reactant.
3. For kinetic analysis, signal intensities of individual peaks from time-resolved spectra can be plotted as a function of time. Kinetic parameters can be determined by fitting the data using a kinetic model that includes a description of the effects of radio frequency pulses and spin relaxation [33].
4. Utilizing the methods above, real-time monitoring of chemical or biological processes such as enzyme reactions [32, 42], protein folding [8], and cellular metabolism [43–45] can be performed. For the study of enzyme reactions, substrates are hyperpolarized, and the NMR signals are acquired after admixing of non-polarized enzymes. The protein-folding process can be observed by injecting hyperpolarized denatured polypeptides into a buffer solution favoring native structure. In cell metabolism studies, metabolites, such as pyruvate, glucose, fructose, fumarate, or other compounds, are hyperpolarized, and mixed with live-cell suspensions in the NMR spectrometer [9]. In all the cases, observing the time-dependent evolution of signals allows determination of kinetics or mechanisms of the underlying process.

3.3.2 Time Evolution in Non-equilibrium Spin States

Several DNP–NMR methods comprising measurements of spin relaxation or spin polarization transfer can be used for investigating interactions between a target molecule such as a protein, and an interaction partner such as a ligand or solvent component. For example, hyperpolarization can be carried out on ^{13}C , ^1H , or ^{19}F nuclear spins of ligand and the evolution of NMR signals of the ligand or a target protein, the latter occurring due to the nuclear Overhauser effect (NOE), can be observed.

1. Relaxation measurements on hyperpolarized samples can be performed in a single dissolution. T_1 relaxation of low natural abundance nuclei can be measured using a series of small flip-angle RF pulses [46]. T_2 relaxation decay of hyperpolarized ligands can be obtained by employing a Carr Purcell Meiboom Gill (CPMG) pulse sequence [47] to identify binding of ligands

to proteins and determine a dissociation constant (K_D), at micro- or sub-micromolar concentrations [48]. Long-lived spin states (LLS) of ligands can under certain conditions also be created to preserve spin polarization during sample injection [49]. The relaxation of LLS can further be measured to detect ligand binding [50]. Competitive binding using a reporter ligand can be utilized more widely in the T_2 and T_{LLS} measurements, since it requires only a well polarizable reporter ligand. [51, 52]. $T_{1\rho}$ relaxation can be also measured using spin-lock [22]. The T_1 , T_2 , T_{LLS} , and $T_{1\rho}$ relaxation time constants can be determined by fitting the data to a single exponential or another appropriate function.

2. The transfer of ligand hyperpolarization to the target molecule or to a second ligand via NOE can be used to study intermolecular interactions between the protein and the ligand. Signal build-up and decay allows the characterization of binding sites of proteins and ligands [53–56]. Instead of hyperpolarizing ligands, hyperpolarization of water can be utilized to identify ligand binding [57] by transferring polarization to the protein through exchangeable protons and NOE. This method can be used to study protein–water interactions and to potentially investigate exchange rates or cross relaxation rates [58–60], providing structural information.

3.3.3 Spin-Correlation and Other Experiments

Multiple experiments are available for obtaining 2D or pseudo-2D correlation information in a single D-DNP experiment. Multi-dimensional D-DNP NMR spectra can be acquired from sequential 2D experiments [61], ultrafast COSY/HSQC experiments [62], or 2D slice-selective COSY experiments [63]. Multiple-scan DNP experiments using flow injection combined with Hadamard spectroscopy [15] or off-resonance decoupling schemes [64] can be also utilized for investigating atomic connectivity or the measurement of J-coupling constants. D-DNP is also applicable to diffusion experiments, allowing a rapid measurement of diffusion coefficients without signal averaging [65, 66]. Diffusion and relaxation information can be obtained simultaneously using ultrafast Laplace NMR correlation experiments [67].

4 Notes

1. Gd^{3+} compounds, such as gadolinium diethylene triamine pentaacetic (Gd-DTPA), can be added to the stock solution with a concentration of 1–2 mM. Gd^{3+} doping has been found to improve the achievable polarization level, especially for ^{13}C polarization using OXO63 as the polarizing agent [68, 69].

2. TEMPOL, OXO63, Finland, and water-soluble BDPA can be used for aqueous solutions, while TEMPO or BDPA can be used for organic solvents.
3. Hazardous liquids, such as those that are toxic, those that result in toxic by-products upon heating, or those that could violently decompose, should be avoided.
4. The temperature of the injected sample can be measured by carrying out test injections. For test injections, use the same dissolution solvent as for the experiments and set the desired vapor pressure for the dissolution. The target vapor pressure can be chosen to adjust the temperature of the injected sample, within limits. The temperature can be measured using a thermocouple, or NMR chemical shift.
5. Wear appropriate personal protection equipment when loading a sample into the DNP insert, as well as when working with high-pressure gas for injection.
6. Microwave sweep spectra should be recorded periodically for microwave frequency calibration. The time dependence of polarization build-up in the solid state can also be measured to determine the appropriate polarization time.
7. The efficiency of polarizing low- γ nuclei, such as ^{13}C , can be significantly improved by combining Hartmann-Hahn cross polarization (CP) with dissolution DNP. Transferring polarization from electron spins to protons, and further to ^{13}C nuclei, has been shown to reduce the polarization build-up time, as well as to achieve higher polarization level [39–41].

References

1. Walker TG, Happer W (1997) Spin-exchange optical pumping of noble-gas nuclei. *Rev Mod Phys* 69:629–642
2. Bowers CR, Weitekamp DP (1986) Transformation of symmetrization order to nuclear-spin magnetization by chemical reaction and nuclear magnetic resonance. *Phys Rev Lett* 57: 2645–2648
3. Bargon J, Fischer H (2014) Kernresonanz-Emissionslinien während rascher Radikalreaktionen. *Z Naturforsch A* 22:1556–1562
4. Abragam A, Goldman M (1978) Principles of dynamic nuclear polarisation. *Rep Prog Phys* 41: 395
5. Jacques V, Neumann P, Beck J, Markham M, Twitchen D, Meijer J, Kaiser F, Balasubramanian G, Jelezko F, Wrachtrup J (2009) Dynamic polarization of single nuclear spins by optical pumping of nitrogen-vacancy color centers in diamond at room temperature. *Phys Rev Lett* 102:57403
6. Ardenkjær-Larsen JH, Fridlund B, Gram A, Hansson G, Hansson L, Lerche MH, Servin R, Thaning M, Golman K (2003) Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR. *Proc Natl Acad Sci USA* 100: 10158–10163
7. Golman K, in't Zandt R, Thaning M (2006) Real-time metabolic imaging. *Proc Natl Acad Sci USA* 103:11270–11275
8. Chen H-Y, Ragavan M, Hilty C (2013) Protein folding studied by dissolution dynamic nuclear polarization. *Angew Chem Int Ed* 52: 9192–9195
9. Meier S, Jensen PR, Karlsson M, Lerche MH (2014) Hyperpolarized NMR probes for biological assays. *Sensors* 14:1576–1597
10. Höfer P, Carl P, Guthausen G, Prisner T, Reese M, Carlomagno T, Griesinger C, Bennati M (2008) Studies of dynamic nuclear polarization with nitroxides in aqueous solution. *Appl Magn Reson* 34:393–398

- Ardenkjær-Larsen JH, Laursen I, Leunbach I, Ehnholm G, Wistrand L-G, Petersson JS, Goldman K (1998) EPR and DNP properties of certain novel single electron contrast agents intended for oximetric imaging. *J Magn Reson* 133:1–12
- Lumata L, Ratnakar SJ, Jindal A, Merritt M, Comment A, Malloy C, Sherry AD, Kovacs Z (2011) BDPA: an efficient polarizing agent for fast dissolution dynamic nuclear polarization NMR spectroscopy. *Chem Eur J* 17:10825–10827
- Haze O, Corzilius B, Smith AA, Griffin RG, Swager TM (2012) Water-soluble narrow-line radicals for dynamic nuclear polarization. *J Am Chem Soc* 134:14287–14290
- Zhu Y, Chen C-H, Wilson Z, Savukov I, Hilty C (2016) Milli-tesla NMR and spectrophotometry of liquids hyperpolarized by dissolution dynamic nuclear polarization. *J Magn Reson* 270:71–76
- Chen H-Y, Hilty C (2013) Hyperpolarized hadamard spectroscopy using flow NMR. *Anal Chem* 85:7385–7390
- Drago RS (1977) In: Drago RS (ed) *Physical methods in chemistry*. Saunders, Philadelphia
- Capozzi A, Hyacinthe J-N, Cheng T, Eichhorn TR, Boero G, Roussel C, van der Klink JJ, Comment A (2015) Photoinduced nonpersistent radicals as polarizing agents for X-nuclei dissolution dynamic nuclear polarization. *J Phys Chem C* 119:22632–22639
- Eichhorn TR, Takado Y, Salameh N, Capozzi A, Cheng T, Hyacinthe J-N, Mishkovsky M, Roussel C, Comment A (2013) Hyperpolarization without persistent radicals for in vivo real-time metabolic imaging. *Proc Natl Acad Sci USA* 110:18064–18069
- Dollmann BC, Junk MJN, Drechsler M, Spiess HW, Hinderberger D, Münnemann K (2010) Thermoresponsive, spin-labeled hydrogels as separable DNP polarizing agents. *Phys Chem Chem Phys* 12:5879
- McCarney ER, Han S (2008) Spin-labeled gel for the production of radical-free dynamic nuclear polarization enhanced molecules for NMR spectroscopy and imaging. *J Magn Reson* 190:307–315
- Gajan D, Bornet A, Vuichoud B et al. (2014) Hybrid polarizing solids for pure hyperpolarized liquids through dissolution dynamic nuclear polarization. *Proc Natl Acad Sci USA* 111:14693–14697
- Miéville P, Ahuja P, Sarkar R et al (2010) Scavenging free radicals to preserve enhancement and extend relaxation times in NMR using dynamic nuclear polarization. *Angew Chem Int Ed* 49:6182–6185
- Harris T, Bretschneider C, Frydman L (2011) Dissolution DNP NMR with solvent mixtures: substrate concentration and radical extraction. *J Magn Reson* 211:96–100
- Chen H-Y, Hilty C (2015) Implementation and characterization of flow injection in dissolution dynamic nuclear polarization NMR spectroscopy. *ChemPhysChem* 16:2646–2652
- Jóhannesson H, Macholl S, Ardenkjær-Larsen JH (2009) Dynamic nuclear polarization of [$1\text{-}^{13}\text{C}$]pyruvic acid at 4.6 tesla. *J Magn Reson* 197:167–175
- Jannin S, Bornet A, Melzi R, Bodenhausen G (2012) High field dynamic nuclear polarization at 6.7 T: Carbon-13 polarization above 70% within 20 min. *Chem Phys Lett* 549:99–102
- Leggett J, Hunter R, Granwehr J, Panek R, Perez-Linde AJ, Horsewill AJ, McMaster J, Smith G, Köckenberger W (2010) A dedicated spectrometer for dissolution DNP NMR spectroscopy. *Phys Chem Chem Phys* 12:5883–5892
- Bowen S, Hilty C (2010) Rapid sample injection for hyperpolarized NMR spectroscopy. *Phys Chem Chem Phys* 12:5766–5770
- Katsikis S, Marin-Montesinos I, Pons M, Ludwig C, Günther UL (2015) Improved stability and spectral quality in ex situ dissolution DNP using an improved transfer device. *Appl Magn Reson* 46:723–729
- Granwehr J, Panek R, Leggett J, Köckenberger W (2010) Quantifying the transfer and settling in NMR experiments with sample shuttling. *J Chem Phys* 132:244507
- Milani J, Vuichoud B, Bornet A, Miéville P, Mottier R, Jannin S, Bodenhausen G (2015) A magnetic tunnel to shelter hyperpolarized fluids. *Rev Sci Instrum* 86:24101
- Bowen S, Hilty C (2008) Time-resolved dynamic nuclear polarization enhanced NMR spectroscopy. *Angew Chem Int Ed* 47:5235–5237
- Zeng H, Lee Y, Hilty C (2010) Quantitative rate determination by dynamic nuclear polarization enhanced NMR of a Diels–Alder reaction. *Anal Chem* 82:8897–8902
- Bowen S, Ardenkjær-Larsen JH (2014) Enhanced performance large volume dissolution-DNP. *J Magn Reson* 240:90–94
- Ludwig C, Marin-Montesinos I, Saunders MG, Günther UL (2010) Optimizing the polarization matrix for ex situ dynamic nuclear polarization. *J Am Chem Soc* 132:2508–2509
- Lumata L, Merritt ME, Kovacs Z (2013) Influence of deuteration in the glassing matrix on ^{13}C dynamic nuclear polarization. *Phys Chem Chem Phys* 15:7032

37. Batel M, Krajewski M, Weiss K et al (2012) A multi-sample 94 GHz dissolution dynamic-nuclear-polarization system. *J Magn Reson* 214: 166–174
38. Hu S, Larson PEZ, VanCrickinge M et al. (2013) Rapid sequential injections of hyperpolarized [$1-^{13}\text{C}$]pyruvate in vivo using a subkelvin, multi-sample DNP polarizer. *Magn Reson Imaging* 31:490–496
39. Jannin S, Bornet A, Colombo S, Bodenhausen G (2011) Low-temperature cross polarization in view of enhancing dissolution dynamic nuclear polarization in NMR. *Chem Phys Lett* 517: 234–236
40. Bornet A, Melzi R, Perez Linde AJ, Hautle P, van den Brandt B, Jannin S, Bodenhausen G (2013) Boosting dissolution dynamic nuclear polarization by cross polarization. *J Phys Chem Lett* 4:111–114
41. Bornet A, Melzi R, Jannin S, Bodenhausen G (2012) Cross polarization for dissolution dynamic nuclear polarization experiments at readily accessible temperatures $1.2 < T < 4.2$ K. *Appl Magn Reson* 43:107–117
42. Jensen PR, Meier S, Ardenkjær-Larsen JH, Duus JØ, Karlsson M, Lerche MH (2009) Detection of low-populated reaction intermediates with hyperpolarized NMR. *Chem Commun* 34:5168–5170
43. Harris T, Eliyahu G, Frydman L, Degani H (2009) Kinetics of hyperpolarized $^{13}\text{C}_1$ -pyruvate transport and metabolism in living human breast cancer cells. *Proc Natl Acad Sci USA* 106: 18131–18136
44. Meier S, Karlsson M, Jensen PR, Lerche MH, Duus JØ (2011) Metabolic pathway visualization in living yeast by DNP-NMR. *Mol BioSyst* 7: 2834–2836
45. Gallagher FA, Kettunen MI, Hu D-E et al. (2009) Production of hyperpolarized [$1,4-^{13}\text{C}_2$] malate from [$1,4-^{13}\text{C}_2$] fumarate is a marker of cell necrosis and treatment response in tumors. *Proc Natl Acad Sci USA* 106:19801–19806
46. Day IJ, Mitchell JC, Snowden MJ, Davis AL (2007) Applications of DNP-NMR for the measurement of heteronuclear T_1 relaxation times. *J Magn Reson* 187:216–224
47. Meiboom S, Gill D (1958) Modified spin-Echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29:688–691
48. Lee Y, Zeng H, Ruedisser S, Gossert AD, Hilty C (2012) Nuclear magnetic resonance of hyperpolarized fluorine for characterization of protein–ligand interactions. *J Am Chem Soc* 134: 17448–17451
49. Warren WS, Jenista E, Branca RT, Chen X (2009) Increasing hyperpolarized spin lifetimes through true singlet Eigenstates. *Science* 323: 1711–1714
50. Buratto R, Bornet A, Milani J et al (2014) Drug screening boosted by hyperpolarized long-lived states in NMR. *ChemMedChem* 9: 2509–2515
51. Kim Y, Hilty C (2015) Affinity screening using competitive binding with fluorine-19 hyperpolarized ligands. *Angew Chem Int Ed* 54: 4941–4944
52. Buratto R, Mammoli D, Chiarparin E, Williams G, Bodenhausen G (2014) Exploring weak ligand-protein interactions by long-lived NMR states: improved contrast in fragment-based drug screening. *Angew Chem Int Ed* 53: 11376–11380
53. Min H, Sekar G, Hilty C (2015) Polarization transfer from ligands hyperpolarized by dissolution dynamic nuclear polarization for screening in drug discovery. *ChemMedChem* 10: 1559–1563
54. Wang Y, Ragavan M, Hilty C (2016) Site specific polarization transfer from a hyperpolarized ligand of dihydrofolate reductase. *J Biomol NMR* 65:41–48
55. Lee Y, Zeng H, Mazur A, Wegstroth M, Carlomagno T, Reese M, Lee D, Becker S, Griessinger C, Hilty C (2012) Hyperpolarized binding pocket nuclear overhauser effect for determination of competitive ligand binding. *Angew Chem Int Ed* 51:5179–5182
56. Lerche MH, Meier S, Jensen PR, Baumann H, Petersen BO, Karlsson M, Duus JØ, Ardenkjær-Larsen JH (2010) Study of molecular interactions with ^{13}C DNP-NMR. *J Magn Reson* 203:52–56
57. Chappuis Q, Milani J, Vuichoud B, Bornet A, Gossert AD, Bodenhausen G, Jannin S (2015) Hyperpolarized water to study protein–ligand interactions. *J Phys Chem Lett* 6:1674–1678
58. Harris T, Szekely O, Frydman L (2014) On the potential of hyperpolarized water in biomolecular NMR studies. *J Phys Chem B* 118: 3281–3290
59. Olsen G, Markhasin E, Szekely O, Bretschneider C, Frydman L (2016) Optimizing water hyperpolarization and dissolution for sensitivity-enhanced 2D biomolecular NMR. *J Magn Reson* 264:49–58
60. Kim J, Liu M, Chen H-Y, Hilty C (2015) Determination of intermolecular interactions using polarization compensated heteronuclear Overhauser effect of hyperpolarized spins. *Anal Chem* 87:10982–10987
61. Zeng H, Bowen S, Hilty C (2009) Sequentially acquired two-dimensional NMR spectra from hyperpolarized sample. *J Magn Reson* 199: 159–165
62. Frydman L, Blazina D (2007) Ultrafast two-dimensional nuclear magnetic resonance spectroscopy of hyperpolarized solutions. *Nat Phys* 3: 415–419

63. Panek R, Granwehr J, Leggett J, Köckenberger W (2010) Slice-selective single scan proton COSY with dynamic nuclear polarisation. *Phys Chem Chem Phys* 12:5771
64. Zhang G, Schilling F, Glaser SJ, Hilty C (2013) Chemical shift correlations from hyperpolarized NMR using a single SHOT. *Anal Chem* 85: 2875–2881
65. Koelsch BL, Keshari KR, Peeters TH, Larson PEZ, Wilson DM, Kurhanewicz J (2013) Diffusion MR of hyperpolarized ^{13}C molecules in solution. *Analyst* 138:1011
66. Schilling F, Düwel S, Köllisch U, Durst M, Schulte RF, Glaser SJ, Haase A, Otto AM, Menzel MI (2013) Diffusion of hyperpolarized ^{13}C -metabolites in tumor cell spheroids using real-time NMR spectroscopy. *NMR Biomed* 26: 557–568
67. Ahola S, Zhivonitko VV, Mankinen O, Zhang G, Kantola AM, Chen H-Y, Hilty C, Koptug IV, Telkki V-V (2015) Ultrafast multidimensional Laplace NMR for a rapid and sensitive chemical analysis. *Nat Commun* 6:8363
68. Ardenkjaer-Larsen JH, Macholl S, Johanneson H (2008) Dynamic nuclear polarization with trityls at 1.2 K. *Appl Magn Reson* 34: 509–522
69. Lumata L, Merritt ME, Malloy CR, Sherry AD, Kovacs Z (2012) Impact of Gd^{3+} on DNP of $[1-^{13}\text{C}]$ pyruvate doped with Trityl OX063, BDPA, or 4-Oxo-TEMPO. *J Phys Chem A* 116: 5129–5138

Determination of Protein ps-ns Motions by High-Resolution Relaxometry

Samuel F. Cousin, Pavel Kadeřávek, Nicolas Bolik-Coulon,
and Fabien Ferrage

Abstract

Many of the functions of biomacromolecules can be rationalized by the characterization of their conformational energy landscapes: the structures of the dominant states, transitions between states and motions within states. Nuclear magnetic resonance (NMR) spectroscopy is the technique of choice to study internal motions in proteins. The determination of motions on picosecond to nanosecond timescales requires the measurement of nuclear spin relaxation rates at multiple magnetic fields. High sensitivity and resolution are obtained only at high magnetic fields, so that, until recently, site-specific relaxation rates in biomolecules were only measured over a narrow range of high magnetic fields. This limitation was particularly striking for the quantification of motions on nanosecond timescales, close to the correlation time for overall rotational diffusion. High-resolution relaxometry is an emerging technique to investigate picosecond—nanosecond motions of proteins. This approach uses a high-field NMR spectrometer equipped with a sample shuttle device, which allows for the measurement of the relaxation rate constants at low magnetic fields, while preserving the sensitivity and resolution of a high-field NMR spectrometer. The combined analysis of high-resolution relaxometry and standard high-field relaxation data provides a more accurate description of the dynamics of proteins, in particular in the nanosecond range. The purpose of this chapter is to describe how to perform high-resolution relaxometry experiments and how to analyze the rates measured with this technique.

Key words Nuclear magnetic resonance, Protein dynamics, Nuclear spin relaxation, High-resolution relaxometry, Model free

1 Introduction

Nuclear magnetic resonance (NMR) can provide unique information about internal motions of biomolecules [1] with atomic resolution both in solution and in solid state using measurements of nitrogen-15 [2–5], carbon-13 [6–8], phosphorus-31 [9], or deuterium relaxation rates [10–15]. Both backbone and side-chain

Samuel F. Cousin, Pavel Kadeřávek and Nicolas Bolik-Coulon contributed equally to this work.

motions of biomolecules can be described over a broad range of timescales, from tens of picoseconds to seconds.

Nuclear spin relaxation is induced by fluctuations of orientation-dependent interactions (dipole-dipole coupling, quadrupolar coupling, etc.). These fluctuations can be described by time correlation functions $C(t)$. A correlation function is unique for each interaction contributing to relaxation. The Bloch-Wangsness-Redfield (BWR) [16–19] relaxation theory shows that the NMR relaxation rates depend on linear combinations of a set of values of the spectral density function $J(\omega)$, which is the Fourier transform of the correlation functions $C(t)$. For example, the relaxation rates of an isolated ^{15}N - ^1H spin system depend on the values of the spectral density function $J(\omega)$ at only five frequencies: $\omega = 0$; $-B_0\gamma_{\text{N}}$; $-B_0\gamma_{\text{H}}$; and $-B_0(\gamma_{\text{H}} \pm \gamma_{\text{N}})$, where B_0 is the static magnetic field intensity, while γ_{H} and γ_{N} are the gyromagnetic ratios of spins ^1H and ^{15}N , respectively. In order to quantify complex motions over two to three orders of magnitude of correlation times (typically from tens of picoseconds to a few nanoseconds), one needs to analyze sets of relaxation rates acquired at multiple magnetic fields, and hence on different NMR spectrometers.

Sensitivity and resolution are often limiting factors for protein NMR. The lowest magnetic field used in biomolecular studies is usually $B_0 = 11.75$ T, or more rarely 9.4 T. Thus, sampling the spectral density function at low (nonzero) frequencies is difficult. This makes nanosecond timescale motions, which are best sampled at low frequencies, difficult to characterize. High-resolution relaxometry was introduced [20–23] as a solution to overcome these limitations.

The principle of high-resolution relaxometry is to use the stray field of a high-field NMR magnet as a variable field (Fig. 1). Low magnetic fields are reached by displacing the sample to a chosen position in the stray field. This type of experiment is performed on a

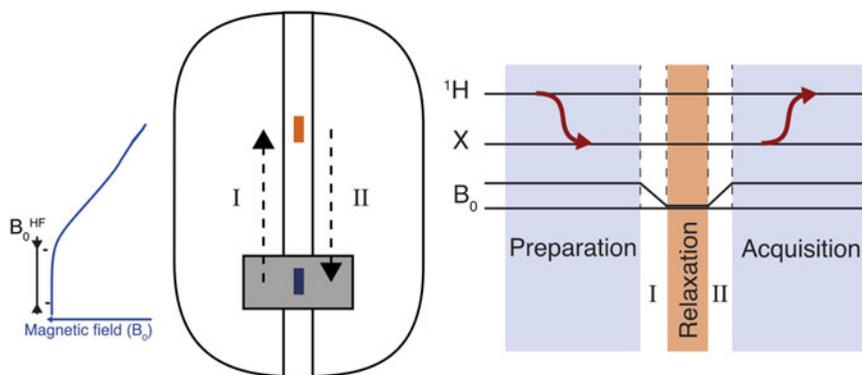


Fig. 1 Principle of high-resolution relaxometry. After the preparation of the desired spin term at high field, the sample travels in the stray field of the magnet, to reach a position with a lower magnetic field for the relaxation delay. Then, the sample is moved back to the high-field center for detection

commercial NMR spectrometer equipped with a shuttle device used to move the sample. Polarization delays, manipulations of spin system, and detection are performed at the high-field position. The sample is moved to a desired low field for the relaxation period. Such an experimental design ensures high sensitivity due to the polarization of nuclear spins at high field as well as the maximum resolution due to detection at high field. A single NMR spectrometer may be then used to measure relaxation rates on a broad range of magnetic fields, typically, from ~ 100 mT to the magnetic field inside the superconducting coil (>10 T).

Moving the sample requires a sample shuttle apparatus, which can be either mechanical [21, 22], or pneumatic [20, 23]. Once the sample is moved above the magnetic center, it is outside the RF coils. Thus, there is no possibility to control cross-relaxation pathways during the shuttle transfers and the relaxation delay. This leads to deviations from a mono-exponential decay with a desired auto-relaxation rate, which must be taken into account. The analysis of high-resolution relaxometry data requires good understanding of spin dynamics and relaxation pathways over the entire range of magnetic fields explored, from 100 mT to ~ 20 T.

Here, we present procedures on how to perform and analyze data from high-resolution relaxometry experiments. After a brief presentation of basic theory, we describe the experimental protocol for the measurement of protein backbone amide ^{15}N relaxation rates over several orders of magnitude of magnetic field using high-resolution relaxometry. We follow with a presentation of the analysis of these relaxation data to quantify protein backbone motions in the picosecond-nanosecond range.

2 Theory

A full description of relaxation theory is beyond the scope of this chapter, and it can be found elsewhere [24, 25]. We provide a brief introduction of the theoretical background necessary for understanding the analysis of relaxation rates obtained from high-resolution relaxometry, which includes a key correction step. As an example, the theory is described for an isolated ^{15}N - ^1H spin pair. It should be stressed that relaxation processes depend on the details of the spin system, e.g., the number of spins and their geometric arrangement. A different set of equations and different relaxation matrices should be used for other spin systems but the general framework will be similar.

The relaxation matrix for the complete set of spin operators of the given system is block-diagonal, due to the secular approximation [24, 25]. Only the spin terms that belong to the same subspace

as the spin term created at the beginning of the relaxation period should be taken into account. No RF pulses can be applied at low field, thus, only the relaxation of populations (e.g., longitudinal polarizations and spin orders that remain secular in the absence of RF fields) has to be considered here.

2.1 Isolated ^{15}N - ^1H Spin Pair

In the case of an isolated ^{15}N - ^1H spin pair, the relaxation of the expectation value for the ^{15}N longitudinal magnetization toward equilibrium can be described by a set of differential equations represented in a matrix form in the following way:

$$\frac{d}{dt} \begin{pmatrix} \langle N_z \rangle - \langle N_z \rangle^{\text{eq}} \\ \langle H_z \rangle - \langle H_z \rangle^{\text{eq}} \\ \langle 2N_z H_z \rangle \end{pmatrix} = - \begin{pmatrix} R_{1,\text{N}} & \sigma & \delta_{\text{N}} \\ \sigma & R_{1,\text{H}} & \delta_{\text{H}} \\ \delta_{\text{N}} & \delta_{\text{H}} & R_{\text{NH}} \end{pmatrix} \begin{pmatrix} \langle N_z \rangle - \langle N_z \rangle^{\text{eq}} \\ \langle H_z \rangle - \langle H_z \rangle^{\text{eq}} \\ \langle 2N_z H_z \rangle \end{pmatrix} \quad (1)$$

where $R_{1,\text{N}}$ and $R_{1,\text{H}}$ are the longitudinal auto-relaxation rates of the nitrogen-15 nucleus and the proton; R_{NH} is the auto-relaxation of the two-spin order, σ is the dipolar cross-relaxation rate between the nitrogen-15 and proton polarization, δ_{N} and δ_{H} are cross-correlated cross-relaxation rates caused by the cross correlations between the ^{15}N - ^1H dipole-dipole interaction and the chemical shift anisotropy of the nitrogen-15 nucleus and the proton, respectively. These relaxation rates can be written as [26, 27]:

$$R_{1,\text{N}} = \frac{d_{\text{NH}}^2}{4} \left(3J_{\text{NH}}(\omega_{\text{N}}) + 6J_{\text{NH}}(\omega_{\text{H}} + \omega_{\text{N}}) + J_{\text{NH}}(\omega_{\text{H}} - \omega_{\text{N}}) \right) + c_{\text{N}}^2 J_{\text{N}}(\omega_{\text{N}}) \quad (2)$$

$$R_{1,\text{H}} = \frac{d_{\text{NH}}^2}{4} \left(3J_{\text{NH}}(\omega_{\text{H}}) + 6J_{\text{NH}}(\omega_{\text{H}} + \omega_{\text{N}}) + J_{\text{NH}}(\omega_{\text{H}} - \omega_{\text{N}}) \right) + c_{\text{H}}^2 J_{\text{H}}(\omega_{\text{H}}) \quad (3)$$

$$\sigma = \frac{d_{\text{NH}}^2}{4} \left(6J_{\text{NH}}(\omega_{\text{H}} + \omega_{\text{N}}) - J_{\text{NH}}(\omega_{\text{H}} - \omega_{\text{N}}) \right) \quad (4)$$

$$\delta_{\text{N}} = \sqrt{3} c_{\text{N}} d_{\text{NH}} J_{\text{NH},\text{N}}(\omega_{\text{N}}) \quad (5)$$

$$\delta_{\text{H}} = \sqrt{3} c_{\text{H}} d_{\text{NH}} J_{\text{NH},\text{H}}(\omega_{\text{H}}) \quad (6)$$

$$R_{\text{NH}} = \frac{3}{4} d_{\text{NH}}^2 \left(J_{\text{NH}}(\omega_{\text{H}}) + J_{\text{NH}}(\omega_{\text{N}}) \right) + c_{\text{N}}^2 J_{\text{N}}(\omega_{\text{N}}) + c_{\text{H}}^2 J_{\text{H}}(\omega_{\text{H}}) \quad (7)$$

where d_{NH} describes the intensity of the dipolar interaction between the spins ^{15}N and ^1H , and c_i is the constant for the chemical shift anisotropy of nucleus i :

$$c_i = \frac{B_0 \gamma_i \Delta \sigma_i}{\sqrt{3}} \quad (8)$$

$$d_{\text{NH}} = \left(\frac{\mu_0}{4\pi} \right) \frac{\gamma_{\text{N}} \gamma_{\text{H}} \hbar}{r_{\text{NH}}^3} \quad (9)$$

μ_0 is the permeability of vacuum, \hbar is the reduced Planck constant, γ_{N} is the gyromagnetic ratio of ^{15}N , γ_{H} is the gyromagnetic ratio of the proton, and r_{NH} is the distance between the nitrogen-15 nucleus and the proton (1.02 Å). $\Delta \sigma_i$ is the chemical shift anisotropy of the nucleus i , and B_0 is the intensity of the static magnetic field at the sample position. $J_k(\omega)$ is the value of the spectral density function corresponding to the auto-correlation of interaction k at the frequency ω , and $J_{k,l}(\omega)$ is the value of the spectral density function corresponding to the cross-correlation of interactions k and l . The subscript NH stands for the dipolar coupling between the two nuclei, while H (or N) stand for the chemical shift anisotropy of the proton (or the nitrogen-15 nucleus).

Detailed presentations of the models that lead to the functional form of the spectral density function can be found elsewhere. Some models assume certain types of motion, which leads to a particular functional form of the spectral density function. For example, motions can be represented by the Gaussian Axial Fluctuation (GAF) [29] model, which was inspired by molecular dynamics simulations and described motions of the peptide plane with fluctuations of orientation angles around a set of principal axes following a Gaussian distribution. Another type of approach does not rely on a particular type of motion, but assumes a general parametric form of the spectral density function is selected. This Model Free approach [30] (MF) is generally the method of choice to describe the internal dynamics of folded proteins.

Most approaches to the analysis of relaxation rates in terms of picosecond-nanosecond motions assume statistical independence of overall tumbling and internal motions. The slowly relaxing local structure [31, 32] (SRLS) model has been suggested to take into account a coupling between the internal and the overall molecular motion. SRLS describes local motions as diffusion in a potential imposed by the molecular environment. The description of motions of the highly flexible fragments of proteins or intrinsically disordered proteins (IDP) is also challenging. A model assuming a distribution of correlation times of reorientational motions has been suggested [33–35] for that purpose. Other approaches reconstitute the spectral density function with a set of discrete correlation times [28, 36].

The analysis of high-resolution relaxometry rates presented here can be used for most of these modes of analyses of dynamics. However, we will present the principle of the analysis with the well-known model-free and extended model-free [37] approaches. The justification and criteria for the selection of the appropriate form of the spectral density function is similar to the standard NMR analysis of relaxation rates and should be based on statistical tests and physical arguments [38–42].

2.2 Protons in the Vicinity of the ^{15}N - ^1H Spin Pair

The amide ^{15}N - ^1H spin pairs should not be considered isolated to describe properly the relaxation of spin terms that involve the amide proton.

Protons in the vicinity of the amide proton influence significantly the relaxation of some spin terms and should be considered in the analysis. It has been shown that the effects of all additional protons can be sufficiently well accounted for if one considers two virtual protons placed in the vicinity of the amide proton (Fig. 2) [43, 44, 23]. The choice of distance between the additional protons and the amide proton (d_{HH}) depends on the level of the deuteration and compactness of the local structure. For folded proteins, this effective distance varies between 2.1 Å (protonated sample) and 2.9 Å (deuterated sample). The relaxation matrix (Eq. 1) should be expanded to include the effects of additional protons (the relevant sub-space dimension increases from 3 to 7):

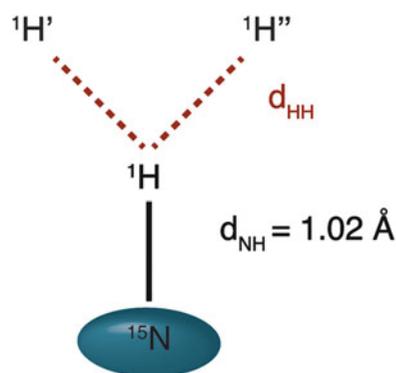


Fig. 2 NH spin system considering two protons in the vicinity

$$\frac{d}{dt} \begin{pmatrix} \langle N_z \rangle - \langle N_z \rangle^{eq} \\ \langle H_z \rangle - \langle H_z \rangle^{eq} \\ \langle 2N_z H_z \rangle \\ \langle H'_z \rangle - \langle H'_z \rangle^{eq} \\ \langle 2N_z H'_z \rangle \\ \langle H''_z \rangle - \langle H''_z \rangle^{eq} \\ \langle 2N_z H''_z \rangle \end{pmatrix} = - \begin{pmatrix} R_{1,N} & \sigma & \delta_N & 0 & 0 & 0 & 0 \\ \sigma & R'_{1,H} & \delta_H & \sigma' & 0 & \sigma'' & 0 \\ \delta_N & \delta_H & R'_{NH} & 0 & \sigma' & 0 & \sigma'' \\ 0 & \sigma' & 0 & R_{1,H'} & 0 & 0 & 0 \\ 0 & 0 & \sigma' & 0 & R_{NH'} & 0 & 0 \\ 0 & \sigma'' & 0 & 0 & 0 & R_{1,H''} & 0 \\ 0 & 0 & \sigma'' & 0 & 0 & 0 & R_{NH''} \end{pmatrix} \begin{pmatrix} \langle N_z \rangle - \langle N_z \rangle^{eq} \\ \langle H_z \rangle - \langle H_z \rangle^{eq} \\ \langle 2N_z H_z \rangle \\ \langle H'_z \rangle - \langle H'_z \rangle^{eq} \\ \langle 2N_z H'_z \rangle \\ \langle H''_z \rangle - \langle H''_z \rangle^{eq} \\ \langle 2N_z H''_z \rangle \end{pmatrix} \quad (10)$$

where the additional terms compared to Eq. 1 are

$$R'_{1,H} = R_{1,H} + \frac{d_{HH'}^2}{2} \left(3J_{HH'}(\omega_H) + 6J_{HH'}(2\omega_H) + J_{HH'}(0) \right) \quad (11)$$

$$R'_{NH} = R_{NH} + \frac{d_{HH'}^2}{2} \left(3J_{HH'}(\omega_H) + 6J_{HH'}(2\omega_H) + J_{HH'}(0) \right) \quad (12)$$

$$R_{1,H'} = R_{1,H''} = R_{1,H} + \frac{d_{HH'}^2}{2} \left(3J_{HH'}(\omega_H) + 6J_{HH'}(2\omega_H) + J_{HH'}(0) \right) \quad (13)$$

$$R_{NH'} = R_{NH''} = R_{NH} + \frac{d_{HH'}^2}{2} \left(3J_{HH'}(\omega_H) + 6J_{HH'}(2\omega_H) + J_{HH'}(0) \right) \quad (14)$$

$$\sigma' = \sigma'' = \frac{d_{HH'}^2}{4} \left(6J_{HH'}(2\omega_H) - J_{HH'}(0) \right) \quad (15)$$

where $d_{HH''} = d_{HH'}$ describes the intensity of the dipolar interaction between the amide proton and the additional protons in the vicinity (*see* Eq. 9). Note that the two additional protons were chosen to relax as amide protons. A slightly more thorough approach would be to consider two aliphatic protons.

Note, that Eq. 10 represents only a simplified relaxation matrix of the spin system, in which only the dipole-dipole interactions between the amide proton and the protons in its vicinity are considered, while all other interactions and all dipole-dipole cross-correlations are neglected.

3 Materials

3.1 Instrumentation

1. A high-field NMR spectrometer ($B_0 \geq 11.75$ T) equipped with a high-resolution relaxometry device and software for controlling the shuttling compatible with the NMR operating program (e.g., a recent version of Topspin for Bruker spectrometers). In our case, we use a pneumatic shuttle system described in detail elsewhere [23]. The shuttling system is composed of shuttling controller, a sample shuttle adapted to the system, and a shuttle guide to constrain the motion of the sample shuttle. The position chosen for the relaxation delay is set by a shuttle stopper, which is tightly fixed to the top of the Dewar.
2. A gaussmeter with a detection range up the magnetic field in the center of the NMR spectrometer to measure the whole magnetic field profile inside the bore of the NMR spectrometer.

3.2 Sample

1. A sample container adapted to the shuttle apparatus. Note that different sample shuttles require different tubes (here we use a pneumatic shuttle device developed by Bruker [23]).
2. A solution of protein specifically or uniformly labeled with nitrogen-15. It must be considered that the sensitivity of relaxometry experiments is generally lower than the sensitivity of typical high-field relaxation experiments because of relaxation during the transfer of the sample to low field and back to the high field. In our hands, we have favored sample concentrations of 0.8 mM and above. Attention should be paid to the homogeneity of the sample, which must be preserved during the whole relaxometry experiment. We recommend degassing the sample in order to prevent the formation of a bubble inside the sample during shuttling. Degassing is performed by placing the sample solution in a mild vacuum. Increase the applied vacuum slowly to avoid the sample splashing and its potential loss. Apply the vacuum as long as bubbles are released from the solution. Immediately after the end of degassing seal the sample.
3. A temperature calibration sample. Typically, a methanol sample is used for this purpose (either a pure unlabeled methanol sample [45] or fully deuterated methanol sample [46–48]).
4. It is advisable to use an internal temperature sensor. For example, deuterated acetic acid may be used as a buffering agent and the chemical shift difference between the signals of deuterium from the methyl group of acetic acid and HOD in the buffer can be monitored.

3.3 Software

The analysis of relaxation data described here relies on the use of well-known NMR software, software for curve fitting, and a house-

developed python-based program dedicated to the analysis of relaxometry data.

1. nmrPipe [49]: software designed for processing and analysis of NMR spectra <https://www.ibbr.umd.edu/nmrpipe/>.
2. CurveFit: software to fit the relaxation decays and estimate the errors of the fitted parameters <http://www.cumc.columbia.edu/dept/gsas/biochem/labs/palmer/software/curvefit.html>.
3. Grace: software to display the results of the Curvefit analysis (<http://plasma-gate.weizmann.ac.il/Grace/>).
4. ROTDIF: software used to determine the overall diffusion tensor [40, 50] (<http://gandalf.umd.edu/FushmanLab/pdsw.html>).
5. ICARUS: software developed [23] to analyze relaxation rates measured by relaxometry http://paris-en-resonance.fr/code/code_samples.php (the program requires python 3.5 for which documentation is available at <http://www.python.org>, including the following libraries: Tkinter, matplotlib, scipy and numpy). This version of ICARUS is written so that users can adapt the relaxation matrix for the specific spin system under study. Here, we only describe the procedure for an isolated ^{15}N - ^1H pair. This version of ICARUS includes a graphical user interface, which allows the user to import data sets and choose model of the spectral density function easily.

3.4 Conventional Experimental Data

1. A set of high-field relaxation rates (longitudinal relaxation rate R_1 , transverse relaxation rate R_2 , dipolar cross-relaxation rate σ) measured at least at one high magnetic field [51].
2. Analysis of these relaxation rates using ROTDIF [40] to extract the diffusion tensor for the global tumbling of the entire molecule.

4 Methods

4.1 Mapping of the Magnetic-Field Profile

Measure the dependence of the magnetic field on the distance from the high-field magnetic center along the axis of the superconducting magnet. These measurements should be performed just once since the magnetic field is very stable. The measurements have to be accurate for the entire trajectory where the sample travels, in order to be able to simulate relaxation during the entire experiment (this is necessary for the analysis). It is recommended to measure the intensity of the magnetic field with steps of 5 mm or less. The profile of the stray field may also be provided by the manufacturer of your spectrometer upon request.

4.2 Temperature Calibration

The temperature of the sample should be carefully calibrated and remain stable during the experiment. In particular, relaxation rates depend on the viscosity of water, which is sensitive to temperature. Commercially available high-resolution relaxometry systems do not allow yet the control of the temperature within the bore of the Dewar. Currently, experiments carried out close to room temperature offer the best control of the temperature of the sample. The temperature is calibrated using a sample that displays two NMR signals with a chemical shift difference that varies reliably with the temperature. Ideally, the compound used to determine the temperature is added as an internal standard to the protein sample, so that we can monitor the temperature during the experiment.

The compound used as an internal temperature standard needs to fulfill several conditions:

- The temperature reference must be soluble in the buffer.
- It must not interact with the protein sample.
- Molecules must be stable.
- It must provide NMR signals with a clear temperature dependence of the monitored chemical shift difference.
- The signals of the temperature reference compound should not overlap with the protein signals.

In our case, we used a buffer composed of 50 mM of deuterated acetic acid (AcOH), and the difference between the deuterium signals of HOD and of the methyl group of AcOH was monitored. It is recommended to verify the dependence of the chemical shift difference with respect to the temperature for a new sample:

1. Introduce the external reference (e.g., perdeuterated methanol sample) in the spectrometer.
2. Temperature stabilization should take about 10 min, if the initial temperature of the sample is <5 °C from the desired value. Match and tune the proton circuit and adjust shims to reach sufficient homogeneity of the magnetic field.
3. Run a single pulse 1D proton experiment using a $1 \mu\text{s}$ ^1H pulse at full power.
4. Process the spectrum, read the chemical shifts difference $\Delta\delta$ between the hydroxyl and methyl protons.
5. Calculate the temperature, if using a methanol- d_4 sample, use the formula [48]:

$$T = -16.7467\Delta\delta^2 - 52.5130\Delta\delta + 419.1381.$$
6. Change the nominal temperature inside the spectrometer and repeat **steps 2–5** until the desired temperature is reached.
7. Repeat **steps 2–6** with temperatures deviating by $+4$ °C, $+2$ °C, -2 °C, and -4 °C from the target temperature.

8. Insert the protein sample with the internal temperature standard (or a protein buffer sample without the protein) inside the NMR spectrometer.
9. Adjust shims to reach sufficient homogeneity of the magnetic field, match and tune the probe.
10. Perform the experiment (1D proton or deuterium spectrum) for temperature calibration with the internal standard.
11. Process the spectrum and read the difference between the selected chemical shifts.
12. Determine the chemical shifts difference at all temperatures calibrated by the standard temperature calibration sample in **steps 1–7**.
13. Determine the relationship between the measured chemical shifts difference and the temperature. Usually, a linear dependence is a good approximation in such a small range of temperatures.

This relationship between the temperature and the chemical shift difference should be used for the temperature calibration before every relaxation measurement on this sample:

1. Insert the protein sample with the internal temperature standard (or a sample with buffer but no protein) inside the NMR spectrometer.
2. Adjust shims to reach sufficient homogeneity of the magnetic field, match and tune the probe.
3. Set up the desired temperature and follow all the steps for the preparation of the relaxometry experiment (Subheading 4.3), use exclusively the shortest possible relaxation delay. Set up the number of dummy scans to 100.
4. Start the relaxometry experiment (Subheading 4.3) and monitor the temperature during the experiment, because the shuttling process may influence the sample temperature. Note, how many scans are necessary to obtain a stable temperature. Use this value as the number of dummy scans.
5. If monitoring the temperature during the experiment is not possible, you can use an alternative approach: start the experiment, stop it when 100 dummy scans are finished and immediately after, run the 1D experiment for temperature calibration. Note the temperature. Double the number of dummy scans to confirm the stability of the temperature. If the temperature is not stable, increase the number of dummy scans to 300. Note the number of dummy scans. Adjust the temperature so that the desired temperature is reached after the dummy scans.
6. Repeat **steps 4** and **5** with the longest expected relaxation delay. The calibrated temperature should not differ more than

0.3 K; otherwise, the system temperature stabilization setup must be adjusted (e.g., the steady gas flow in the high field probe, or pressures settings in a pneumatic system) or the longest relaxation delay should be reduced. The number of dummy scans in the relaxometry experiment (Subheading 4.3, item 14) should be large enough so that the temperature is stable when the first signal is recorded with both relaxation delays.

4.3 Setting Up Low-Field Relaxation Experiments

Nuclear magnetic relaxation dispersion profiles display the variation of relaxation rates with the magnetic field. The analysis of protein dynamics requires that the details of the dependence of the relaxation rate with the field are obtained. It is recommended to record the relaxometry experiment with at least 8 (and up to ~20, depending on the spectrometer time available) different sample positions (i.e., magnetic fields) spanning the range of low fields 0.1–10 T. We advise users to analyze each relaxation experiment as soon as it is recorded in order to select the most appropriate magnetic field for the next measurement. This strategy leads to an optimal sampling of the relaxation dispersion profile.

The sample position at low field, where the relaxation takes place, is adjusted before each experiment and should be verified regularly (if possible) until the end of each experiment.

1. Choose the magnetic field at which relaxation will be measured. Set up the low-field stop position for the sample. The magnetic field should be chosen from the table of magnetic field vs. distance from the center of the magnet (*see* Subheading 4.1).
2. Generally, the shuttle transfers should be as short as possible in order to avoid unnecessary loss of sensitivity. We recommend transfer delays shorter than 150 ms, and if possible shorter than 100 ms. If the pneumatic shuttle device is used, the shuttling speed is adjusted by gas pressures settings at the high- and low-field positions. Pressure settings allow to either blow (positive) or suck (negative) the gas with a pressure up to 5 bar and down to –0.5 bar. It is recommended to set the positive pressure always slightly larger than the negative one, in order to prevent ambient dust to enter the system. The example gas pressure settings used in the Bruker pneumatic shuttling device for a sample of a diameter 3.2 mm and length 2 cm are summarized in Table 1.
3. Calibrate the temperature as described in Subheading 4.2. We recommend the calibration of the temperature for every experiment with a different low-field position.
4. Insert the sample inside the spectrometer. Let the temperature equilibrate for 10 min. Match and tune the probe and shim the magnet.

Table 1
Settings of the pressures for the high-resolution relaxometry on the Bruker pneumatic device

	Pressure at high-field position (mbar)	Pressure at low-field position (mbar)
Shuttling from high field to low field	350	-300
Shuttling from low field to high field	-250	300
Sample at low field	200	-100
Sample at high field	-40	100

- Record a 1D proton spectrum with a short (1 μ s) excitation pulse and with the carrier frequency on resonance with the signal of water (~ 4.7 ppm), use the minimum receiver gain (RG = 1 for Bruker NMR spectrometers). Process the 1D spectrum. Then, perform the same experiment with the expected pulse length for a 360° flip angle. Process the spectrum with the same phase as used for the first spectrum. Vary the length until a null signal is obtained. Finally, use this value to calculate the length of the 90° ^1H pulse (one quarter of the duration of the 360° pulse).
- Prepare a 1D version of a ^{15}N - ^1H Heteronuclear Single Quantum Coherence (HSQC) experiment with a heteronuclear gradient echo. Set the calibrated length of the proton pulse and the expected value of a ^{15}N 90° pulse, place the proton carrier frequency on resonance with the signal of water (~ 4.7 ppm) and the ^{15}N carrier frequency in the middle of the amide region (~ 118 ppm), set the number of dummy scans to 16, and the number of scans from 2 to 16, depending on the sensitivity. Adjust the receiver gain value. Run the experiment, process the spectrum, and adjust its phase. If the signal-to-noise is too small, increase the number of scans accordingly, but not the number of dummy scans.
- Double the duration of the first ^{15}N 90° pulse in the sequence. Run the experiment and process it, using the phase optimized in the previous step. Optimize the length of this pulse to get a null signal. Halve the length of the pulse to obtain a ^{15}N 90° pulse.
- If the spectral width and carrier frequency of a 2D ^{15}N - ^1H HSQC of the sample has not been optimized yet set up a 2D HSQC experiment with a ~ 40 ppm spectral width and carrier frequency 118 ppm in the indirect dimension. The broad spectral width generally ensures that no peaks except arginine NeHe signals will be aliased. Run the experiment, process it, and

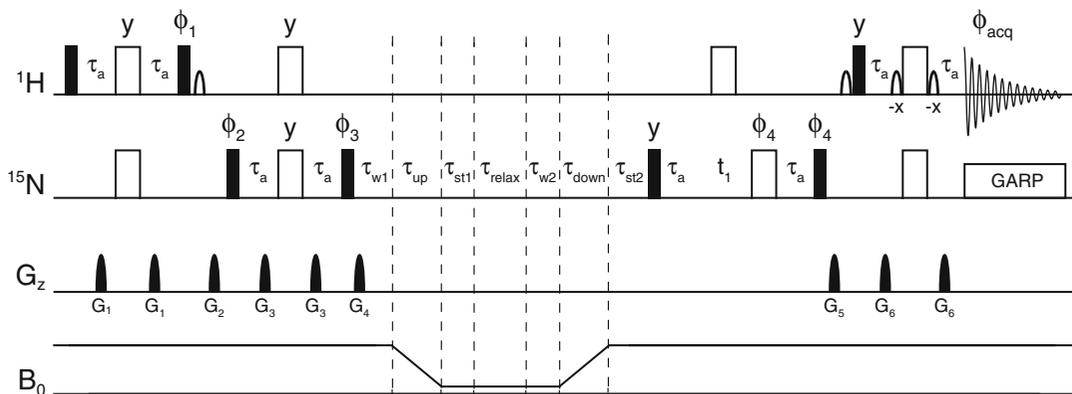


Fig. 3 Pulse sequence used for the measurement of longitudinal nitrogen-15 relaxation of amide nitrogen nuclei in proteins at various low fields and with recovery delay and detection at high field. *Narrow (filled) and wide (open) rectangles* represent 90° and 180° pulses, respectively. Pulses are applied along the x -axis of the rotating frame unless otherwise specified. The *bell-shaped* pulses represent 1.2 ms sinc pulses. All delays τ_a are set to be equal to $1/4J_{\text{NH}}$, with $J_{\text{NH}} = -92$ Hz. The sample is moved from the high- to the low-field position during the delay τ_{up} and back during the delay τ_{down} . The delays τ_{w1} and τ_{w2} correspond to the response delays of the shuttling device. The short delay $\tau_{\text{st1}} = \sim 40$ ms is required by the system, the stabilization delay $\tau_{\text{st2}} = 200$ ms allows for convection currents and vibrations to settle. Pulsed field gradients G_z have smoothed rectangular amplitude profiles and 1 ms durations. Their peak amplitudes are $G_1 = 12.5$ G/cm; $G_2 = 20$ G/cm; $G_3 = 5.75$ G/cm; $G_4 = 10.25$ G/cm; $G_5 = 20$ G/cm and $G_6 = 7.5$ G/cm. The phase cycles were $\phi_1 = \{\{y\}_4, \{-y\}_4\}$, $\phi_2 = \{x, -x\}$, $\phi_3 = \{\{y\}_8, \{-y\}_8\}$, $\phi_4 = \{\{x\}_2, \{-x\}_2\}$, $\phi_{\text{acq}} = \{x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x\}$

optimize the ^{15}N carrier frequency and the spectral width so that the spectral width will be as small as possible, but no folded peak will interfere with any other signals and no peak will be closer than 0.5 ppm to the edge of the spectrum.

9. Set up a relaxation experiment with the pulse sequence shown in Fig. 3 and set the delays and gradients according to the description in Fig. 3.
10. Set durations of ^{15}N and ^1H pulses according to the calibrations.
11. Set the optimized ^{15}N carrier frequency and spectral width.
12. Set the proton shape pulses according to the spectrometer-specific protocol.
13. Choose an inter-scan delay of at least 2 s, preferably 3 s.
14. Set the number of dummy scans so that the temperature is stable before the first scan is recorded (*see* Subheading 4.2). In any case the number of dummy scans should not be smaller than 100.

15. Set the number of scans to 8 and run the experiment to acquire the first increment with the shortest possible relaxation delay τ_{\min} (first line in the VDLIST for Bruker systems) and then stop it. If the signal-to-noise ratio is not sufficient increase the number of scans accordingly.
16. Set the estimated length of the relaxation delay τ_{\max} , which should lead to a decay of the magnetization to about 1/3 of the intensity obtained for the shortest relaxation delay. Run the experiment and stop it after the first increment is recorded. Process the first increment and compare its intensity with the intensity of the first increment measured with the shortest relaxation delay. Optimize τ_{\max} until the required intensity decay is reached.
17. If the maximum relaxation delay exceeds the one tested during the temperature calibration, the stability of the temperature should be verified.
18. Estimate an average relaxation rate $R = \ln(3)/(\tau_{\max} - \tau_{\min})$.
19. Prepare the list of relaxation delays. The number of various relaxation delays N should be between 4 and 8. It is also recommended to repeat two relaxation delays in the series. The series of relaxation delays can be calculated using $\tau_n = \tau_{\min} + (1/R)\ln((3N - 3)/(3N - 2n - 1))$, where τ_n is the n th relaxation delay. It is recommended to mix the order of the relaxation delays in the list, for example according to a pattern: $\tau_1, \tau_N, \tau_2, \tau_{N-1}, \dots$. Note that if there are protein residues with peculiar dynamics, these relaxation delays might not be optimal for them. A second experiment should be performed with the relaxation delays adjusted specifically for them. Once this feature is known, the list of relaxation delays can be modified to include additional short delays for fast-relaxing residues and long delays for slow-relaxing residues.
20. Start the experiment. When the first increments corresponding to all defined relaxation delays are acquired, process them and verify that the intensity decays between spectra recorded with consecutive delays are more or less constant, as expected.

4.4 Data Processing

1. The spectra corresponding to the individual relaxation delays are acquired in an interleaved manner. The first step of the analysis is to split the resulting file containing all FIDs in order to separate them by relaxation delay (use NMRpipe [49] or a python routine in Topspin extending the inbuilt *split* procedure (available here: http://paris-en-resonance.fr/code/code_samples.php)).
2. Processing of the spectra is performed using the NMRPipe program [49]. We recommend a sine-bell window function, with a shift parameter between 0.5 and 0.35 in both

dimensions. Truncation artifacts should be avoided because they might contaminate intensities of other peaks.

3. Pick the peaks in the most intense spectrum. Once the peaks are assigned, export the peak list and save it as peakX_0.tab, where X is the index number of the spectrum.
4. Extract the peak intensities using the nonlinear line shape analysis tool (nlinLS), provided with nmrPipe: Measure full width in points for both proton (WH) and ^{15}N dimension (WN) and run the following line in the command line:


```
nlinLS -in peakX_0.tab -out peakX_1.tab -data spectrumX.ft2 -w WH/6 WN/6 -delta X_AXIS WH/10 Y_AXIS WN/10.
```
5. It is likely that the nlinLS program will initially provide a series of error messages. Most often they indicate a poor convergence of the fit. In spite of these error messages, use the output file as an input for a second round of analysis with nlinLS. The process usually requires one to two iterations (up to five might be necessary in difficult cases) for convergence providing a proper fit with no error message. If the width of a peak is diverging from typical values, add additional constraints of the peak width parameters, WX and WY. In order to improve the convergence, peaks may also be excluded or added from clusters. The final list obtained with no error messages should be used as the first input for the analysis of all other spectra.
6. Perform the iterative procedure described in **steps 4** and **5** for every spectrum.
7. Extract peak intensities from the peakX_n.tab file produced by the last (nth) run of the program nlinLS. The value of the noise of each spectrum is usually a good estimate of the uncertainty in the intensities.
8. Prepare the inputs for the relaxation decay fitting using the script inputcurvefit from the package Curvefit.
9. Fit the decays of the peak intensities of all residues with the relaxation delay using the script batch_curve.in from the program Curvefit. A mono-exponential function can be used as a fitting function, even though the decay is not perfectly mono-exponential.
10. Plot the experimental dependence of the peak intensities on the relaxation delay in the program Grace and verify the quality of the fit for each residue. An example is shown in Fig. 4.

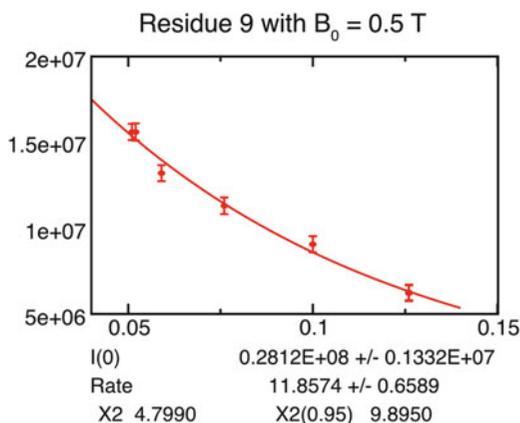


Fig. 4 Example of relaxation decay as shown by Grace during the Curvefit procedure. The data correspond to the measurement of longitudinal relaxation of residue 9 in Ubiquitin, at 0.5 T

5 Data Analysis Using ICARUS

5.1 Principles Used in the Analysis

ICARUS stands for Iterative Correction for the Analysis of Relaxation Under Shuttling. When the sample is outside the high-field probe, it is impossible to control cross-relaxation pathways. The measured intensity decays do not correspond to pure longitudinal relaxation, but they have a complex multi-exponential form. The program ICARUS is designed to take into account the effects of (auto- and cross-correlated) cross-relaxation pathways during the relaxometry experiment and correct the measured relaxation rates accordingly to obtain the ^{15}N auto-relaxation rate $R_{1,\text{N}}$.

The evolution of the spin system is simulated for the part of the experiment that spans the delays $\tau_{w1} + \tau_{\text{up}} + \tau_{\text{st1}} + \tau_{\text{relax}} + \tau_{w2} + \tau_{\text{down}} + \tau_{\text{st2}}$ (see Fig. 3). The integration of the Master equation during the transfer between the magnetic center and the position for relaxation is divided into small time steps Δt . The position of the sample, hence the field, is calculated for each of the time steps. Δt should be short enough to consider that the magnetic field is constant during Δt . The full propagator is then obtained as the product of the propagators for all delays Δt .

ICARUS computes the relaxation matrices at all positions of the sample using selected models of the spectral density functions $J(\omega)$ and estimated parameters of internal dynamics for all residues. In the first iteration ICARUS uses the parameters for internal dynamics optimized from the accurate high-field relaxation rates (nitrogen-15 R_1 , R_2 and steady-state NOE) exclusively (the high-field relaxation rates are measured within the probe allowing the suppression of undesired cross-relaxation pathways). The overall spin propagators for all relaxation delays are used to calculate the

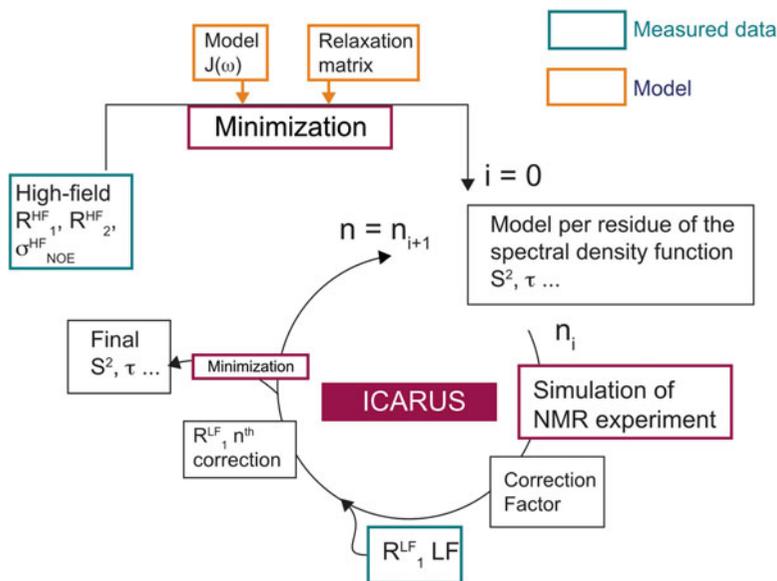


Fig. 5 Schematic representation of the iterative protocol for ICARUS, designed for the correction of low-field relaxation rates measured using the relaxometry experiment in Fig. 3. The outputs of the analysis are the parameters of the model of internal dynamics, here the order parameter S^2 and the effective correlation time τ

apparent relaxation decay in relaxometry experiments. This decay is fitted using a mono-exponential function, as in the protocol used for the analysis of the experimental decay. ICARUS calculates the ratio between the rate obtained in the fit of the simulated decay and the simulated $R_{1,N}$ rate at the low-field sample position derived from Eq. 2. This ratio is called “correction factor.” The experimental rate is then multiplied by the correction factor to obtain the scaled (or corrected) experimental low-field relaxation rates. The ensemble of corrected relaxation rates is then used, alongside the accurate high-field rates, as the input for the fit of the parameters for internal dynamics. The next iteration of ICARUS uses this new set of parameters to simulate the evolution of the spin system during the shuttling experiments. This iterative correction protocol has to be performed a few times, until the convergence of internal dynamics parameters is reached (twice is often enough). Figure 5 shows a schematic representation of the ICARUS protocol.

5.2 Preparation of Input Files

The program ICARUS requires several input files:

1. FunctionsFile.py: characteristics of the studied spin system and form of the spectral density function.
2. Magnetic field profile.
3. Experimental relaxation rates measured at high field (accurate values, no correction needed).

4. Experimental relaxation rates measured with relaxometry, which need to be corrected.
5. Parameters for the sample shuttle trajectories.
6. Value(s) of CSA(s).

The format of the input files is described in the following sections.

5.2.1 The Function File

This file contains all the physical properties of the spin system and molecule necessary for ICARUS. An example file is shown below, adapted for the analysis of backbone ^{15}N - ^1H spin pairs in the case of a model-free approach. Up to two additional protons can be added to the ^{15}N - ^1H spin system, using the “vicinity” variable equal to 0, 1 or 2. The file includes the definition of the geometry as well as the amplitudes and orientations of dipole-dipole and CSA tensors. Finally, this file contains the definition of the form of the spectral density functions used for the system. The model-free [30] form can be used. In the example below, the extended model-free form [37] of $J(\omega)$ is selected.

The analytical expressions of the relaxation rates acquired at high field (i.e., within a probe) are also included. In the example, nitrogen-15 R_1 , R_2 , and σ (from Eq. 4), measured in a steady-state nuclear Overhauser effect (NOE) experiment, are given for a ^{15}N - ^1H pair.

If the user chooses to use predefined gradients during the minimization process (see below), the function file needs to include the expressions for the first derivatives of the spectral density functions with respect to the optimized parameters (order parameters and correlation times in the given example).

Different spin systems and models of spectral density function can be used for ICARUS. The characteristics of the spin system and models are defined in the function file. The function file must follow the Python syntax and it must contain the following elements:

- A list called `RelaxationRates`, which contains strings corresponding to the defined relaxation rates in the file (see the example below, #flag 1).
- The position of the longitudinal relaxation rate R_1 in the relaxation matrix, called `PositionAuto`. It is 0 if it is the first element of the matrix, 1 if it is the second diagonal element, etc. (see the example below, #flag 2).
- The variable `pA` is the population of site A, if two sites in exchange with different overall and/or internal dynamics parameters are considered. `pA` should be set between 0.0 and 1.0 (`pA = 1.0` if there is only one population) (see the example below, #flag 3).

- The variable Anisotropy1 describes the rotational diffusion tensor. If it is symmetric, Anisotropy1 should be set to “NO” if it is axially symmetric to “Axial” and if it is completely asymmetric to “Full” (see the example below, #flag 4). If another state is considered, the variable Anisotropy2 should also be defined (see the example below, #flag 5).
- A dictionary called Names, which includes the names of variables that will be fitted in the analysis. A name between quotation marks defines a class of variables (e.g., order parameters or correlation times) and is followed by the series of corresponding names in between brackets. These names are used throughout the code of the function file (see the example below, #flag 6).
- The variable NeedRex should be set to “YES” if the effect of chemical exchange of amide protons with the solvent is considered, and to “NO” otherwise (see the example below, #flag 7).
- The definitions of relaxation rates measured at high fields (R_1 , R_2 , σ). The variables of these functions are the magnetic field, a vector X, which will contain the values of the optimized parameters specified in Names, and additional parameters, including the overall diffusion tensor. These functions are named using the RelaxationRates elements and by adding the suffix “calculation” to the name of the variable (for example R1calculation for the R1) (see the example below, #flag 8).
- The function RelaxMatrix describing the relaxation matrix used for the calculation of the evolution of the spin system during the relaxometry experiments. The same variables as for the relaxation rates are required. The spin term that is generated at high field before the transfer to low field must be set at the PositionAuto position (see the example below, #flag 9).
- All functions or variables required by the relaxation rates and matrix functions (spectral density function definition, interaction constants, etc.).

An example of FunctionsFile.py. This file describes the case of a ^{15}N - ^1H group with two protons in the vicinity. Here, only one population is considered ($p_A = 1.0$). The overall diffusion tensor is isotropic. The model-free formalism is used. The optional definition of the gradients for the basin-hopping algorithm is also included. Examples may be downloaded from the following URL: http://www.paris-en-resonance.fr/code/code_samples.php

```
# -*- coding: utf-8 -*-
import math
#Constants
hbar = 1.05457173e-34
mu = 4.0e-7 * math.pi
#Gyromagnetic ratio
GammaH = 42.576e6 * 2.0 * math.pi
GammaN = -4.316e6 * 2.0 * math.pi
```

```

# Defined relaxation rates
RelaxationRates = ["NOE", "R1", "R2"] #flag 1
# Variables required by ICARUS
PositionAuto = 0 #flag 2
pA = 1.0 #flag 3
Anisotropy1 = 'NO' #flag 4
Anisotropy2 = 'NO' #flag 5
Names = {'OrderParam': ['Sf2'], 'CorrTimes': ['tauloc']} #flag 6
NeedRex = 'YES' #flag 7
# Geometry constraints
vicinity = 2 #0, 1, or 2
#Distances
rhn = 1.02e-10
rhh = 2.1e-10
#Dipolar coefficients
delta_NH = (mu / (4.0 * math.pi)) * hbar * GammaH * GammaN / rhn**3
delta_HH = (mu / (4.0 * math.pi)) * hbar * GammaH * GammaH / rhh**3
#Angles
theta_HxxHN = math.pi / 2.0 #angle between the x axis of CSA and
NH bond
theta_HyyHN = 99.0 * math.pi / 180.0 #angle between the y axis of
CSA and NH bond
theta_NNH = 18.0 * math.pi / 180.0
# Spectral Density Function
#Model free on NH
def J(omega, X, tauc):
    Sf2 = X[0]
    tauloc = X[1]
    tauc = tauc[0]
    tau1 = tauc * tauloc / (tauc + tauloc)
    r0 = tauc / (1.0 + (omega * tauc)**2)
    r1 = tau1 / (1.0 + (omega * tau1)**2)
    spec = 2.0 / 5.0 * Sf2 * r0 + 2.0 / 5.0 * (1.0-Sf2) * r1
    return spec
def J_deriv(omega, X, tauc):
    Sf2 = X[0]
    tauloc = X[1]
    tauc = tauc[0]
    tau1 = tauc * tauloc / (tauc + tauloc)
    r0 = tauc / (1.0 + (omega * tauc)**2)
    r1 = tau1 / (1.0 + (omega * tau1)**2)
    dJMF_Sf2 = 2.0 / 5.0 * (r0 - r1)
    dJMF_tauloc = 2.0 / 5.0 * (1.0-Sf2) * ((tau1/tauloc)**2) * (1.0 -
(omega * tau1)**2) / (1.0 + (omega * tau1)**2)**2
    return [dJMF_Sf2, dJMF_tauloc]
# Relaxation Rates #flag 8 (and lines below)
def R1calculation(B0, X, tauc, DELTA, Rex, Angles):

```

```

DELTA_N = DELTA[0]/math.sqrt(3)
R1calc = (delta_NH**2) / 4.0 * (J((GammaN-GammaH)*B0, X, tauc)
+ 3.0 * J(GammaN * B0, X, tauc) + 6.0 * J((GammaN+GammaH)*B0, X,
tauc)) +\
(DELTA_N**2) * ((GammaN * B0)**2) * J(GammaN*B0, X, tauc)
return R1calc
def R2calculation(B0, X, tauc, DELTA, Rex, Angles):
DELTA_N = DELTA[0]/math.sqrt(3)
R2calc = (delta_NH**2) / 8.0 * (4.0 * J(0.0, X, tauc) + J((GammaN-
GammaH)*B0, X, tauc) + 3.0 * J(GammaN*B0, X, tauc) + 6.0 * J
(GammaH*B0, X, tauc) + 6.0 * J((GammaN+GammaH)*B0, X, tauc)) +\
(DELTA_N**2) * ((GammaN * B0)**2) / 6.0 * (4.0 * J(0.0, X, tauc) + 3.0 *
J(GammaN*B0, X, tauc))
return R2calc
def NOEcalculation(B0, X, tauc, DELTA, Rex, Angles):
DELTA_N = DELTA[0]/math.sqrt(3)
SigmaCalc = (delta_NH**2) / 4.0 * (6.0 * J((GammaN+GammaH)*B0,
X, tauc) - J((GammaN-GammaH)*B0, X, tauc))
return SigmaCalc
def R1calculation_deriv(B0, X, tauc, DELTA, Rex, Angles):
DELTA_N = DELTA[0]/math.sqrt(3)
f1 = J_deriv((GammaN-GammaH)*B0, X, tauc)
f2 = J_deriv(GammaN * B0, X, tauc)
f3 = J_deriv((GammaN+GammaH)*B0, X, tauc)
f4 = J_deriv(GammaN*B0, X, tauc)
dR1_Sf2 = (delta_NH**2) / 4.0 * (f1[0] + 3.0 * f2[0] + 6.0 * f3
[0]) + (DELTA_N**2) * ((GammaN * B0)**2) * f4[0]
dR1_tauloc = (delta_NH**2) / 4.0 * (f1[1] + 3.0 * f2[1] + 6.0 *
f3[1]) + (DELTA_N**2) * ((GammaN * B0)**2) * f4[1]
return [dR1_Sf2, dR1_tauloc]
def R2calculation_deriv(B0, X, tauc, DELTA, Rex, Angles):
DELTA_N = DELTA[0]/math.sqrt(3)
f1 = J_deriv(0.0, X, tauc)
f2 = J_deriv((GammaN-GammaH)*B0, X, tauc)
f3 = J_deriv(GammaN*B0, X, tauc)
f4 = J_deriv(GammaH*B0, X, tauc)
f5 = J_deriv((GammaN+GammaH)*B0, X, tauc)
f6 = J_deriv(0.0, X, tauc)
f7 = J_deriv(GammaN*B0, X, tauc)
dR2_Sf2 = (delta_NH**2) / 8.0 * (4.0 * f1[0] + f2[0] + 3.0 * f3
[0] + 6.0 * f4[0] + 6.0 * f5[0]) + (DELTA_N**2) * ((GammaN * B0)
**2) / 6.0 * (4.0 * f6[0] + 3.0 * f7[0])
dR2_tauloc = (delta_NH**2) / 8.0 * (4.0 * f1[1] + f2[1] + 3.0 *
f3[1] + 6.0 * f4[1] + 6.0 * f5[1]) + (DELTA_N**2) * ((GammaN *
B0)**2) / 6.0 * (4.0 * f6[1] + 3.0 * f7[1])
return [dR2_Sf2, dR2_tauloc]
def NOEcalculation_deriv(B0, X, tauc, DELTA, Rex, Angles):
DELTA_N = DELTA[0]/math.sqrt(3)

```

```

f1 = J_deriv((GammaN+GammaH)*B0, X, tauc)
f2 = J_deriv((GammaN-GammaH)*B0, X, tauc)
dSigma_Sf2 = (delta_NH**2) / 4.0 * (6.0 * f1[0] - f2[0])
dSigma_tauloc = (delta_NH**2) / 4.0 * (6.0 * f1[1] - f2[1])
return [dSigma_Sf2, dSigma_tauloc]
# Relaxation Matrix #flag 9 (and lines below)
def RelaxMatrix(B0, X, tauc, DELTA, Rex, Angles):
    DELTA_N = DELTA[0]/math.sqrt(3)
    sigma_Hxx, sigma_Hyy, sigma_Hzz = DELTA[1]/math.sqrt(3), DELTA
    [2]/math.sqrt(3), DELTA[3]/math.sqrt(3)
    rhoH = Rex[0] + 1.0/4.0 * (delta_NH**2) * (J((GammaH-GammaN)*B0,
    X, tauc) + 3.0 * J(GammaH*B0, X, tauc) + 6.0 * J((GammaH+GammaN)
    *B0, X, tauc)) +\
    float(vicinity) * 1.0/4.0 * (delta_HH**2) * (J(0.0, X, tauc) + 3.0 *
    J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) +\
    ((GammaH*B0)**2) * (sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2 -
    sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sig-
    ma_Hzz) * J(GammaH*B0, X, tauc)
    sigmaNH = 1.0/4.0 * (delta_NH**2) * (6.0 * J((GammaH+GammaN)*B0, X,
    tauc) - J((GammaH-GammaN)*B0, X, tauc))
    deltaH = math.sqrt(3) * delta_NH * GammaH*B0 * ((sigma_Hxx-
    sigma_Hzz) * ((3.0 * ((math.cos(theta_HxxHN))**2) - 1.0) / 2.0)
    + (sigma_Hyy - sigma_Hzz) * ((3.0 * ((math.cos(theta_HyyHN))
    **2) - 1.0) / 2.0)) * J(GammaH*B0, X, tauc)
    sigma = float(bool(vicinity)) * 1.0/4.0 * (delta_HH**2) * (6.0
    * J(2.0*GammaH*B0, X, tauc) - J(0.0, X, tauc))
    rhoN = 1.0/4.0 * (delta_NH**2) * (J((GammaH-GammaN)*B0, X,
    tauc) + 3.0 * J(GammaN*B0, X, tauc) + 6.0 * J((GammaH+GammaN)
    *B0, X, tauc)) +\
    ((DELTA_N*GammaN*B0)**2) * J(GammaN*B0, X, tauc)
    deltaN = math.sqrt(3) * delta_NH * DELTA_N * GammaN * B0 * ((3.0
    * (math.cos(theta_NNH))**2 - 1.0) / 2.0) * J(GammaN*B0, X, tauc)
    rhoNH = Rex[0] + 3.0/4.0 * (delta_NH**2) * (J(GammaH*B0, X,
    tauc) + J(GammaN*B0, X, tauc)) +\
    float(vicinity) * 1.0/4.0 * (delta_HH**2) * (J(0.0, X, tauc) +
    3.0 * J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) +\
    ((GammaH*B0)**2) * (sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2
    - sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sig-
    ma_Hzz) * J(GammaH*B0, X, tauc) +\
    ((DELTA_N*GammaN*B0)**2) * J(GammaN*B0, X, tauc)
    rhoHi = float(bool(vicinity)) * (1.0/4.0 * (delta_NH**2) * (J
    ((GammaH-GammaN)*B0, X, tauc) + 3.0 * J(GammaH*B0, X, tauc) +
    6.0 * J((GammaH+GammaN)*B0, X, tauc)) +\
    float(vicinity) / 4.0 * (delta_HH**2) * (J(0.0, X, tauc) + 3.0 *
    J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) +\
    ((GammaH*B0)**2) * ((sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2
    - sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sig-

```

```

ma_Hzz) * J(GammaH*B0, X, tauc))
rhoNH = float(bool(vicinity)) * (3.0/4.0 * (delta_NH**2) * (J
(GammaH*B0, X, tauc) + J(GammaN*B0, X, tauc)) + \
float(vicinity) / 4.0 * (delta_HH**2) * (J(0.0, X, tauc) + 3.0 *
J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) + \
((GammaH*B0)**2) * (sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2 -
sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sigma_Hzz)
* J(GammaH*B0, X, tauc) + \
((DELTA_N*GammaN*B0)**2) * J(GammaN*B0, X, tauc))
rhoHw = float(bool(vicinity)-float(bool(vicinity))) * (1.0/4.0 *
(delta_NH**2) * (J((GammaH-GammaN)*B0, X, tauc) + 3.0 * J(Gam-
maH*B0, X, tauc) + 6.0 * J((GammaH+GammaN)*B0, X, tauc)) + \
float(vicinity) / 4.0 * (delta_HH**2) * (J(0.0, X, tauc) + 3.0 *
J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) + \
((GammaH*B0)**2) * (sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2
- sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sig-
ma_Hzz) * J(GammaH*B0, X, tauc))
rhoNHw = float(bool(vicinity)-float(bool(vicinity))) * (3.0/
4.0 * (delta_NH**2) * (J(GammaH*B0, X, tauc) + J(GammaN*B0, X,
tauc)) + \
float(vicinity) / 4.0 * (delta_HH**2) * (J(0.0, X, tauc) + 3.0 *
J(GammaH*B0, X, tauc) + 6.0 * J(2.0*GammaH*B0, X, tauc)) + \
((GammaH*B0)**2) * (sigma_Hxx**2 + sigma_Hyy**2 + sigma_Hzz**2 -
sigma_Hxx*sigma_Hyy - sigma_Hxx*sigma_Hzz - sigma_Hyy*sig-
ma_Hzz) * J(GammaH*B0, X, tauc) + \
((DELTA_N*GammaN*B0)**2) * J(GammaN*B0, X, tauc))
M = [[rhoN, sigmaNH, deltaN, 0.0, 0.0, 0.0, 0.0], \
[sigmaNH, rhoH, deltaH, sigma, 0.0, sigma, 0.0], \
[deltaN, deltaH, rhoNH, 0.0, sigma, 0.0, sigma], \
[0.0, sigma, 0.0, rhoHi, 0.0, 0.0, 0.0], \
[0.0, 0.0, sigma, 0.0, rhoNHi, 0.0, 0.0], \
[0.0, sigma, 0.0, 0.0, 0.0, rhoHw, 0.0], \
[0.0, 0.0, sigma, 0.0, 0.0, 0.0, rhoNHw]]
return M

```

5.2.2 The Magnetic Field Profile

This file contains the values of the magnetic field along the axis of the magnet determined as described in Subheading 4.1. The first column contains the distance (in meters) to the high-field magnetic center. The second column displays the value of the magnetic field (in Tesla). A single tab separates the two columns. Table 2 shows sample lines of the file in the case of our 600 MHz spectrometer.

5.2.3 Accurate Relaxation Rates Measured with the Control of Cross-Relaxation Pathways

One file for each relaxation rate and each field must be prepared. Each file must contain three columns separated by a single tab: the amino acid number, the relaxation rate, the uncertainty/error. These files will be loaded one by one when running ICARUS.

Table 2
Organization of the file containing the magnetic field profile

0	14.1
0.001	14.1
...	...
0.698	0.0703
0.699	0.0699
0.700	0.0693

Only a few lines are shown. The first column indicates the height (in meter) above the magnetic center. The second column gives the corresponding magnetic field (in Tesla)

Table 3
Organization of a file containing high-field relaxation rates, here R_1 rates (s^{-1}) recorded at 18.8 T

2	1.458	0.022
3	1.575	0.024
4	1.617	0.024
5	1.49	0.022

If some data are missing, the user must put NA at the corresponding position (only for the relaxation rate and the uncertainty). Table 3 shows an example of such a file.

5.2.4 Low-Field Longitudinal Relaxation Rates

The experimental rates are organized in columns starting from the first line. The amino-acid numbers are in the first column. The following columns are organized by pair, with the relaxometry relaxation rates and experimental errors. These pairs of columns have to be ordered from the highest to the lowest magnetic field reached during the relaxation period τ_{relax} (see pulse sequence in Fig. 3). If any relaxation rate is missing, write NA in the given positions for this rate and its uncertainty. Columns should be separated by tabulators. Note that the magnetic fields are not specified in this file but in the trajectory file (see below). An example is given in Table 4.

5.2.5 Experimental Parameters

This file contains one line for each relaxometry experiment. The lines must be sorted according to the magnetic field at the sample position during τ_{relax} . Columns have to be organized according to the following pattern, all time-related variables are expected to be in milliseconds:

Table 4
Organization of a file containing relaxometry relaxation rates. Uncorrected longitudinal relaxation rates (and corresponding uncertainties) at 5, 3, 2, and 1.4 T are shown from left to right

2	4.169	0.089	5.796	0.108	7.364	0.158	8.539	0.151
3	4.469	0.107	6.067	0.128	7.598	0.190	9.534	0.197
4	4.531	0.140	6.074	0.150	7.719	0.217	9.630	0.258
5	4.122	0.103	5.769	0.127	7.453	0.189	9.079	0.191

1. Experiment number, chosen by the user for reference.
2. Maximum distance (in meters) to the high-field position, reached by the sample during the relaxometry experiment. The distances must be ordered in an ascending order: lowest (highest field) on the first line and largest distance (lowest field) on the last line.
3. Stabilization time after the shuttling to low field (delay τ_{st1} in Fig. 3).
4. Stabilization time after the shuttling to high field (delay τ_{st2} in Fig. 3).
5. The waiting time at high-field position corresponding to the response time of the shuttling device before the motion starts (delay τ_{w1} in Fig. 3).
6. The duration of the sample transfer from high field to low field (delay τ_{up} in Fig. 3).
7. The waiting time at low field corresponding to the response time of the shuttling device before the motion starts (delay τ_{w2} in Fig. 3).
8. The duration of the sample transfer from low field to high field (in millisecond) (delay τ_{down} in Fig. 3).
9. The relaxation delays at low field (delay τ_{relax} in Fig. 3). These values do not have to be ordered and their number can differ for different magnetic fields.

Table 5 shows an example in a case where four experiments were recorded.

5.2.6 CSA Values

This file contains the values of the CSA for each residue. The first column should be the amino acid number, and the following column(s) correspond(s) to the value(s) of the required CSA. In the FunctionsFile given as an example, the CSA of the amide nitrogen is required (defined by a single value for the anisotropy of this axially symmetric CSA), as well as the CSA of the proton (the three components of the tensor are necessary to define this fully

Table 5

Organization of a file containing the relaxometry experimental parameters. The first line (A to P) is only here for reference and should not be included in the file

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0.27	0	0.1	26	35.6	40	36.7	110	10	250	25	190	50	15	80
2	0.311	0	0.1	26	38.6	40	39.6	110	5	250	25	190	50	15	80
3	0.344	0	0.1	26	40.9	40	41.9	110	10	250	30	190	50	15	80
4	0.374	0	0.1	26	42.8	40	43.8	100	10	160	40	130	70	13	40

The columns give: experiment number (*A*), distance to the high-field magnetic center during the relaxation delay (*B*, in meter), stabilization time after the transfer to low field τ_{st1} (*C*, in millisecond), stabilization time after the transfer to high field τ_{st2} (*D*, in millisecond), delay before the shuttle leaves the high-field position τ_{w1} (*E*, in millisecond), transfer time from high field to low field τ_{up} (*F*, in millisecond), delay before the shuttle leaves the low-field position τ_{w2} (*G*, in millisecond), transfer time from low field to high field τ_{down} (*H*, in millisecond), relaxation delays τ_{relax} at low field (*I–P*, in millisecond)

Table 6

Example of a file containing the CSA values. The first line (A to E) is only here for reference and should not be included in the file

A	B	C	D	E
2	164e-6	14.6e-6	8.2e-6	2.1e-6
3	164e-6	14.6e-6	8.2e-6	2.1e-6
4	164e-6	14.6e-6	8.2e-6	2.1e-6
5	164e-6	14.6e-6	8.2e-6	2.1e-6

The residue number is given in column *A*. Column *B* provides the anisotropy of the amide nitrogen-15 nucleus. Columns *C–E* give the three components of the amide proton CSA tensor [56].

asymmetric CSA). Table 6 shows an example for this file, compatible with the provided FunctionsFile. This file can be customized and include as many principal values and angles as called in the FunctionsFile.

5.2.7 Exchange Rate Constants

This file is required only if the variable NeedRex is set to “YES” in the function file. In this file, the first column should be the amino acid number and the exchange rates are given in the second column, separated by tab. Note that these rate constants do not describe all chemical exchange contributions to line-broadening but specifically the exchange rate of the proton with the solvent.

5.3 ICARUS

Once all the input files are prepared, the program can be started by writing in the terminal:

```
python [path]/ICARUSx.x.py
where x.x is the version of the used ICARUS program.
Then, follow these steps:
```

1. Set the fitting parameters.

Set the number of iterations in ICARUS. It is recommended to start with at least three iterations and check the evolutions of the correction factors between consecutive iterations to identify convergence.

Values for the overall rotational diffusion tensor should be provided (eventually for the two considered populations).

Choose, according to the type of the used shuttling device, whether the sample is transferred between high and low fields with constant speed (pneumatic system) or constant acceleration during the first half of the motion and constant deceleration during the second half (motorized system).

Choose the time step Δt used for the evaluation of the propagator during the sample transfer. The shorter the Δt is, the more accurate the calculation. The recommended value of the time step is $\Delta t = 1$ ms. The choice of this value depends on the speed of shuttling.

Choose how many basin-hopping iterations you want to perform and if you want to use the user-defined gradients during the basin hopping minimization algorithm (the gradients must be defined in the function file to enable this option). A higher number decreases the risk that only a local optimum of the fitted function is found. If gradients are used, 50 iterations of basin hopping might be enough; otherwise, set the number of basin-hopping iterations to at least 500 (*Note*: this leads to longer calculations but is often necessary).

Set the number of random sampling points (recommended value: 10000).

2. Define the allowed ranges of the fitted parameters of the spectral density function (correlation times must be in seconds).
3. If it is available, enter the 4-letter PDB ID of the protein of interest.
4. Load all the required files. For the files containing the high-field rates, provide the type of rate and the field at which they were measured. Once a file is loaded, a button labeled “Add high-field rates” appears. Click on this button to add the next file with high-field relaxation rates.
5. Select if you want to perform a Jackknife statistical test in the first iteration of the procedure (*see* Fig. 5, $i = 0$) in which only high-field relaxation rates are included. (Recommended, and only available if more than one experiment of the same kind has been recorded).
6. Tick if the Jackknife statistical test is required to be performed on the low-field datasets. We strongly recommend using this option in order to identify possible outliers.

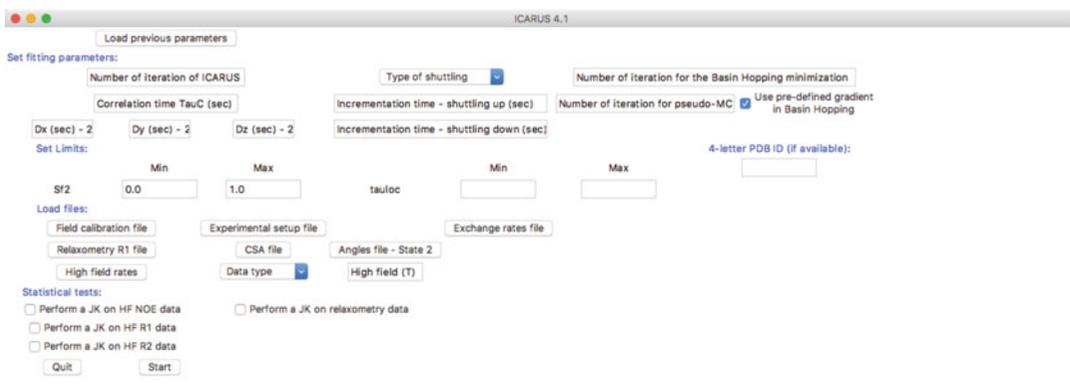


Fig. 6 Example of the graphical user interface (GUI) of ICARUS used for the analysis of ^{15}N relaxation using a model-free formalism (the same functions file as given in the example above). The “Type of shuttling” menu allows choosing between a constant speed (pneumatic system) or a constant acceleration (mechanic system) type of motion for the sample. The “Data type” for the high-field rates loading part allows choosing between all the relaxation rates for which an expression is given in the functions file. After loading a first high-field rate file, a button will appear to add another file

All of these steps can be shortcut by the use of the “Load previous parameters” top left button. As shown below (Subheading 5.4), after each run of ICARUS, a file is created which contains all the required information to start another run with the same configuration of parameters.

Figure 6 shows the window that opens when ICARUS is started for the analysis of ^{15}N relaxation using the model-free formalism (one order parameter and one correlation time). Here, we consider two states, the first one with a symmetric overall diffusion tensor and the other one with a fully asymmetric one.

5.4 Output Files of ICARUS

ICARUS creates a new folder in the working directory where the program was started to store all output files. The name of the folder is the date when the program was started (format `yy_mm_dd`). If a directory with the same name and path already exists, a number is added to the name (format `yy_mm_dd_x`) to avoid overwriting any other files or directories. All results, output files, and files containing the information about the program setup are stored in this folder:

1. The subdirectory `InputFiles` includes:
 - (a) `FunctionsFile.py`—the function file (Subheading 5.2.1).
 - (b) `FieldCalibration.txt`—the magnetic field along the sample trajectory (Subheading 5.2.2).
 - (c) `HFRelaxationRates`—a subdirectory containing the input files with accurate (usually, high-field) rates (Subheading 5.2.3).
 - (d) `RelaxometryR1.txt`—the input file with relaxometry relaxation rates (Subheading 5.2.4).

- (e) ExpSetUp.txt—the information about the sample shuttle trajectories (Subheading 5.2.5).
 - (f) CSA.txt—the input file for the CSA values (Subheading 5.2.6).
 - (g) Eventually, the file containing the exchange rates (Subheading 5.2.7).
 - (h) Parameters.txt—file summarizing the chosen setup in the GUI. This file can be loaded when ICARUS is started using the “Load previous parameters” button (Subheading 5.3) and fills all the fields automatically, which is practical when ICARUS is run several times with similar configurations. This allows the user to perform consecutive runs with limited modifications of the setup (Subheading 5.3).
 - (i) PrintFinalRelaxationMatrix.py. This python script created at the end of ICARUS allows the user to visualize the relaxation matrix. To launch the script, go in the Input-Files folder and write in terminal: `python PrintFinalRelaxationMatrix.py`. A window will open that allows the user to choose a field (in Tesla) and an amino acid number. After clicking on the “calculate” button, another window will open showing the values of the relaxation rates defined in the FunctionsFile and the whole relaxation matrix.
2. HFieldsFit: A subdirectory with figures displaying comparisons between high-field experimental data and the back-calculated values after the first iteration of ICARUS—i.e., a fit with high-field data exclusively. The figures are stored in subdirectories named after the types of relaxation rates. A file containing the values of the fitted parameters as well as the χ^2 values is created. An additional directory JKresults is created in this subdirectory if a Jackknife statistical test was selected (Subheading 5.3—step 5). Inside this folder a set of subdirectories is created with names according to the relaxation rates selected for the Jackknife test. Each directory contains figures in postscript (.ps) format, which display the variations of given parameters with the steps of the Jackknife analysis. A file is created to indicate which datasets were suppressed in all the Jackknife steps. The step number zero corresponds to the analysis with all accurate (high-field) rates.
3. FittingResults: a subdirectory containing the results:
- CorrectionFactors_RoundN.txt (N being the number of the corresponding iteration of ICARUS): a text file containing the correction factors for the relaxation rates measured at low field for all residues after the N th round of ICARUS.
 - FinalRelaxationData.txt: a text file containing the values of the relaxation rates at all high fields and corrected relaxometry rates at the end of ICARUS.

FinalFittedParameters.txt: a text file containing the fitted parameters of internal dynamics and corresponding χ^2 after the last iteration of ICARUS (the results are visualized in the PlotParameters subdirectory, see below).

The PDB subdirectory contains files that can be run in PyMOL and named the following way: PDBID_parameter.pml where PDBID is the provided PDB ID, parameter is either Chi2 or a fitted parameter name. You can open one of these files in PyMOL. It will color the structure according to the value of the chosen parameter. The user should include the raw PDB file in this folder (do not rename it!).

4. The subdirectory PlotParameters includes:
 - (a) Postscript figures with the fitted parameters of the spectral density function after the last iteration of the ICARUS program. The name of the file before the postscript extension “.ps” reflects the parameter presented in the plot.
 - (b) Chi2.ps—postscript figure visualizing χ^2 for all residues.
 - (c) JKresults: a subdirectory with the figures described in (a) and (b) for all data sets variants calculated during the Jackknife statistical test (if selected in the setup: Subheading 5.3—steps 5–6). The Jackknife step number corresponds to the rate removed in the input file.
5. Fit_of_B0.ps: a plot of experimental dependence of the magnetic field on the distance from the magnetic center (Subheading 4.1). A polynomial function fitted to this profile and used throughout the calculation in the program is also included.
6. FitAllResidues: A subdirectory with figures displaying comparisons between the experimental data and the back-calculated values after the last iteration of the ICARUS program. The figures are stored in subdirectories named after the high-field rates. The relaxometry rates are plotted together with their corrected values and high-field longitudinal relaxation rates. A subdirectory called Profiles also contains the measured, corrected (for relaxometry rates only), and back-calculated rates at each magnetic field throughout the sequence of the protein.

5.5 Results for the Analysis of High-Resolution Relaxometry Data in Ubiquitin

Published high-resolution relaxometry data [23] were analyzed with the new version of the program ICARUS. The results for backbone internal dynamics are presented in Fig. 7. The analysis was based on the longitudinal ^{15}N relaxation rates, and steady-state NOE data recorded at three high fields (14.1, 18.8, 22.3 T) as well as eight longitudinal relaxation rates measured at low field with the sample shuttle. Two additional protons in the vicinity of the ^{15}N - ^1H spin pair were taken into account (using a distance $d_{\text{HH}'} = 2.1 \text{ \AA}$) for the ICARUS analysis. As in our first analysis of this dataset [23], order parameters in Fig. 7b show that high-

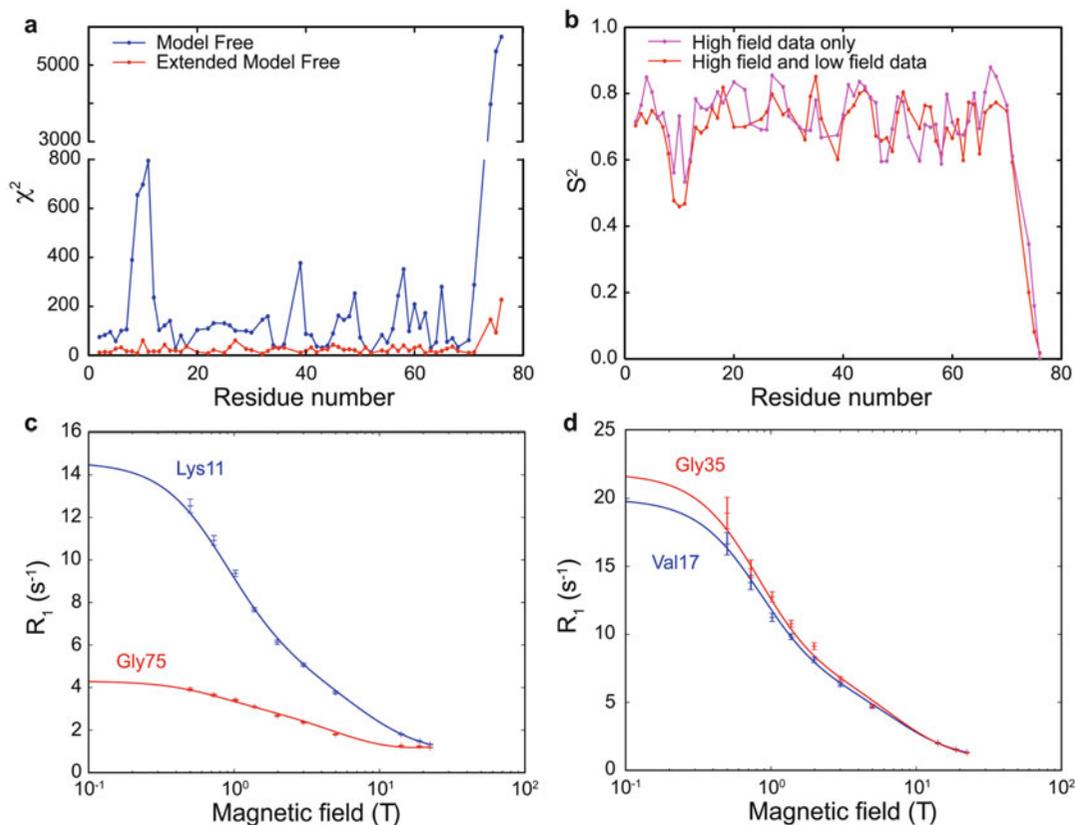


Fig. 7 Results of the analysis of relaxation rates measured for backbone ^{15}N - ^1H spin pairs in Ubiquitin. (a) χ^2 values of the fit using the Model Free (*blue*) or Extended Model Free (*red*) form of the spectral density function. (b) Order parameters obtained after fitting the high-field relaxation rates exclusively (*magenta*) and all recorded relaxation rates (*red*) using the Extended Model Free form of the spectral density function. (c) and (d) show nuclear magnetic relaxation dispersion profiles with the experimental longitudinal relaxation rates measured at high fields (14.1, 18.8, and 22.3 T) and corrected longitudinal rates measured at low field for four residues. The *solid lines* represent the fitted nuclear magnetic relaxation dispersion profiles. We show two examples of dynamic residues (Lys11 and Gly75) as well as two more rigid residues (Val17 and Gly35)

resolution relaxometry reveals more nanosecond timescale motions in flexible regions than a conventional analysis of high-field relaxation. The analysis based on the relaxometry data is more sensitive to nanoseconds motions, which is documented by a significantly lower order parameter S^2 in the $\beta 1$ - $\beta 2$ turn (in particular residues 9–11) in agreement with studies using different techniques [52–55].

Acknowledgments

This work was funded by the European Research Council (ERC) under the European Community Seventh Framework Program (FP7/2007–2013), ERC Grant Agreement 279519 (2F4BIODYN).

References

- Charlier C, Cousin SF, Ferrage F (2016) Protein dynamics from nuclear magnetic relaxation. *Chem Soc Rev* 45:2410–2422. doi:10.1039/c5cs00832h
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N -inverse detected heteronuclear NMR-spectroscopy - application to staphylococcal nuclease. *Biochemistry* 28:8972–8979
- Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104(8):3623–3640
- Peng JW, Wagner G (1992) Mapping of spectral density-functions using heteronuclear NMR relaxation measurements. *J Magn Reson* 98(2):308–332
- Tjandra N, Feller SE, Pastor RW, Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J Am Chem Soc* 117:12562–12566
- Agarwal V, Xue Y, Reif B, Skrynnikov NR (2008) Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: a similarity revealed. *J Am Chem Soc* 130(49):16611–16621. doi:10.1021/ja804275p
- Duchardt E, Schwalbe H (2005) Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR ^{13}C relaxation. *J Biomol NMR* 32(4):295–308. doi:10.1007/s10858-005-0659-x
- Ferrage F, Pelupessy P, Cowburn D, Bodenhausen G (2006) Protein backbone dynamics through ^{13}C - $^{13}\text{C}^{\alpha}$ cross-relaxation in NMR spectroscopy. *J Am Chem Soc* 128(34):11072–11078. doi:10.1021/ja0600577
- Rinntenthal J, Richter C, Nozinovic S, Furtig B, Lopez JJ, Glaubitz C, Schwalbe H (2009) RNA phosphodiester backbone dynamics of a perdeuterated cUUCGg tetraloop RNA from phosphorus-31 NMR relaxation analysis. *J Biomol NMR* 45(1-2):143–155. doi:10.1007/s10858-009-9343-x
- Muhandiram DR, Yamazaki T, Sykes BD, Kay LE (1995) Measurement of ^2H $T_{1\rho}$ relaxation times in uniformly ^{13}C -labeled and fractionally ^2H -labeled proteins in solution. *J Am Chem Soc* 117:11536–11544
- Lee AL, Flynn PF, Wand AJ (1999) Comparison of ^2H and ^{13}C NMR relaxation techniques for the study of protein methyl group dynamics in solution. *J Am Chem Soc* 121(12):2891–2902. doi:10.1021/ja983758f
- Millet O, Muhandiram DR, Skrynnikov NR, Kay LE (2002) Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuterium in ^{13}C -labeled and fractionally ^2H -enriched proteins in solution. *J Am Chem Soc* 124:6439–6448
- Skrynnikov NR, Millet O, Kay LE (2002) Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. *J Am Chem Soc* 124(22):6449–6460. doi:10.1021/ja012498q
- Sheppard D, Li D-W, Brueschweiler R, Tugarinov V (2009) Deuterium spin probes of backbone order in proteins: ^2H NMR relaxation study of deuterated carbon alpha sites. *J Am Chem Soc* 131(43):15853–15865. doi:10.1021/ja9063958
- Liao X, Long D, Li D-W, Brueschweiler R, Tugarinov V (2012) Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. *J Phys Chem B* 116(1):606–620. doi:10.1021/jp209304c
- Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of magnetic resonance in one and two dimensions. Clarendon Press, Oxford
- Kumar A, Grace RCR, Madhu PK (2000) Cross-correlations in NMR. *Prog NMR Spectrosc* 37(3):191–319
- Goldman M (2001) Formal theory of spin-lattice relaxation. *J Magn Reson* 149:160–187
- Nicholas MP, Eryilmaz E, Ferrage F, Cowburn D, Ghose R (2010) Nuclear spin relaxation in isotropic and anisotropic media. *Prog Nucl Magn Reson Spectrosc* 57:111
- Redfield AG (2003) Shuttling device for high-resolution measurements of relaxation and related phenomena in solution at low field, using a shared commercial 500 MHz NMR instrument. *Magn Reson Chem* 41(10):753–768. doi:10.1002/mrc.1264
- Redfield AG (2012) High-resolution NMR field-cycling device for full-range relaxation and structural studies of biopolymers on a shared commercial instrument. *J Biomol NMR* 52(2):159–177. doi:10.1007/s10858-011-9594-1
- Chou CY, Chu ML, Chang CF, Huang TH (2012) A compact high-speed mechanical sample shuttle for field-dependent high-resolution solution NMR. *J Magn Reson* 214:302–308. doi:10.1016/j.jmr.2011.12.001
- Charlier C, Khan SN, Marquardsen T, Pelupessy P, Reiss V, Sakellariou D,

- Bodenhausen G, Engelke F, Ferrage F (2013) Nanosecond time scale motions in proteins revealed by high-resolution NMR relaxometry. *J Am Chem Soc* 135(49):18665–18672. doi:10.1021/ja409820g
24. Goldman M (1988) Quantum description of high resolution NMR in liquids. Clarendon Press, Oxford
 25. Abragam A (1961) Principles of nuclear magnetism. Oxford University Press, Oxford
 26. Redfield AG (1965) Theory of relaxation processes. *Adv Magn Reson* 1:1–32
 27. Cavanagh J, Fairbrother WJ, Palmer AG III, Rance M, Skelton NJ (2006) Protein NMR spectroscopy: principles and practice. Academic Press, San Diego, CA
 28. Khan SN, Charlier C, Augustyniak R, Salvi N, Déjean V, Bodenhausen G, Lequin O, Pelu-pessy P, Ferrage F (2015) Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation: a simple array of correlation times. *Biophys J* 109:988
 29. Brüschweiler R, Wright PE (1994) NMR order parameters of biomolecules - a new analytical representation and application to the gaussian axial fluctuation model. *J Am Chem Soc* 116:8426–8427
 30. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules I. Theory and range of validity. *J Am Chem Soc* 104:4546–4559
 31. Polimeno A, Freed JH (1995) Slow motional ESR in complex fluids - the slowly relaxing local-structure model of solvent cage effects. *J Phys Chem* 99(27):10995–11006. doi:10.1021/j100027a047
 32. Meirovitch E, Shapiro YE, Polimeno A, Freed JH (2006) Protein dynamics from NMR: the slowly relaxing local structure analysis compared with model-free analysis. *J Phys Chem A* 110(27):8366–8396
 33. Buevich AV, Baum J (1999) Dynamics of unfolded proteins: incorporation of distributions of correlation times in the model free analysis of NMR relaxation data. *J Am Chem Soc* 121(37):8671–8672
 34. Buevich AV, Shinde UP, Inouye M, Baum J (2001) Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J Biomol NMR* 20(3):233–249. doi:10.1023/a:1011243116136
 35. Ochsenbein F, Neumann JM, Guittet E, Van Heijenoort C (2002) Dynamical characterization of residual and non-native structures in a partially folded protein by ^{15}N NMR relaxation using a model based on a distribution of correlation times. *Protein Sci* 11(4):957–964. doi:10.1110/ps.4000102
 36. Abyzov A, Salvi N, Schneider R, Maurin D, Ruigrok RWH, Jensen MR, Blackledge M (2016) Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent NMR relaxation. *J Am Chem Soc* 138(19):6240–6251. doi:10.1021/jacs.6b02424
 37. Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM (1990) Deviations from the simple 2-parameter model-free approach to the interpretation of Nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* 112(12):4989–4991
 38. Mandel AM, Akke M, Palmer AG III (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246:144–163
 39. Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* 16(1):23–28
 40. Berlin K, Castaneda CA, Schneidman-Duhovny D, Sali A, Nava-Tudela A, Fushman D (2013) Recovering a representative conformational ensemble from underdetermined macromolecular structural data. *J Am Chem Soc* 135(44):16595–16609. doi:10.1021/ja4083717
 41. d’Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR* 40(2):107–119. doi:10.1007/s10858-007-9214-2
 42. d’Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR* 40(2):121–133. doi:10.1007/s10858-007-9213-3
 43. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
 44. Ferrage F, Piserchio A, Cowburn D, Ghose R (2008) On the measurement of ^{15}N - $\{^1\text{H}\}$ nuclear overhauser effects. *J Magn Reson* 192(2):302–313. doi:10.1016/j.jmr.2008.03.011

45. Ammann C, Meier P, Merbach AE (1982) A simple multi-nuclear NMR thermometer. *J Magn Reson* 46(2):319–321. doi:[10.1016/0022-2364\(82\)90147-0](https://doi.org/10.1016/0022-2364(82)90147-0)
46. Lacey ME, Webb AG, Sweedler JV (2002) On-line temperature-monitoring in a capillary electrochromatograph frit using microcoil NMR. *Anal Chem* 74(17):4583–4587. doi:[10.1021/ac025741s](https://doi.org/10.1021/ac025741s)
47. Berger S, Braun S (2004) 200 and more NMR experiments: a practical course. Wiley-VCH, Weinheim
48. Findeisen M, Brand T, Berger S (2007) A ^1H -NMR thermometer suitable for cryoprotocols. *Magn Reson Chem* 45(2):175–178. doi:[10.1002/mrc.1941](https://doi.org/10.1002/mrc.1941)
49. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
50. Fushman D, Varadan R, Assfalg M, Walker O (2004) Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog Nucl Magn Reson Spectrosc* 44(3-4):189–214. doi:[10.1016/j.pnmrs.2004.02.001](https://doi.org/10.1016/j.pnmrs.2004.02.001)
51. Ferrage F (2012) Protein dynamics by ^{15}N nuclear magnetic relaxation. *Meth Mol Biol* 831:141–163
52. Lakomek NA, Carlomagno T, Becker S, Griesinger C, Meiler J (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J Biomol NMR* 34(2):101–115
53. Lakomek NA, Walter KFA, Fares C, Lange OF, de Groot BL, Grubmüller H, Bruschweiler R, Munk A, Becker S, Meiler J, Griesinger C (2008) Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J Biomol NMR* 41(3):139–155. doi:[10.1007/s10858-008-9244-4](https://doi.org/10.1007/s10858-008-9244-4)
54. Salmon L, Bouvignies G, Markwick P, Lakomek N, Showalter S, Li DW, Walter K, Griesinger C, Bruschweiler R, Blackledge M (2009) Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. *Angew Chem Int Ed Engl* 48(23):4154–4157. doi:[10.1002/anie.200900476](https://doi.org/10.1002/anie.200900476)
55. Salmon L, Bouvignies G, Markwick P, Blackledge M (2011) Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry* 50(14):2735–2747. doi:[10.1021/bi200177v](https://doi.org/10.1021/bi200177v)
56. Loth K, Pelupessy P, Bodenhausen G (2005) Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. *J Am Chem Soc* 127(16):6062–6068

Chapter 10

Characterizing Protein Dynamics with NMR $R_{1\rho}$ Relaxation Experiments

Francesca Massi and Jeffrey W. Peng

Abstract

The measurement of $R_{1\rho}$, the longitudinal relaxation rate constant in the rotating frame, is one of the few available methods to characterize the μ s-ms functional dynamics of biomolecules. Here, we focus on ^{15}N $R_{1\rho}$ experiments for protein NH groups. We present protocols for both on- and off-resonance ^{15}N $R_{1\rho}$ measurements needed for relaxation dispersion studies, and describe the data analysis for extracting kinetic and thermodynamic parameters characterizing the motional processes.

Key words Dynamic process, Chemical exchange, ^{15}N spin relaxation, Rotating frame relaxation, NMR spectroscopy, Millisecond-to-microsecond dynamics, Relaxation dispersion

1 Introduction

Over the past 60 years, the number of atomic-resolution structures of proteins, nucleic acids, and their complexes has grown without pause, due to x-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, and cryo-electron microscopy [1]. This persistent structural focus reflects a core principle of biochemistry: “structure dictates function.” Yet, it has also been long recognized that biomolecules have internal motions [2–5]. Biomolecules are generally not limited to “the” experimentally determined structure, but instead, sample a distribution of structures (conformations). Moreover, mounting evidence shows that conformational dynamics can be crucial for function [6–9]. Hence, the characterization of biomolecular dynamics in solution is an essential complement to structural studies for a more complete understanding of function.

NMR spectroscopy is a powerful technique to determine both the structure and the dynamics of biomolecules in solution [10, 11]. For studies of dynamics, spin relaxation measurements are a rich source of dynamic information. There now exists a broad range of relaxation experiments to study functional motions on both

“fast” (picosecond-nanosecond; ps-ns) and “slow” (microsecond-millisecond; μ s-ms) time scales [10, 12, 13].

This chapter concerns relaxation experiments for studying μ s motions: measurements of $R_{1\rho}$, the longitudinal relaxation rate constant in the rotating frame. Our focus is on applications to protein dynamics, although these methods have also been developed and successfully employed to study nucleic acid dynamics [14, 15].

Protein motions on the μ s-ms time scale are often large-scale conformational rearrangements, such as loop and domain motions supporting catalysis, ligand binding, or allosteric regulation. $R_{1\rho}$ relaxation experiments can determine the kinetics of these dynamic processes. Moreover, in many favorable cases, $R_{1\rho}$ relaxation experiments can give a full characterization of the process, including not only kinetic parameters, but also thermodynamic and structural parameters pertaining to the exchanging conformations [13, 16–18].

$R_{1\rho}$ pulse sequences have been developed for a variety of protein nuclei, including ^{15}N , ^1H in amide groups [19–21], and ^{13}C in CO [22], C^α [23], methyl groups [24, 25], and aromatic side chains [26]. The ability to monitor both backbone and side-chain nuclei provides a broader description of the motional process in both space (distribution of internal sites) and time (range of time scales), and thus, a more accurate understanding of their role in function.

In this chapter, we focus on ^{15}N $R_{1\rho}$ measurements for protein NH groups. We begin with a brief overview of key theoretical concepts underlying $R_{1\rho}$ relaxation. We then give a detailed description of both on- and off-resonance ^{15}N $R_{1\rho}$ relaxation measurements for relaxation dispersion studies. We conclude with a description of the data analysis for extracting kinetics and thermodynamics parameters. We note that while our focus is on ^{15}N , the key points are also applicable to ^{13}C . The latter has the additional challenge of controlling homonuclear ^{13}C - ^{13}C coupling artifacts.

1.1 Theory

In this section, we provide the minimal theoretical background of $R_{1\rho}$ relaxation necessary to understand the pulse schemes described further below. A more detailed description of the theory and practice of $R_{1\rho}$ relaxation methods has been presented elsewhere [13].

We consider a μ s-ms dynamic process in which a protein undergoes equilibrium exchange between two conformational states, A and B. Examples of such two-state exchange include ligand binding and local unfolding. The corresponding two-state equilibrium is



The symbols k_1 and k_{-1} are the forward and reverse rate constants, respectively. Their sum, $k_{ex} = k_1 + k_{-1}$, is the net exchange

rate constant that characterizes the overall rapidity of exchange [27], and is useful in predicting the effect of the exchange on NMR parameters.

The resonance frequencies of nuclear spins depend on their local magnetic environments through their chemical shifts, which in turn depend on local conformation. Thus, in Eq. 1, a subset of protein spins within or near the site of conformational dynamics would have different resonance frequencies ω_A in state A versus ω_B in state B. Protein exchange between the A and B states toggles the spin resonance frequencies between ω_A and ω_B . Hence, the spin precession frequencies become stochastic functions of time, which accelerates the de-phasing of transverse magnetization and increases the transverse relaxation rate constant, R_2 :

$$R_2 = R_2^0 + R_{ex} \quad (2)$$

In Eq. (2), R_2 is the measured transverse relaxation rate constant, R_2^0 accounts for transverse relaxation from non-exchange origins (e.g., dipolar and chemical shift anisotropy relaxation mechanisms), and R_{ex} represents the contribution of the chemical exchange process.

$R_{1\rho}$ relaxation experiments can quantify R_{ex} to provide information about the exchange dynamics. They achieve this by applying a continuous-wave radio-frequency (RF) field, also called the spin-lock field, with strength ω_{SL} at a carrier frequency ω_{RF} . Seen from the rotating frame (Fig. 1), the spins experience an effective field, ω_e , which is the vector sum of the static spin-lock field with magnitude ω_{SL} in the transverse plane, and the resonance offset field, $\Omega_i = \omega_i - \omega_{RF}$ along the z-axis. The effective field ω_e has the tilt angle $\theta = \arctan(\omega_{SL}/\Omega)$, and magnitude $\omega_e = (\Omega^2 + \omega_{SL}^2)^{1/2}$ [28]. $R_{1\rho}$ experiments lock bulk magnetization along the direction of ω_e , and measure its relaxation rate constant, $R_{1\rho}$, along ω_e [28]. The measured value of $R_{1\rho}$ depends on the orientation and magnitude of ω_e , which may be changed by the spin-lock strength ω_{SL} , its offset Ω , or both.

In the case of exchange between conformations A and B (Eq. 1), the magnetization of each state will be locked around its distinct effective field, ω_{eA} and ω_{eB} , with tilt angles $\theta_A = \arctan(\omega_{SL}/\Omega_A)$ and $\theta_B = \arctan(\omega_{SL}/\Omega_B)$, per Fig. 1 [28]. As the two effective fields become nearly collinear, one observes a decrease in the extent of magnetization de-phasing due to chemical exchange and consequent line broadening. The collinear condition is approximated in two limits: (a) the magnitude of the resonance offset far exceeds the breadth of resonance frequencies (both Ω_A and $\Omega_B \gg \Delta\omega$ where $\Omega_A = \omega_A - \omega_{RF}$, $\Omega_B = \omega_B - \omega_{RF}$ and $\Delta\omega = \omega_A - \omega_B$); (b) the spin-lock strength ω_{SL} far exceeds the breadth of resonance frequencies ($\omega_{SL} \gg \Delta\omega$). In limit (a), the ω_{eA} and ω_{eB} field vectors have similar tilt angles near the “z” axis; in limit (b), they are both nearly aligned near the transverse plane.

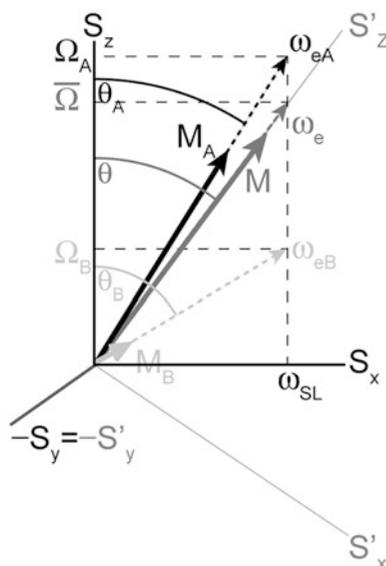


Fig. 1 Representation of the vectors for the applied RF field (ω_{SL}), magnetizations (M_A , M_B and M), effective fields (ω_{eA} , ω_{eB} and ω_e) and of resonance offsets (Ω_A , Ω_B , and $\bar{\Omega}$) for the two exchanging conformational states, A and B, and for their average in an $R_{1\rho}$ experiment. The tilt angle θ , defined as $\theta = \arctan(\omega_1/\bar{\Omega})$, gives relative orientation of the laboratory frame (S_x, S_y, S_z shown in *black*) and tilted frame (S'_x, S'_y, S'_z shown in *grey*). The orientations of the effective fields of the two exchanging states are given by the tilt angles $\theta_A = \arctan(\omega_{SL}/\Omega_A)$ and $\theta_B = \arctan(\omega_{SL}/\Omega_B)$

$R_{1\rho}$ experiments measure the dependence of $R_{1\rho}$ on the spinlock strength, ω_{SL} , and the frequency offset, Ω_i . The results—plots of $R_{1\rho}$ versus the effective field parameters—are called dispersion profiles. Critically, the shape of these profiles provides information about the kinetics and thermodynamics of the exchange process. The values of ω_e applied in the experiments determine the time scale of the motional process probed, $k_{ex} \approx \omega_e/2\pi$. Typical $\omega_e/2\pi$ values range between 25 and 6000 Hz, corresponding to motions within the μ s–ms time scales.

For two-state exchange, there are several model functions that have been developed and successfully applied [13]. They have been derived under different conditions and using different approximations. One of the simplest and most accurate expressions for $R_{1\rho}$ is that of Miloushev and Palmer [29]:

$$R_{1\rho} = R_1 \cos^2 \theta + (R_2^0 + R_{ex}) \sin^2 \theta \quad (3)$$

$$R_{ex} = \frac{p_A p_B \Delta \omega^2 k_{ex}}{\frac{\omega_{eA}^2 \omega_{eB}^2}{\omega_e^2} + k_{ex}^2 - \sin^2 \theta p_A p_B \Delta \omega^2 \left[1 + \frac{2k_{ex}^2 (p_A \omega_{eA}^2 + p_B \omega_{eB}^2)}{\omega_e^2 k_{ex}^2 + \omega_{eA}^2 \omega_{eB}^2} \right]} \quad (4)$$

Equations (3) and (4) assume that the non-exchange relaxation rate constants (the longitudinal rate constant R_1 , and R_2^0) are the

same for both A and B. The exchange parameters k_{ex} , $\Delta\omega$, p_A , and p_B are defined as before, and

$$\begin{aligned}\theta &= \arctan \frac{\omega_{SL}}{\bar{\Omega}} \\ \bar{\Omega} &= p_A \Omega_A + p_B \Omega_B \\ \omega_e &= \sqrt{\bar{\Omega}^2 + \omega_{SL}^2} \\ \omega_{eA} &= \sqrt{\Omega_A^2 + \omega_{SL}^2} \\ \omega_{eB} &= \sqrt{\Omega_B^2 + \omega_{SL}^2}\end{aligned}\quad (5)$$

When the A and B populations are skewed ($p_A > p_B$), Eq. (4) simplifies to that given by Trott and Palmer [18]

$$R_{ex} = \frac{p_A p_B \Delta\omega^2 k_{ex}}{k_{ex}^2 + \omega_{eA}^2 \omega_{eB}^2 / \omega_e^2} \quad (6)$$

Skewed populations ($p_A > p_B$) are common in proteins, as even small differences in free energy between the exchanging states translate into large population differences.

When the populations are highly skewed ($p_A \gg p_B$), $\bar{\Omega} \approx \Omega_A$ and $\omega_e \approx \omega_A$, and Eq. (6) simplifies further to [18]

$$R_{ex} = \frac{p_A p_B \Delta\omega^2 k_{ex}}{k_{ex}^2 + \omega_{eB}^2}, \quad (7)$$

In the hi-skew limit, R_{ex} maximizes when the applied RF field matches the Larmor frequency of the minor state B. Thus, the Larmor frequency of the minor state B, often invisible, can be determined by measuring R_{ex} as a function of the offset at a single magnetic field strength.

A different simplification occurs when the exchange is fast ($k_{ex} \gg \Delta\omega$), and so $\omega_{eA}\omega_{eB}/\omega_e \approx \omega_e$. Eq. (6) then reduces to [28, 30]

$$R_{ex} = \frac{p_A p_B \Delta\omega^2 k_{ex}}{k_{ex}^2 + \omega_e^2} \quad (8)$$

A comparison of Eqs. (6)–(8) shows that the fast exchange limit (Eq. (8)) allows for the extraction of two parameters: k_{ex} and the product $p_A p_B \Delta\omega^2$. When the exchange process is intermediate-to-slow, however, it is possible to extract all the parameters, k_{ex} , p_A , p_B , $\Delta\omega$, and ω_B , the frequency of the minor state (typically invisible because $p_A > p_B$). Kay and coworkers have shown that, in this motional regime, k_{ex} and p_A and p_B values obtained via a single spin-lock field strength $\omega_1 \sim k_{ex}$ and static magnetic field strength showed modest deviations of 35% or less from reference values obtained via measurements at multiple static fields [17]. This is an

advantage of the $R_{1\rho}$ over CPMG experiments, for which accurate characterization of exchange requires data from at least two different static magnetic field strengths to avoid artifacts from the inter-parameter correlations inherent in single field data analysis [31]. Nonetheless, a rigorous statistical assessment of the quality of fit, by generating confidence limits on the parameter estimates, is essential when using single static magnetic field $R_{1\rho}$. Such limits can be determined using resampling methods like the bootstrap [32–34].

Finally, for those interested in moving beyond the two-state scheme, we point to the recent results of Palmer and coworkers [35], which give $R_{1\rho}$ for N-exchange coupled states with one dominant state.

2 Materials

2.1 NMR Samples

1. Methanol or ethylene glycol for temperature calibration [36].
2. Uniformly labeled ^{15}N protein of interest at concentration > 0.5 mM. If working with large proteins, deuteration of non-exchangeable protons will be necessary to reduce ^1H dipolar relaxation. In this case, uniformly ^{15}N labeled and fractionally ^2H labeled protein at concentration > 0.5 mM will be needed. Protein concentrations higher than 0.5 mM are required to obtain high signal-to-noise ratio ($S/N \geq 40\text{--}50$) without using extensive spectrometer time. High S/N is necessary to ensure accurate data analysis and fitting.

Careful determination of the sample conditions to achieve maximal sample stability is important in $R_{1\rho}$ relaxation studies, as data collection can take several days.

2.2 NMR Tubes

Both the regular and Shigemi tubes can be used. Formation of bubbles during prolonged data collection could compromise the quality of the data. This is a particular nuisance when using Shigemi tubes; because of their smaller sample volume, the bubbles are in the detection volume. Degassing samples before data collection will minimize bubble formation.

3 Methods

3.1 Temperature Calibration

Temperature calibration involves measuring chemical shift differences of selected resonances in standard samples including neat methanol (for $T < 25$ °C) and ethylene glycol (for $T > 25$ °C) [36]. Temperature calibration is essential for consistent and reproducible dispersion data. One reason, of course, is that the exchange-related quantities sensed by dispersion (chemical shifts, populations, and rate constants) may have significant temperature

dependence. Another reason is that full extraction of exchange parameters often requires dispersion data collected at multiple static magnetic field strengths, and thus, on different spectrometer systems. Such is the case for fast chemical exchange processes, which require dispersion data at two static fields for an accurate determination of kinetic and thermodynamic parameters. To optimize the consistency of dispersion data collected at multiple static fields (spectrometers), one should verify the consistency of sample temperatures on the different spectrometers. Temperature calibration can expose possible system-dependent differences between the nominal sample temperature readout and the actual sample temperature.

3.2 Pulse Calibration

^1H and ^{15}N pulses should be properly calibrated. First, carry out sample shimming and tuning/matching of the ^1H and ^{15}N channels. Then, set the ^1H carrier on the water signal and collect a 1D ^1H spectrum. Set the ^1H pulse length to that expected for a 360° flip angle, and then vary the pulse length to obtain a null signal. The 90° pulse length is the optimized 360° pulse length divided by four. The ^{15}N pulse can be calibrated by collecting a 1D version of a $^{15}\text{N} - ^1\text{H}$ HSQC experiment (keep the ^{15}N t_1 fixed at one value). Multiply the first ^{15}N 90° pulse length by two to get an approximate 180° pulse. Optimize this pulse length to get a null signal. The ^{15}N 90° pulse length is the optimized pulse length divided by two.

3.3 Calibration of RF Field Strength

The amplitude of the ω_{SL} RF field can be calibrated by collecting two 2D $^{15}\text{N} - ^1\text{H}$ experiments, one without decoupling during the acquisition period, the other using an off-resonance continuous wave (CW) decoupling pulse during the acquisition period [10]. The reduced scalar coupling constant measured in the presence of the CW decoupling, J_{eff} , is related to the resonance offset, Ω , and field strength, ω_{SL} , by

$$\tan \theta = \sqrt{\left(J_0/J_{eff}\right) - 1} = \omega_{SL}/\Omega \quad (9)$$

where J_0 is scalar coupling measured in the absence of decoupling. By measuring the scalar coupling constant for every resonance in the presence and absence of the off-resonance decoupling pulse, the ratio J_0/J_{eff} can be determined. Per Eq. (9), the RF field strength ω_{SL} can be determined by plotting $\tan \theta$ versus $1/\Omega$. An alternative method measures the distribution of the ^{15}N $\omega_{SL} = -\gamma B_1$ field, using the spin-lock as the indirect t_1 dimension in a 2-D refocused HSQC spectrum [37].

3.4 Optimization of ^{15}N Carrier Frequency, Spectral Width, and Number of t_1 Increments.

Collect a preliminary 2D HSQC experiment to optimize the ^{15}N carrier frequency, spectral width, and number of t_1 increments. The ^{15}N carrier frequency should be optimized to be in the middle of the spectral region of interest. The spectral width should be the smallest possible to ensure that: (a) aliased peaks do not overlap with non-aliased peaks; (b) a ~ 1 ppm region without resonances resides on each side of the spectrum. The number of t_1 time increments (complex time points) in the ^{15}N dimension should be high enough to obtain good resolution of all the peaks of interest.

3.5 $R_{1\rho}$ Relaxation Dispersion Experiments for ^{15}N Nuclei

Here, we describe several experiments to measure $R_{1\rho}$ relaxation of ^{15}N nuclei. $R_{1\rho}$ experiments have been categorized as: (a) on- and off-resonance experiments depending on the size of the offset, $\Omega_i = \omega_i - \omega_{RF}$; (b) strong and weak-field experiments, depending on the magnitude of the applied RF field, ω_{SL} . High spin-lock fields are necessary for fast exchange processes with short time scales, while low spin-lock fields are essential for characterizing intermediate-to-slow exchange processes occurring on longer time scales. The 2D pulse sequences for on- and off-resonance experiments that use both strong and weak fields are shown in Fig. 2a–c. The pulse sequences shown in Fig. 2a, b are heteronuclear single quantum coherence (HSQC) experiments with a spin lock relaxation period of total length T [38, 39]. Fig. 2c shows a transverse relaxation optimized spectroscopy (TROSY) [40] experiment appropriate for larger proteins (e.g., $M_r > 20,000$), which selects the slowly relaxing component of a J -coupled multiplet to measure $R_{1\rho}$ [41]. Fig. 2d shows a 1D pulse sequence that uses selective Hartmann-Hahn polarization transfers [42] to measure the $R_{1\rho}$ relaxation rate of a single resonance of interest [17]. This 1D scheme can also be used for on- and off-resonance experiments with strong and weak spin-lock field strengths [17, 43].

All the experiments share the following features:

1. In all the $R_{1\rho}$ relaxation experiments, the ^{15}N magnetization needs to be aligned along the direction of the effective field prior to the application of the spin-lock field. In the experiments shown in Fig. 2, alignment is achieved with different approaches. In Fig. 2a and c alignment is achieved using an adiabatic pulse [38, 44]. For weak spin-lock fields, the requisite adiabatic ramp can become prohibitively long because of significant relaxation losses; thus, these experiments are appropriate for spin-lock fields $> \sim 450$ Hz [45].

Alternatively, a pulse/free precession element can be used as shown in Fig. 2b [46, 47]: $90^\circ_\phi - \chi - 90^\circ_{\phi-\pi/2}$ where $\phi = x$ or y , $\chi = 1/\omega_{SL}$. Good alignment ($\leq 5\%$ error) is obtained when $|\Omega|/\omega_{SL} \leq 0.4$. For this reason, when using a weak spin-lock field, one focuses on a small region of the spectrum for accurate results. Multiple 2D $R_{1\rho}$ experiments, with different spin-lock

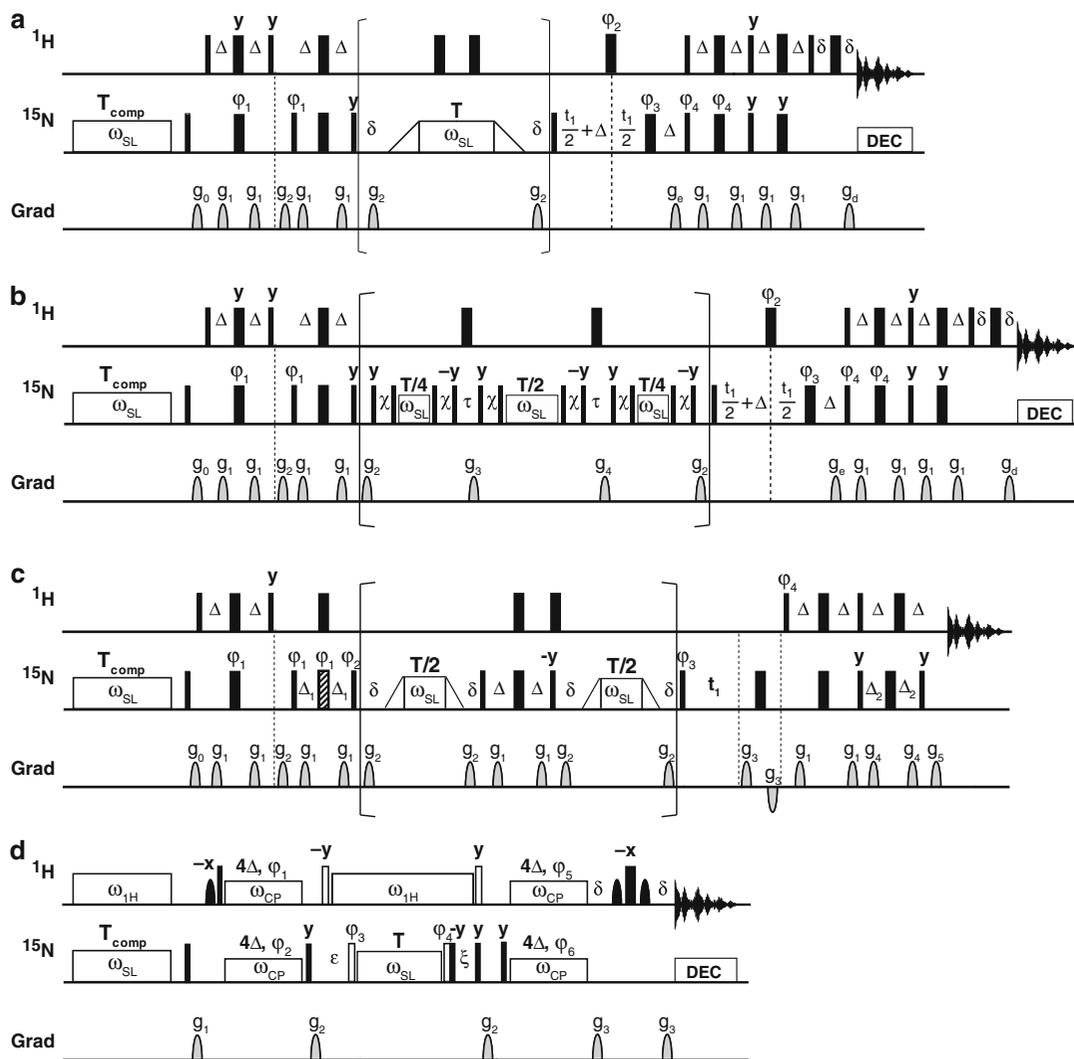


Fig. 2 Pulse sequences for ^{15}N $R_{1\rho}$ relaxation measurements. Narrow and wide filled bars indicate 90° and 180° pulses, respectively. Open rectangular boxes denote ^{15}N and ^1H spin-lock and CW irradiation periods. All pulses and spin-locks are applied with phase x unless otherwise indicated. Delays are $\Delta = 1/4J^{NH} = 2.7$ ms, and δ is long enough to encompass the enclosed gradients and gradient recovery delays. The total length of the spin-lock period is T . The duration of the heat compensating spin lock at the beginning of each scan, T_{comp} , is discussed in Subheading 3.5 [13]. In a-c all ^1H pulses are applied with the carrier frequency centered on the water resonance with the exception of those applied during the periods shown inside the brackets, which are applied in the center of the amide region. (a) and (b) The phase cycle is as follows: $(\varphi_1) x, -x$; $(\varphi_2) 4(x), 4(-x)$; $(\varphi_3) 2(x), 2(y), 2(-x), 2(-y)$; $(\varphi_4) x$; $(\varphi_{rec}) x, -x, -x, x$. The gradients g_e and g_d are used for gradient coherence selection. Gradient PEP coherence selection is obtained by inverting the sign of gradient g_e and phase φ_4 [53, 54]. ^{15}N decoupling during acquisition is performed using a 1.25 kHz GARP sequence [55]. (a) Off-resonance $R_{1\rho}$ experiment. Only within the bracketed period, is the ^{15}N carrier frequency moved by the desired offset, Ω , from spectral center (generally ~ 118 ppm). The triangles represent the tanh/tan adiabatic sweeps used to rotate the magnetization from the z -axis to the direction of the effective field, and back to the z -axis [56]. The tanh/tan adiabatic pulse lengths are 4–10 ms. The gradients are sine shaped, with durations and strengths as follows: $g_0(z)$ (2.0 ms, 9 G/cm), $g_1(z)$ (0.2 ms, 4 G/cm), $g_2(z)$ (1.5 ms, 10 G/cm), $g_e(xyz)$

carrier offsets, might be needed to accurately measure the relaxation rates of all the spins of interest. This scheme (Fig. 2b) has been used to accurately measure $R_{1\rho}$ rate constants using spin-lock field strength as low as 150 Hz for ^{15}N spins of amide groups in proteins, where the scalar coupling constant $J_{NH} \sim 92$ Hz [39].

In the 1D experiment (Fig. 2d), only a single resonance is probed at time; it is therefore easy to align the resonance of interest along the direction of the effective field. Two pulses, $90^\circ_y \theta^\circ_\phi$ are applied before the spin-lock period, where $\phi = -y$ if the offset of the resonance from the spin-lock carrier frequency, Ω , is negative, $\phi = y$ if Ω is positive, and $\theta^\circ = \arctan \omega_{SL}/\Omega$.

2. During the spin-lock relaxation period, cross-relaxation caused by cross-correlation between dipole-dipole and chemical shift anisotropy relaxation mechanisms should be suppressed. The experiments in Fig. 2 achieve this suppression by applying ^1H 180° pulses (Fig. 2a, b), or a ^1H CW field (Fig. 2d).
3. Temperature control is an important component of any relaxation experiment, because conformational dynamics can be

Fig. 2 (continued) (2.0 ms, 40 G/cm), $g_d(xyz)$ (0.2 ms, 40 G/cm). **(b)** On-resonance $R_{1\rho}$ experiment. The delay $\chi = 1/\omega_{SL} - 4\tau_{90}/\pi$, where ω_{SL} is the spin-lock field strength and τ_{90} is the length of the 90° pulse. The delays τ are set to be just long enough to encompass the gradient, gradient recovery delay, and the 180° ^1H pulse. The gradients are sine shaped with durations and strengths as follows: $g_0(z)$ (2.0 ms, 9 G/cm), $g_1(z)$ (0.2 ms, 4 G/cm), g_2 (1.5 ms, 10 G/cm), $g_3(xyz)$ (1.0 ms, 25 G/cm), $g_4(xyz)$ (1.0 ms, 20 G/cm), $g_e(xyz)$ (2.0 ms, 40 G/cm), $g_d(xyz)$ (0.2 ms, 40 G/cm). **(c)** Off-resonance TROSY selected experiment. Delays are $\Delta_1 = 1.3$ ms and $\Delta_2 = 2.49$ ms. The hatched bar is a composite $90^\circ_y - 180^\circ_x - 90^\circ_y$ pulse. The phase cycle for Bruker spectrometers is φ_1 $4(x, -x)$, φ_2 $4(135^\circ)4(315^\circ)$, φ_3 $2(y, y, -y, -y)$, φ_4 $-y$, φ_5 $-x$, and φ_{rec} $(x, -x, -x, x, -x, x, x, -x)$; for Varian spectrometers, y -phases should be inverted. Gradient pulses are sine-shaped and have the following durations and amplitudes: $g_1(z)$ (0.5 ms, 6 G/cm), $g_2(z)$ (1.0 ms, 10 G/cm), $g_3(z)$ (0.9 ms, 28 G/cm), $g_4(xyz)$ (1 ms, 20 G/cm), $g_5(xyz)$ (0.182 ms, 28 G/cm). **(d)** Selective one-dimensional on- and off-resonance $R_{1\rho}$ experiment. The CW fields applied synchronously on ^1H and ^{15}N during the polarization transfer period have strength $\omega_{CP}/2\pi = 90$ Hz and duration $4\Delta = 1/J^{NH} \approx 10.8$ ms. Open bars are pulses with a tip angle equal to $\arctan(\omega_{SL}/\Omega)$ for ^{15}N or $\arctan(\omega_{1H}/\Omega_H)$ for ^1H , where ω_{SL} is the ^{15}N spin-lock field strength, Ω is the ^{15}N resonance offset from the spin lock carrier, ω_{1H} is the strength of the ^1H CW field applied during the spin-lock period, Ω_H is the ^1H resonance offset from the water frequency. The strength of the ^1H CW field ω_{1H} is $\omega_{1H}/2\pi \approx 3900$ Hz. Shaped ^1H pulses are 90° pulses selective for water (typically of ~ 1.5 ms duration), used for Watergate water suppression [57]. After the first water selective pulse, the ^1H carrier is moved from the water frequency to the amide resonance of interest; it is returned to the water frequency prior to the Watergate element. For all ^{15}N pulses, the ^{15}N carrier is placed on the amide resonance of interest with the exception of the ^{15}N spin-lock period, T , during which the ^{15}N carrier is placed on the desired off-resonance position, Ω . ^{15}N decoupling during acquisition is performed using a 1 kHz WALTZ16 sequence [58]. The period ε and the optional delay ξ are described in the text, Subheading 3.7. The phase cycle is as follows: φ_1 $8(y)$, $8(-y)$; φ_2 $-x, x$; φ_3 $4(x), 4(-x)$; φ_4 $2(x), 2(-x)$; φ_{rec} $x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, x, -x, -x, x$; $\varphi_3 = -y$ and $\varphi_4 = y$ for $\Omega < 0$; $\varphi_3 = y$ and $\varphi_4 = -y$ for $\Omega > 0$. Gradient pulses have the following durations and amplitudes: g_1 (1 ms, 3 G/cm), g_2 (50 μs , 15 G/cm), g_3 (500 μs , 20 G/cm)

greatly influenced by temperature [48]. Temperature control is particularly important in a $R_{1\rho}$ experiment, because the application of strong RF fields can cause significant sample heating. For this reason, the same total RF power should be applied to the sample for each recorded free-induction decay. This is achieved by applying a compensating spin-lock field between scans, at the beginning of the experiments during the recycle delay. The duration should be $T_{\text{comp}} = T_{\text{max}} - T$, where T is the particular spin-lock length during relaxation period, and T_{max} is the maximum over all spin-lock T values used to measure $R_{1\rho}$ (see Fig. 2). To compensate for the different spin-lock strengths ω_{SL} , T_{comp} should obey $T_{\text{comp}}\omega_{SL\text{max}}^2 + T\omega_{SL}^2 = \text{const}$ [13]. Critically, the compensating spin-lock field should occur during the recycle delay such that the recovery of proton polarization during that the inter-scan delay is constant. The compensating ^1H spin-lock field used in Fig. 2d must be applied far enough off-resonance to avoid perturbing ^1H magnetization.

3.6 Protocols for 2D $R_{1\rho}$ Measurements

2D $R_{1\rho}$ pulse schemes are presented in Fig. 2a–c and described in Subheading 3.5. These are challenging experiments because RF-heating from sustained pulsing (spin-locks, decoupling) may degrade the sample, form bubbles, or even damage the probe. Check with the probe manufacturer for the maximum probe power allowed. Exercise further caution when working with cryogenically cooled probes; sustained pulsing even at allowed powers may warm up the probe. Modern cryogenic probes can employ spin-lock field strengths high enough to investigate faster kinetic processes [49].

1. Pulse durations and amplitude should be set according to spectrometer specifications and as defined in the pulse sequences depicted in Fig. 2a–c. Delays should be set according to the pulse sequence, as indicated in Figs. 2a–c.
2. Set the recycle delay sufficiently long to ensure recovery of the ^1H magnetization, generally, 2–3 s.
3. For a concentrated sample start with eight scans; increase the number of scans to 16–32 for low concentration samples. The signal-to-noise ratio should be as high as possible (>40), as the quality of the fitted $R_{1\rho}$ rates heavily depends on the accuracy of the measured peak intensities.
4. When ready to collect the data, set number of “dummy scans” in the first experiment to be high, (256 of 512) to ensure proper equilibration of the temperature in the experiment.
5. Collect a first 1D experiment with the relaxation time, T , set to zero. Increase the value of the relaxation delay and adjust it until the intensities are approximately 30% of what measured with $T = 0$. This should be used as the maximum relaxation delay, T_{max} .

6. Set the relaxation delays to be used in the experiments, between 0 and T_{\max} . The best sampling of the decay curve is achieved when using relaxation times that correspond to an even decay of the intensities. These times points can be obtained using an average $R_{1\rho}$ rate, $R_{1\rho}^{av}$, obtained from T_{\max} as

$$I(t) = I_0 \exp(-R_{1\rho} t) \tag{10}$$

where I_0 is the intensity measured when $T = 0$. Thus,

$$R_{1\rho}^{av} = -\frac{\ln(0.3)}{T_{\max}}. \tag{11}$$

Six to eight time points are sufficient to generate good relaxation decay curves. For example, when using eight relaxation time points they can be calculated as

$$0, \quad -\frac{\ln 0.9}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.8}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.7}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.6}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.5}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.4}{R_{1\rho}^{av}}, \quad -\frac{\ln 0.3}{R_{1\rho}^{av}} = T_{\max}.$$

Generally at least two time points, first and last, are duplicated. As the decay of the magnetization is mono-exponential, only two relaxation times are strictly needed to derive $R_{1\rho}$ rates. Nevertheless, it is good practice to use more than two relaxation time points to build the relaxation decay curves to check for deviation from a single exponential curve, indicative of a problem with the experimental settings. Oscillations of the spin-locked magnetization, for example, might be symptomatic of poor alignment of the magnetization along the effective field.

To minimize the error due to environmental fluctuations such as temperature, magnetic field, etc., collect the data in an interleaved manner, i.e., collecting all the relaxation times for each time increment (t_1 , i.e., the indirect dimension) of the 2D spectrum, before moving onto the next t_1 increment. If the data are collected in the traditional non-interleaved manner (i.e., collect an entire 2D for one time point before moving onto the next one), avoid recording the 2Ds with a monotonically increasing T . Instead, use a shuffled order with several duplicates, such as $T_1, T_{\max}, T_2, T_7, T_3, T_6, T_4, T_5, T_1, T_{\max}, T_3$. The subscripts indicate the relative duration of the spin-lock period, $T_1 = 0 < T_2 < T_3 < \dots < T_{\max}$. Collecting the 2D spectra in this order allows the detection of systematic artifacts (i.e., bubbles formation, sample decay) that would otherwise go unnoticed, and produce erroneously high relaxation rates.

7. Process the time-domain data with the software of choice, to produce a 2D spectrum for each sampled relaxation delay T . For each spectrum, record the intensities or the volumes of all the peaks. This should produce for each peak a list of intensities (or volumes) versus relaxation delay T . Fit the intensities

(or volumes) to a single exponential decay to derive the $R_{1\rho}$ rate as described by Eq. (10).

8. To build the relaxation dispersion curve (Fig. 3), $R_{1\rho}$ rates are measured as a function of the effective field, ω_e , by varying the spin-lock field amplitude, ω_{SL} , and offset, Ω , *see* Eqs. (3)–(5). Repeat **steps 5–7** using different values of ω_{SL} and/or Ω . Use on-resonance experiments, Fig. 2b–c, with different values of ω_{SL} and off-resonance experiment, Fig. 2a and c, using different Ω and ω_{SL} . The larger the value of Ω , the larger the contribution of R_1 to $R_{1\rho}$ rates (Eq. (3)–(5)), thus Ω values probed in off-resonance experiments are rarely larger than three times the value of ω_{SL} , corresponding to tilt angles, θ , of $\sim 18^\circ$. The typical spin-lock field strengths ω_{SL} are as follows: 450–2000 Hz for the experiments in Fig. 2a and c; 150–2000 Hz in Fig. 2b, as discussed in Subheading 3.5. One must ensure that the highest field strength ω_{SL} stays within the recommended specifications of the probe. Generally, 15–30 $R_{1\rho}$ rates are determined as a function of the effective field, ω_e , by varying ω_{SL} and Ω .

3.7 Protocols for 1D $R_{1\rho}$ Measurements

The 1D $R_{1\rho}$ pulse sequence is shown in Fig. 2d and described in Subheading 3.5. As discussed in Subheading 3.6, extreme caution should be used when selecting the spin-lock field strength and the duration of the spin-lock period to ensure that they do not exceed the tolerance of the probe. Pulse durations and amplitudes and all delays should be as defined in Fig. 2d, and checked for their compliance with spectrometer and probe specifications. Additional key points are below.

1. The recycle delay should be long enough to ensure recovery of the ^1H magnetization, generally 2–3 s.
2. Set the number of scans to obtain a signal-to-noise ratio >40 .
3. Per Subheading 3.6, set the maximum duration of the relaxation time, T_{max} , such that the peak intensity decays to $\sim 30\%$ of that measured when the relaxation time $T = 0$. T_{max} may differ from resonance to resonance. To define an appropriate relaxation decay curve, select the 6–8 relaxation times to be between 0 and T_{max} as discussed in Subheading 3.6.
4. The strength of the $^1\text{H}/^{15}\text{N}$ RF fields, ω_{CB} , used in the selective Hartman-Hahn polarization transfer should be set to ~ 90 Hz, a value that has been shown to be optimal for polarization transfer [17, 42]. The power levels corresponding to these ^1H and ^{15}N pulses can be estimated from the ^1H 90° pulse and from the RF field strength calibration discussed in Subheading 3.3. The pulse duration is $4\Delta = 1/J \approx 10.8$ ms. The power level and the duration of the ^1H and ^{15}N pulses can be then fine-tuned to maximize the signal intensity of a sharp well-resolved resonance in a 1D experiment with a relaxation time $T = 0$.

5. The period ε allows for equilibration between the two exchanging states and should be $\sim 3/k_{ex}$ [17, 43].
6. The optional delay ξ uses as similar strategy as the delay χ in sequence Fig. 2b to eliminate unwanted resonances with overlapping ^1H resonances, and a similar ^{15}N resonance. Set $\xi = \pi/2\delta$, where δ is the ^{15}N offset (in rad/s) of the unwanted peak from the resonance of interest.
7. Collect on-resonance data by setting the offset from the carrier frequency, $\Omega = 0$. Measure ~ 20 values of the spin-lock field strength, $\omega_{SL}/2\pi$, covering a range from ~ 50 to 2000 Hz, making sure they comply with the probe specifications. For each ω_{SL} , measure $R_{1\rho}$ for a series of relaxation delays, T , (generally between 6 and 8, with at least two replicates, between 0 and T_{max}). Measure the intensity of the peak obtained for each T , and fit the intensities to a single exponential decay to extract the $R_{1\rho}$ rate of the resonance of interest, as discussed for the 2D experiments in Subheading 3.6.
8. Collect off-resonance data by varying both ω_{SL} and the offset from the carrier frequency, Ω . A set of 3–5 values of ω_{SL} is generally sufficient for off-resonance experiments. For each ω_{SL} , take measurements for ~ 20 –30 values of Ω , from $-\Omega_{max}$ to Ω_{max} . The Ω_{max} value is typically \sim three times the value of ω_{SL} because for $\Omega > \Omega_{max}$, $R_{1\rho}$ rates are dominated by R_1 (see Eqs. (3)–(5)).

3.8 Data Fitting

Collecting the $R_{1\rho}$ dispersion curve (shown in Fig. 3), using both on-resonance and off-resonance experiments with weak and strong RF fields, is the first step in the characterization of the exchange dynamics. Subsequent analysis of the $R_{1\rho}$ dispersion profile gives

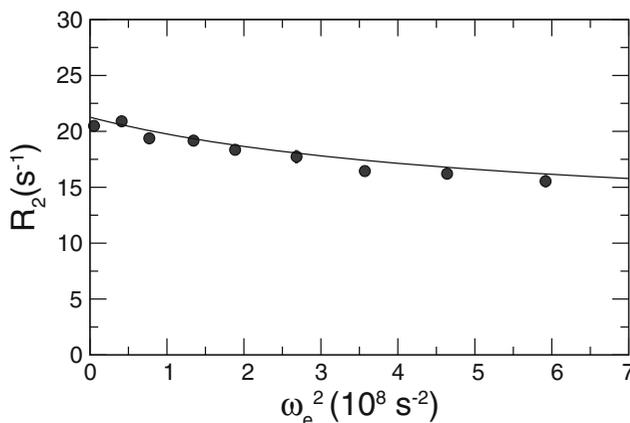


Fig. 3 Transverse relaxation rate constant, R_2 , is shown as a function of the effective field, ω_e , at 280 K for N25 of ubiquitin. The dispersion curve was collected at a static magnetic field of 14.1 T. The solid line shows the result of the fit, using Eq. (8), that yielded a value of $k_{ex} = (2.3 \pm 0.2) \times 10^4 \text{ s}^{-1}$ [16]

information about the exchange rate constant, k_{ex} [13]. The analysis involves fitting the dispersion profile to those predicted by model functions. For a two-state exchange process, Eqs. (3)–(8) describe the dependence of $R_{1\rho}$ from the effective field, ω_e . These expressions invoke different approximations and are valid under different conditions, described in Subheading 1.1. These functions contain kinetic and thermodynamic parameters of the exchange process, which are adjusted to best fit the $R_{1\rho}$ dispersion profile, as judged by a χ^2 metric [32].

To start the fit it will be necessary to make an initial estimate of the exchange parameters and of the particular equation that best describes the process, discussed in Subheading 1.1. The parameters are then optimized via χ^2 minimization driven by standard algorithms such as the Levenberg-Marquardt nonlinear least-squares algorithm [32]. The quality of the fit is assessed using a χ^2 analysis [32]. A poor fit of the data may indicate that the conditions of validity of the equation used in the fit are not satisfied, or that the dynamical process requires a more complex model than the 2-state chemical exchange equilibrium. Several expressions have been derived describing the dependence of $R_{1\rho}$ from ω_e for more complex exchange processes involving more than 2 states [13, 35, 50] that can be used to fit the data when the 2-state exchange model fails to represent the experimental data. Statistical tests (i.e., F-test [32], Akaike information criterion [51]) can be used for the selection of the most appropriate model describing the exchange process. Uncertainties in fitted parameters can be estimated by jackknife or bootstrap methods [52].

Acknowledgments

This work was supported by National Institutes of Health Grants GM098763 and GM117008 (F.M.) and by National Science Foundation Grant MCB-1616741 (J.W.P). We thank Arthur G. Palmer III (Columbia University) for making available the data presented in Fig. 3.

References

1. Berman HM (2008) The protein data Bank: a historical perspective. *Acta Crystallogr A* 64(Pt 1):88–95. doi:10.1107/S0108767307035623
2. Weber G (1972) Ligand binding and internal equilibria in proteins. *Biochemistry* 11 (5):864–878. doi:10.1021/bi00755a028
3. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975) Dynamics of ligand binding to myoglobin. *Biochemistry* 14(24):5355–5373
4. Wagner G, DeMarco A, Wuthrich K (1976) Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. ^1H NMR studies. *Biophys Struct Mech* 2(2):139–158
5. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267 (5612):585–590
6. Wand AJ (2013) The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* 23 (1):75–81. doi:10.1016/j.sbi.2012.11.005

7. Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13(6):748–757
8. Berendsen HJ, Hayward S (2000) Collective protein dynamics in relation to function. *Curr Opin Struct Biol* 10(2):165–169
9. Yon J, Perahia D, Ghelis C (1998) Conformational dynamics and enzyme activity. *Biochimie* 80(1):33–42
10. Palmer AG 3rd, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Meth Enzymol* 339:204–238
11. Kay LE (1998) Protein dynamics from NMR. *Biochem Cell Biol* 76(2–3):145–152
12. Palmer AG (1997) Probing molecular motion by NMR. *Curr Opin Struct Biol* 7(5):732–737
13. Palmer AG 3rd, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* 106(5):1700–1719. doi:[10.1021/cr0404287](https://doi.org/10.1021/cr0404287)
14. Xue Y, Kellogg D, Kimsey IJ, Sathyamoorthy B, Stein ZW, McBairty M, Al-Hashimi HM (2015) Characterizing RNA excited states using NMR relaxation dispersion. *Methods Enzymol* 558:39–73. doi:[10.1016/bs.mie.2015.02.002](https://doi.org/10.1016/bs.mie.2015.02.002)
15. Zhao B, Zhang Q (2015) Characterizing excited conformational states of RNA by NMR spectroscopy. *Curr Opin Struct Biol* 30:134–146. doi:[10.1016/j.sbi.2015.02.011](https://doi.org/10.1016/j.sbi.2015.02.011)
16. Massi F, Grey MJ, Palmer AG 3rd (2005) Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR $R_{1\rho}$ relaxation experiments. *Protein Sci* 14(3):735–742. doi: [10.1110/ps.041139505](https://doi.org/10.1110/ps.041139505)
17. Korzhnev DM, Orekhov VY, Kay LE (2005) Off-resonance $R_{1\rho}$ NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *J Am Chem Soc* 127(2):713–721. doi:[10.1021/ja0446855](https://doi.org/10.1021/ja0446855)
18. Trott O, Palmer AG (2002) $R_{1\rho}$ relaxation outside of the fast-exchange limit. *J Magn Reson* 154(1):157–160
19. Eichmüller C, Skrynnikov NR (2005) A new amide proton $R_{1\rho}$ experiment permits accurate characterization of microsecond time-scale conformational exchange. *J Biomol NMR* 32(4):281–293. doi: [10.1007/s10858-005-0658-y](https://doi.org/10.1007/s10858-005-0658-y)
20. Lundström P, Akke M (2005) Off-resonance rotating-frame amide proton spin relaxation experiments measuring microsecond chemical exchange in proteins. *J Biomol NMR* 32(2):163–173. doi:[10.1007/s10858-005-5027-3](https://doi.org/10.1007/s10858-005-5027-3)
21. Ishima R, Wingfield PT, Stahl SJ, Kaufman JD, Torchia DA (1998) Using amide ^1H and ^{15}N transverse relaxation to detect millisecond time-scale motions in perdeuterated proteins: application to HIV-1 protease. *J Am Chem Soc* 120(40):10534–10542
22. Mulder FAA, Akke M (2003) Carbonyl ^{13}C transverse relaxation measurements to sample protein backbone dynamics. *Magn Reson Chem* 41(10):853–865.
23. Lundström P, Akke M (2005) Microsecond protein dynamics measured by $^{13}\text{C}^\alpha$ rotating-frame spin relaxation. *ChemBioChem* 6(9):1685–1692. doi: [10.1002/cbic.200500086](https://doi.org/10.1002/cbic.200500086)
24. Brath U, Akke M, Yang D, Kay LE, Mulder FA (2006) Functional dynamics of human FKBP12 revealed by methyl ^{13}C rotating frame relaxation dispersion NMR spectroscopy. *J Am Chem Soc* 128(17):5718–5727. doi: [10.1021/ja0570279](https://doi.org/10.1021/ja0570279)
25. Weininger U, Blissing AT, Hennig J, Ahlner A, Liu Z, Vogel HJ, Akke M, Lundström P (2013) Protein conformational exchange measured by ^1H $R_{1\rho}$ relaxation dispersion of methyl groups. *J Biomol NMR* 57(1):47–55. doi: [10.1007/s10858-013-9764-4](https://doi.org/10.1007/s10858-013-9764-4)
26. Weininger U, Brath U, Modig K, Teilum K, Akke M (2014) Off-resonance rotating-frame relaxation dispersion experiment for ^{13}C in aromatic side chains using L-optimized TROSY-selection. *J Biomol NMR* 59(1):23–29. doi: [10.1007/s10858-014-9826-2](https://doi.org/10.1007/s10858-014-9826-2)
27. Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press
28. Desvaux H, Berthault P (1999) Study of dynamic processes in liquids using off-resonance rf irradiation. *Prog Nucl Magn Reson Spectrosc* 35(4):295–340
29. Miloushev VZ, Palmer AG 3rd (2005) $R_{1\rho}$ relaxation for two-site chemical exchange: general approximations and some exact solutions. *J Magn Reson* 177(2):221–227. doi: [10.1016/j.jmr.2005.07.023](https://doi.org/10.1016/j.jmr.2005.07.023)
30. Deverell C, Morgan RE, Strange JH (1970) Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Mol Phys* 18(4):553–559. doi:[10.1080/00268977000100611](https://doi.org/10.1080/00268977000100611)
31. Kovrigin EL, Kempf JG, Grey MJ, Loria JP (2006) Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J Magn Reson* 180(1):93–104. doi: [10.1016/j.jmr.2006.01.010](https://doi.org/10.1016/j.jmr.2006.01.010)
32. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes 3rd edition: the art of scientific computing. Cambridge University Press, Cambridge
33. Efron B (1982) The jackknife, the bootstrap, and other resampling plans. SIAM

34. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. CRC Press, Boca Raton, FL
35. Koss H, Rance M, Palmer AG 3rd (2017) General expressions for $R_{1\rho}$ relaxation for N-site chemical exchange and the special case of linear chains. *J Magn Reson* 274:36–45. doi: [10.1016/j.jmr.2016.10.015](https://doi.org/10.1016/j.jmr.2016.10.015)
36. Amman C, Meier P, Merbach AE (1982) A simple multinuclear NMR thermometer. *J Magn Reson* 46(2):319–321
37. Guennegues M, Berthault P, Desvaux H (1999) A method for determining B_1 field inhomogeneity. Are the biases assumed in heteronuclear relaxation experiments usually underestimated? *J Magn Reson* 136(1):118–126. doi: [10.1006/jmre.1998.1590](https://doi.org/10.1006/jmre.1998.1590)
38. Mulder FAA, de Graaf RA, Kaptein R, Boelens R (1998) An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J Magn Reson* 131(2):351–357
39. Massi F, Johnson E, Wang C, Rance M, Palmer AG 3rd (2004) NMR $R_{1\rho}$ rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* 126(7):2247–2256. doi: [10.1021/ja038721w](https://doi.org/10.1021/ja038721w)
40. Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* (23):12366–12371.
41. Igumenova TI, Palmer AG 3rd (2006) Off-resonance TROSY-selected $R_{1\rho}$ experiment with improved sensitivity for medium- and high-molecular-weight proteins. *J Am Chem Soc* 128(25):8110–8111. doi: [10.1021/ja061692f](https://doi.org/10.1021/ja061692f)
42. Pelupessy P, Chiarparin E, Bodenhausen G (1999) Excitation of selected proton signals in NMR of isotopically labeled macromolecules. *J Magn Reson* 138(1):178–181. doi: [10.1006/jmre.1999.1715](https://doi.org/10.1006/jmre.1999.1715)
43. Hansen AL, Nikolova EN, Casiano-Negroni A, Al-Hashimi HM (2009) Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon $R_{1\rho}$ NMR spectroscopy. *J Am Chem Soc* 131(11):3818–3819. doi: [10.1021/ja8091399](https://doi.org/10.1021/ja8091399)
44. Zinn-Justin S, Berthault P, Guennegues M, Desvaux H (1997) Off-resonance rf fields in heteronuclear NMR: application to the study of slow motions. *J Biomol NMR* 10(4):363–372. doi: [10.1023/A:1018365815186](https://doi.org/10.1023/A:1018365815186)
45. Kim S, Baum J (2004) An on/off resonance rotating frame relaxation experiment to monitor millisecond to microsecond timescale dynamics. *J Biomol NMR* 30(2):195–204
46. Yamazaki T, Muhandiram R, Kay LE (1994) NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly ^{13}C , ^{15}N -labeled proteins: application to $^{13}\text{C}^\alpha$ Carbons. *J Am Chem Soc* 116(18):8266–8278. doi: [10.1021/ja00097a037](https://doi.org/10.1021/ja00097a037)
47. Griesinger C, Ernst RR (1987) Frequency offset effects and their elimination in NMR rotating-frame cross-relaxation spectroscopy. *J Magn Reson* 75(2):261–271
48. Wang AC, Bax A (1993) Minimizing the effects of radio-frequency heating in multidimensional NMR experiments. *J Biomol NMR* 3(6):715–720
49. Ban D, Gossert AD, Giller K, Becker S, Griesinger C, Lee D (2012) Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J Magn Reson* 221:1–4. doi: [10.1016/j.jmr.2012.05.005](https://doi.org/10.1016/j.jmr.2012.05.005)
50. Trott O, Palmer AG (2004) Theoretical study of $R_{1\rho}$ rotating-frame and R_2 free-precession relaxation in the presence of n-site chemical exchange. *J Mag Reson* 170(1):104–112
51. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723
52. Caceci MS (1989) Estimating error limits in parametric curve fitting. *Anal Chem* 61(20)
53. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663–10665
54. Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected 2-dimensional Heteronuclear correlation Nmr-spectroscopy. *J Magn Reson* 93(1):151–170
55. Shaka AJ, Keeler J, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson* 64(3):547–552
56. Mulder F, RA dG, Kaptein R, Boelens R (1998) An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J Mag Reson* 131(2):351–357
57. Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2(6):661–665
58. Shaka AJ, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. *J Magn Reson* 53:313–340

Chapter 11

CPMG Experiments for Protein Minor Conformer Structure Determination

Anusha B. Gopalan, D. Flemming Hansen, and Pramodh Vallurupalli

Abstract

CPMG relaxation dispersion NMR experiments have emerged as a powerful method to characterize protein minor states that are in exchange with a visible dominant conformation, and have lifetimes between ~0.5 and 5 milliseconds (ms) and populations greater than 0.5%. The structure of the minor state can, in favorable cases, be determined from the parameters provided by the CPMG relaxation dispersion experiments. Here, we go through the intricacies of setting up these powerful CPMG experiments.

Key words NMR, Protein, Conformational dynamics, CPMG, Relaxation dispersion

1 Introduction

Under ambient conditions in aqueous solutions protein molecules interconvert between different conformational states that can have different populations and lifetimes [1–3]. Exchange is quite often between a dominant major state and other transiently populated minor states [4, 5]. The populations of the states depend on their relative free energies, while their lifetimes are largely dependent on the size of the free energy barrier that separates them from other states [6]. Peaks in NMR spectra arising from these minor conformers are usually not visible due to their low populations and short lifetimes and these minor conformers have to be detected by manipulating the visible major state magnetization [7]. Minor conformers that have lifetimes between 0.5 and 5 milliseconds (ms) can be detected using the Carr-Purcell-Meiboom-Gill (CPMG) type relaxation dispersion experiments. In the constant time (CT) version of the CPMG experiment, transverse magnetization is allowed to evolve for a fixed time T_{CPMG} , during which refocusing π (180°) pulses are applied [8]. The transverse magnetization can be of the single quantum or multi quantum variety and can originate from different nuclei like ^1H , ^{13}C , and ^{15}N [9, 10].

When there is chemical exchange in the millisecond timescale regime, the intensity of the visible state peak depends on the frequency ν_{CPMG} at which the π pulses are applied. Here, $\nu_{\text{CPMG}} = \frac{1}{4\tau_{\text{CPMG}}}$ where $2\tau_{\text{CPMG}}$ is the time between two π pulses. The effective relaxation rate $R_{2,\text{eff}}(\nu_{\text{CPMG}}) = -\frac{1}{T_{\text{CPMG}}} \ln\left(\frac{I(\nu_{\text{CPMG}})}{I_0}\right)$ depends on ν_{CPMG} , thereby giving rise to the relaxation dispersion curve. Here, $I(\nu_{\text{CPMG}})$ is the peak intensity at a given ν_{CPMG} value and I_0 is the peak intensity in the reference experiment recorded in the absence of the T_{CPMG} delay. The minor state population, lifetime, and chemical shift can be obtained by analyzing the relaxation dispersion curve(s) [7, 11].

In this chapter, we describe how to set up NMR CPMG experiments that probe the dynamics at different protein backbone sites. We also provide a very limited overview on how to analyze the data.

2 Materials

2.1 NMR Samples

Protein samples with different isotope labeling schemes are required to study the dynamics at various protein backbone positions [12]. These include (1) a uniformly $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ enriched sample. (2) A uniformly $^{15}\text{N}/^{13}\text{C}/50\%^2\text{H}$ enriched sample. (3) A $^{15}\text{N}/^{13}\text{C}\alpha$ labeled sample. Sample (1) is used to study dynamics at the amide (^1H , ^{15}N) and carbonyl (^{13}C) positions. Sample (2) is used to study dynamics at $\text{H}\alpha$ position. Sample (3) is used to study dynamics at the $\text{C}\alpha$ position. Although not necessary it is preferable to use a uniformly $^{15}\text{N}/^2\text{H}$ enriched sample for amide $^{15}\text{N}/^1\text{H}$ single quantum (SQ), zero quantum (ZQ), and double quantum (DQ) CPMG experiments. Preliminary experiments to standardize buffer conditions, temperature, etc. can be performed using just a ^{15}N -enriched sample. The concentrations of protein samples used in CPMG studies are usually between 1 and 2 millimolar (mM). Experiments can also be performed with 0.5 mM samples when the spectrometer is equipped with a cryogenically cooled probe.

2.2 NMR Spectrometers

The experiments are usually performed on spectrometers equipped with triple resonance probes. The minimum magnetic field strength used is 11.7 T (500 MHz ^1H resonance frequency). Experiments are usually performed at two well-separated field strengths like 11.7 T and 16.5 T/18.8 T (500 and 700/800 MHz).

2.3 Thermocouple for Temperature Measurement

A thermocouple with a long (~ 3 m) wire connecting it to a digital display is inserted into a standard NMR sample tube. The thermocouple is positioned where the sample center would be and the wire coming out of the top of the tube is attached to the tube using Parafilm to hold the thermocouple in place. The tube can now be inserted into a spinner assembly and loaded into the magnet in the usual manner to measure the sample temperature.

3 Methods

3.1 *Sample Temperature Measurement and Setting*

Temperature affects the kinetics of the processes being studied [6] and has to be kept constant during all the experiments. To keep the sample temperature constant between different spectrometers, it is important to measure the sample temperature in each of them for a given variable temperature (VT) setting.

1. Measure the sample temperature for a given VT setting. The sample temperature can be easily measured using a thermocouple inserted into a standard sample tube as described in Sub-heading 2.2 or using temperature standards like methanol (*see Note 1*).
2. Adjust the setting of the VT unit till the desired temperature is obtained.

3.2 *Setting the ^1H Carrier Positions*

In most of the experiments described here the ^1H radio-frequency (RF) carrier is centered on the water signal. In water samples the water resonance can be quite broad (~ 100 Hz) on a typical triple resonance probe and the peak center cannot be determined accurately. A simple presaturation procedure can be used to obtain the water resonance frequency.

1. Obtain an estimate of the water resonance frequency from a one-scan 1D spectrum obtained using a small flip angle (~ 10 – 15°) pulse applied at the highest allowed power (usually 25–50 kHz).
2. Array the ^1H carrier frequency in 1 Hz steps around the estimated offset from **step 1** and record a series of 1D experiments with a weak 2 s presaturation applied at a very low B_1 field of ~ 100 Hz followed by the small flip angle pulse applied at full power. Process each recorded FID in an absolute value (magnitude) mode. The spectrum with the lowest intensity water peak is obtained when the carrier is at the water resonance frequency.

3.3 *Calibrating Pulsewidths and Powers*

Pulses have to be calibrated accurately to suppress artifacts and maximize the signal-to-noise ratio (S/N) of the experiments. All the pulses should be calibrated every time experiments are set up, including the pulses for ^{13}C and ^{15}N .

1. *^1H square pulses*: Both hard and decoupling ^1H square pulses are calibrated using the water signal.
 - (a) Set the ^1H RF carrier on the water resonance frequency.
 - (b) Record a series of single-scan 1D experiments at the desired B_1 field strength varying the pulsewidth in steps of 1 μs around the pulsewidth for a 360° (2π) pulse. Process the FIDs in absolute value mode. The spectrum

with the lowest water peak intensity is the one where the pulsewidth is closest to the 2π pulse. Subsequently, array the pulsewidth in smaller steps around this value to obtain a more accurate estimate of the pulsewidth. Typical array spacing for the ^1H hard pulses is $0.1\ \mu\text{s}$ but can be larger for longer square pulses like decoupling pulses. The ^1H hard pulses are applied at the highest possible power and typical pulsewidths for hard 90° pulses are between 6.5 and $9\ \mu\text{s}$. CPMG pulses are applied at a lower power to avoid damaging the probe and pulsewidths for the refocusing 180° pulses vary between 21 and $22\ \mu\text{s}$.

2. *^1H selective pulses for water suppression*: Due to radiation damping 90° water selective pulses that bring water magnetization down from the Z axis to the transverse plane are longer than those that take water from the transverse plane to the Z axis [7]. The water selective pulses are calibrated using a two-pulse sequence where the first pulse and the second pulse are 180° out of phase. To calibrate the pulse that brings the magnetization down to the transverse plane from the Z axis, the first pulse is the water selective pulse that has to be calibrated while the second one is a hard 90° proton pulse. To calibrate the pulse that takes water magnetization from the transverse plane to the Z axis, the order of the hard and selective pulses is reversed.
 - (a) Place the ^1H RF carrier on the water resonance and set the hard pulse power and pulsewidth to that of a 90° pulse.
 - (b) Select the shape and set the power for the water selective pulse. Typical shapes include a square pulse, a SEDUCE pulse [13], and the EBURP pulse [14].
 - (c) Record a series of single-scan 1D experiments varying the selective pulsewidth around the expected pulsewidth. The step size can be $10\text{--}20\ \mu\text{s}$ for 1.5 to 2 ms selective pulses and $50\text{--}100\ \mu\text{s}$ for 5 to 10 ms selective pulses. Process the FIDs in the absolute value mode. The minimum water peak intensity is obtained when the selective pulsewidth is equal to the length of the 90° pulse. Alternatively, one can array the power of the selective pulse while keeping the pulsewidth fixed. At 500 MHz the 90° square water selective pulses are ~ 1.5 ms long, SEDUCE pulses are ~ 1.8 ms and the EBURP pulse is ~ 7 ms. These pulses are usually not scaled with field strength.
 - (d) Set the selective pulsewidth and power to the optimized values and array the small angle phase of the selective pulse around 0° in steps of 1° . Process the spectra as above to obtain the phase at which the water intensity is the lowest. For newer spectrometers it is very close to 0° but can be larger for older spectrometers mainly due to nonlinearity of the amplifiers and pulse imperfections.

3. *¹⁵N square pulse:* ¹⁵N square pulses are calibrated using a modified HSQC experiment that contains an extra ¹⁵N pulse that is to be calibrated placed between two Z gradients. This calibration element is placed after the first ¹H → ¹⁵N INEPT transfer. The first INEPT period [15] generates a density matrix element proportional to the 2H_zN_z longitudinal two-spin order magnetization. Applying an extra ¹⁵N (or ¹H) pulse followed by a strong Z gradient will destroy all the magnetization and no signal will be observed when the pulsewidth corresponds to 90° pulse. It is best to use a sensitivity enhanced HSQC with gradient selection [16] for this purpose. A ¹⁵N labeled protein sample in H₂O is required to perform this calibration.
 - (a) Place the ¹H RF carrier on the water resonance and the ¹⁵N carrier at 119 ppm, which is in the middle of the amide region.
 - (b) Set powers and pulsewidths of the ¹H hard and water selective pulses. Set the small angle phase shift of the water selective pulses.
 - (c) Set the ¹⁵N hard and decoupling pulsewidths and powers to reasonable values. Set the shape and the power of the pulse to be calibrated to the desired value. These can be taken from the last calibration performed on the sample or from calibrations performed on a standard sample.
 - (d) Set the pulsewidth of the ¹⁵N pulse to be calibrated to 0 and record the first increment of the HSQC experiment (no t₁ delay) with 8 scans. Process the FID with water suppression and suitable apodization and phase the spectrum focusing on the amide ¹H region. Ensure that water suppression is adequate and that the S/N of the amide protons is adequate. If not one can increase the number of scans. An acquisition time of 64 ms and an interscan delay of 1.5–2 s are usually sufficient.
 - (e) Record a series of 1D spectra as above varying the pulsewidth of the pulse to be calibrated. The pulsewidth is varied in steps of 0.33–0.5 μs for hard and CPMG pulses that are ~35–45 μs long. Process the spectrum exactly as in the previous step. The amide peaks will be positive when the pulsewidth is less than the 90° pulsewidth and negative when the pulsewidth is greater than the 90° pulsewidth. Almost no signal will be seen when the pulsewidth is almost 90°. Hard ¹⁵N 90° pulses are ~35 μs long. To avoid damaging the probe CPMG pulses are applied at a lower power with 180° pulsewidths ranging between 80 and 90 μs.
4. *¹⁵N REBURP pulse:* The refocusing REBURP pulse [14] can be calibrated in a manner similar to the hard pulse because when the pulse is applied for half its duration it behaves like a 90° pulse with respect to starting Z magnetization.

- (a) Set the shape of the pulse to be calibrated to a REBURP pulse and the power to the desired value.
 - (b) Record a series of 1D spectra using the first increment of the HSQC experiment as in **step 3** varying the width of the REBURP pulse. The pulsewidth is varied in steps of 1–2 μs around the estimated 90° value. For example if the desired pulsewidth of the REBURP pulse is 1.5 ms, then the pulsewidth is arrayed in steps of $\sim 2 \mu\text{s}$ around 0.75 ms. Process and compare the series of 1D spectra as in **step 3** to obtain the 90° pulsewidth. Doubling this value will give the pulsewidth of the REBURP.
 - (c) Confirm that the REBURP pulse is accurately calibrated. Set the pulsewidth of the REBURP to the calibrated value. Record two spectra, one with the REBURP pulse power set to 0 and one with the REBURP pulse power set to the desired value. When the calibration is accurate the amide region will be inverted in the second spectrum, with the height of each amide peak in the second spectrum being at least 95% of its height in the first spectrum. The amide selective ^{15}N REBURP pulsewidth is typically ~ 1.4 ms (at 16.65 T) and needs to be scaled according to field strength.
5. *^1H REBURP pulse:* The ^1H REBURP pulse can be calibrated in a manner similar to the ^{15}N REBURP pulse. The ^{15}N REBURP pulse is replaced by the ^1H REBURP pulse in the above sequence and the process is repeated with two small differences. (1) During the ^1H REBURP pulse the ^1H carrier is placed at ~ 7.7 ppm in the middle of the amide proton region. (2) A water selective pulse that is usually applied after the ^1H to ^{15}N INEPT period is turned off. The amide selective ^1H REBURP pulse is generally chosen to be ~ 1.4 ms long (at 16.65 T) and needs to be scaled according to the field strength.
 6. *^{13}C square/shaped pulses via HNCOC experiments:* In uniformly $^{15}\text{N}/^{13}\text{C}$ enriched protein samples dissolved in water the ^{13}C pulses are calibrated in a manner similar to the ^{15}N pulses. Here, the first scan of a gradient selected sensitivity enhanced HNCOC experiment [17] is used instead of the HSQC experiment. A Z gradient followed by the extra ^{13}C pulse to be calibrated followed by a strong Z gradient is inserted after the $^{15}\text{N} \rightarrow ^{13}\text{CO}$ INEPT period and calibrated in a manner identical to the ^{15}N pulses as described above.
 7. *^{13}C square/shaped pulse via HSQC experiments:* ^{13}C pulses can be calibrated using ^{13}C labeled methyl groups in samples which are not dissolved in water or when the carbonyl groups are not ^{13}C enriched. As in the case of the ^{15}N pulse calibration an extra ^{13}C pulse to be calibrated followed by a strong gradient is inserted after the $^1\text{H} \rightarrow ^{13}\text{C}$ INEPT in a standard ^1H - ^{13}C

HSQC experiment. The procedure for calibration is similar to that for the ^{15}N pulse. The ^{13}C carrier is placed in the middle of the methyl region (~ 20.5 ppm) and the pulsewidth is calibrated by looking at methyl proton peak intensities between -0.5 and 1 ppm just as one focused on the amide protons during the ^{15}N calibration as described above.

3.4 Measuring Sample Heating

Experiments with ^1H decoupling can heat the samples and consequently change the kinetics of the process being studied. Hence, the effect of the experiment on the sample temperature has to be measured and compensated by adjusting the VT setting (*see Note 2*).

1. Record a high-resolution 1D WATERGATE spectrum [7, 18] of the amide or methyl protons at 1°C higher than the desired temperature. Methyl protons are preferred but amide protons are used when methyl protons are absent due to ^2H labeling.
2. Record a 1D WATERGATE spectrum of the amide or methyl protons at the desired temperature. The peaks in this spectrum will move compared to the one recorded in the previous step allowing one to calibrate peak position change as a function of temperature.
3. Using the same VT temperature, perform the CPMG experiment of interest for about 5–10 min so that the sample temperature has reached a steady state.
4. Queue a single-scan 1D WATERGATE experiment of the amide/methyl region to run immediately after the CPMG experiment from **step 3** has finished recording. This experiment should have no dummy scans.
5. The peak positions would have changed compared to the spectrum recorded in **step 2** due to sample heating during the CPMG experiment. By comparing the peak positions in the three spectra recorded in **steps 1, 2, and 4** the heating caused by the CPMG experiment can be estimated according to $\frac{\omega_4 - \omega_2}{\omega_1 - \omega_2}$, where ω_1 , ω_2 , and ω_4 are the resonance frequencies of a particular amide/methyl proton in spectra recorded in **steps 1, 2, and 4** respectively. Do this for 3–5 peaks that move consistently and take the average value.
6. The VT value can now be reduced to compensate for the heating.

3.5 CT ^{15}N TROSY CPMG Experiment

The CT ^{15}N TROSY CPMG experiment exploits both the improved relaxation properties of the ^{15}N TROSY component during the T_{CPMG} period and the enhanced resolution of the ^{15}N - ^1H TROSY correlation map. This experiment is necessary to perform ^{15}N CPMG experiments with larger proteins but can also be used with smaller proteins. This experiment has to be performed with uniformly ^{15}N enriched samples. Uniform ^2H labeling will

improve the performance of the experiment but is not required. If the sample is ^{13}C enriched a single carbon pulse is used to refocus the ^{13}C - ^{15}N scalar couplings during the ^{15}N chemical shift evolution.

1. Place the ^1H RF carrier on the water resonance, the ^{15}N carrier at 119 ppm in the middle of the amide region, and the ^{13}C carrier at 118 ppm in between ^{13}CO and $^{13}\text{C}\alpha$ resonances. Values of the delays are available in [19].
2. Set powers and pulsewidths of the ^1H hard and water selective 90° pulses.
3. Set powers and pulsewidths of the ^{15}N 90° hard pulses and CPMG π pulses (*see Note 3*). To avoid damage to the probe, CPMG pulses are applied at lower power with the 180° pulsewidth ranging from 80 to 90 μs . A REBURP π pulse can be applied during the first INEPT period to suppress resonances arising from lysine and arginine sidechains.
4. If the sample is ^{13}C enriched set the power for a ~ 400 μs adiabatic pulse [20] with adequate bandwidth to invert both the carbonyls and alpha carbons.
5. Set the T_{CPMG} delay (*see Note 4*). T_{CPMG} values usually lie between 20 and 40 ms. Values above 40 ms are not chosen as applying a large number of ^{15}N refocusing pulses can damage the probe.
6. Set the number of scans in multiples of the phase cycle to obtain good signal to noise. Typical values are 8 or 16 scans.
7. Decide on the ν_{CPMG} values to use. ν_{CPMG} values that can be used are $\frac{n}{T_{\text{CPMG}}}$ where n is an integer starting from 1 [19, 21]. ν_{CPMG} values up to ~ 1000 Hz can be used with current probes. Experiments are performed for ~ 15 – 20 ν_{CPMG} values ranging from $\frac{1}{T_{\text{CPMG}}}$ to 1000 Hz with two or three repeat values for error estimation. Lower ν_{CPMG} values are sampled more closely.
8. Record ^1H - ^{15}N correlation maps for each of the ν_{CPMG} values and a reference plane without the T_{CPMG} delay in an interleaved manner. The usual experimental time varies from 12 to 24 h (*see Note 5*).

3.6 $CT^{15}\text{N}$ CPMG Experiment with ^1H Decoupling

Here, the amide proton is decoupled from the amide ^{15}N nucleus using ^1H CW decoupling during the T_{CPMG} delay [21]. This experiment is particularly useful while performing CPMG experiments in alignment media and is also significantly less prone to off-resonance artifacts. If the sample is ^{13}C enriched, a single-carbon pulse is used to refocus the ^{13}C - ^{15}N couplings during ^{15}N evolution.

1. Set the ^1H , ^{15}N , and ^{13}C carrier offsets as described in Subheading 3.5, **step 1**. The ^1H carrier is moved to 8.1 ppm in the middle of the amide region when CW ^1H decoupling is performed. Values of delays are available in [21].
2. Set powers and pulsewidths of the ^1H hard, water selective 90° pulses and CW decoupling field. An EBURP1 [21] pulse is used to bring the water magnetization from the Z axis to the transverse plane. The CW ^1H decoupling is applied at 15 kHz at 18.8 T [21, 22] and scaled according to the field strength.
3. Set powers and pulsewidths for the ^{15}N pulses as in **step 3** of Subheading 3.5. Additionally set the power for ^{15}N decoupling during acquisition. At 18.8 T WALTZ-16 ^{15}N decoupling is performed with B_1 of 1.3 kHz [23]. This ^{15}N decoupling field should be scaled with magnetic field strength.
4. If the sample is ^{13}C enriched use an adiabatic pulse for decoupling as described in Subheading 3.5, **step 4**.
5. Set the T_{CPMG} delay as described in Subheading 3.5, **step 5**. T_{CPMG} values usually lie between 20 and 30 ms. Due to heating effects of the ^1H CW field it is prudent to set $T_{\text{CPMG}} \leq 30$ ms.
6. Measure the heating due to the ^1H CW field and compensate for it by adjusting the VT setting. Alternatively *see* **Note 2**.
7. Set the number of scans, choose the ν_{CPMG} values and run the experiment as described in **steps 6, 7, and 8** of Subheading 3.5.

3.7 Amide ^1H CPMG Experiment

The amide ^1H CPMG experiment is used to study dynamics at the amide proton sites. A uniformly ^{15}N labeled protein sample is used. ^2H enrichment is desirable but not necessary. To suppress peaks arising from ROE transfer the experiment is performed on $2\text{H}_x\text{N}_z$ antiphase magnetization. If the sample is ^{13}C enriched, a single-carbon pulse is used to refocus the ^{13}C - ^{15}N couplings during ^{15}N evolution. Details can be found in [24, 25].

1. Set the ^1H , ^{15}N , and ^{13}C carrier offsets as described in Subheading 3.5, **step 1**. The ^1H carrier is moved to 8.1 ppm in the middle of the amide region during the CPMG period. Values of delays are available in [24, 25]. The modifications described in [25] have to be used if the procedure described here is to be used.
2. Set powers and pulsewidths of the ^1H hard, CPMG and water selective 90° pulses. To avoid damaging the probe the ^1H CPMG pulse is applied at a lower power with a 180° pulsewidth of $\sim 21 \mu\text{s}$. A REBURP pulse is used to invert the amide protons without perturbing the α protons in the middle of the CPMG period. The length of the REBURP and the amide carrier frequency need to be optimized to cover as many if not all the amides. A square 180° pulse applied at the same

power as the other CPMG pulses can be used if the sample is labeled with ^2H at the H^α positions.

3. Set powers and pulsewidths of the ^{15}N pulses as described in Subheading 3.6, **step 3**.
4. If the sample is ^{13}C enriched use an adiabatic pulse for decoupling as described in Subheading 3.5, **step 4**.
5. Set the T_{CPMG} delay and number of scans as described in Subheading 3.5, **steps 5** and **6**. T_{CPMG} values usually lie between 20 and 30 ms.
6. Decide on the ν_{CPMG} values to use and run the experiment as described in **steps 7** and **8** of Subheading 3.5. Unlike ^{15}N CPMG experiments ν_{CPMG} values up to 2000 Hz can be used with current probes. Experiments are performed for ~ 15 – 25 ν_{CPMG} values ranging from $\frac{1}{T_{\text{CPMG}}}$ to 2000 Hz with two or three repeat values for error estimation.

3.8 ^{13}CO CPMG Experiment

The carbonyl CPMG experiment is performed using a uniformly $^{15}\text{N}/^{13}\text{C}$ enriched sample.

1. Place the ^1H carrier on the water resonance, the ^{13}C carrier at 176 ppm in the middle of the carbonyl region, and the ^{15}N carrier at 119 ppm in the middle of the amide region. Values of delays are available in [26].
2. Set powers and pulsewidths of the ^1H hard and water selective 90° pulses as described in **step 2** of Subheading 3.5. Additionally, here WALTZ-16 decoupling is carried out and is performed at 6 kHz at 18.8 T.
3. Set powers and pulsewidths of the ^{15}N 90° hard pulses as in **step 3** of Subheading 3.5 and power for the ^{15}N decoupling sequence as in **step 3** of Subheading 3.6.
4. Square ^{13}C 90° pulses are applied at a field strength (given in Hz) of $\frac{\Delta\Omega}{\sqrt{15}}$ where $\Delta\Omega$ is 118 ppm \times dfrq where dfrq is the carbon frequency in MHz. The square ^{13}C 180° pulse is applied at the highest possible power leading to a pulsewidth of ~ 25 μs (20 kHz) for the 180° pulse. The SEDUCE-1 sequence [13] is used to decouple the $^{13}\text{C}\alpha$ from ^{13}CO during the t_1 ^{13}CO chemical shift evolution period. The SEDUCE-1 pulses are cosine modulated at a frequency of 118 ppm with a maximum B_1 of 3.14 kHz at 11.7 T. The length of the SEDUCE-1 90° pulse is ~ 335 μs at 11.7 T and is scaled according to the field strength. Note that due to the cosine modulation decoupling is applied at four times the power. During acquisition the ^1HN - ^{13}CO J coupling is decoupled using the WURST-2 decoupling sequence with a bandwidth of 12 ppm [20]. A maximum B_1 of 0.6 kHz is required at 18.85 T and should be scaled with field strength. A shaped pulse [26]

related to the REBURP pulse is used during the ^{13}C CPMG period. The length of this 180° pulse is $380\ \mu\text{s}$ at 18.85 T and should be scaled according to the magnetic field strength. This pulse is calibrated in a manner similar to the REBURP.

5. Set the T_{CPMG} delay and number of scans according to **steps 5** and **6** of Subheading 3.5. T_{CPMG} values usually lie between 20 and 30 ms.
6. Decide on the ν_{CPMG} values to use as described in **step 7** of Subheading 3.5 keeping in mind that although ν_{CPMG} values up to 1000 Hz can be used with current probes, at lower magnetic field strengths the length of the shaped ^{13}C CPMG pulse places an upper limit on the highest ν_{CPMG} value that can be used. Experiments are performed for ~ 15 ν_{CPMG} values.

3.9 $^1\text{H}_\alpha$ CPMG Experiment

The $^1\text{H}_\alpha$ chemical shift provides useful structural information and is extremely useful for the structure determination of the minor state. However, the H_α nucleus is scalar coupled to amide protons (HN) and H_β nuclei. The scalar coupling to the amide protons can be eliminated by dissolving the sample in D_2O , so that the amide proton is replaced by a ^2H nucleus. However, there is no straightforward way to produce protein molecules that are H_α labeled while being $\sim 100\%$ ^2H enriched at the H_β position. By overexpressing the protein in *E. coli* grown in 50% D_2O minimal media with $^2\text{H}/^{13}\text{C}$ glucose as the sole carbon source, the H_β is 50–88% deuterated reducing the effect of $^3\text{J}_{\text{H}_\alpha\text{H}_\beta}$ coupling. The $^1\text{H}_\alpha$ CPMG experiments consist of a J refocusing element in the middle of the CPMG period [27]. Neglecting relaxation, this J refocusing element uses the $^1\text{J}_{\text{C}_\alpha\text{H}_\alpha}$ coupling to refocus the effects of the $^3\text{J}_{\text{H}_\alpha\text{H}_\beta}$ coupling to first order. In this experiment dispersion profiles are obtained by analyzing the ^1H - $^{13}\text{C}_\alpha$ correlation maps.

1. Place the ^1H RF carrier on the water resonance, the ^{13}C carrier at 58 ppm in the C_α region, and the ^{15}N carrier at 119 ppm in the amide region. Values of delays are available in [27].
2. Set powers and pulsewidths of the ^1H hard and CPMG pulses. CPMG pulses are applied at a lower power as described in Subheading 3.7, **step 2**.
3. Set the power and pulsewidth for the ^{15}N hard pulse if the sample is ^{15}N enriched. As there is no water in the sample the ^{15}N pulsewidth and power have to be set to the last values calibrated on the same sample.
4. Set the power and pulsewidth for the ^{13}C hard pulse and decoupling sequences. During acquisition ^{13}C decoupling is performed using a 2.5 kHz WALTZ-16 decoupling scheme at 18.8 T. The decoupling power has to be scaled with field strength. Carbonyl and $^{13}\text{C}_\beta$ are decoupled from $^{13}\text{C}_\alpha$ during t_1 evolution. ^{13}CO decoupling is performed using a

CAWURST-8 scheme [28, 29] with 10 ppm bandwidth centered at 175 ppm and applied with a maximum field strength of 0.4 kHz at 18.8 T. $^{13}\text{C}\beta$ decoupling is performed using (1) A CAWURST-8 scheme with a 30 ppm bandwidth swept from 15 to 45 ppm centered at 30 ppm with a maximum B_1 of 0.5 kHz at 18.8 T and (2) A second CAWURST-8 scheme with a 30 ppm bandwidth swept from 101 to 71 ppm centered at 86 ppm with a maximum B_1 of 0.5 kHz at 18.8 T. The second field eliminates decoupling artifacts. During the J refocusing element C β decoupling is performed using a WURST-2 scheme [30] centered at 27.5 ppm with a 25 ppm bandwidth and a maximum B_1 of 1.22 kHz at 18.8 T.

5. Set the T_{CPMG} delay (*see Note 4*). T_{CPMG} values can be quite low due to the short $\text{H}\alpha$ T_2 values and usually lie between 15 and 30 ms.
6. Set the number of scans in multiples of the phase cycle to obtain good signal to noise. The sensitivity of these experiments is low due to 50% ^1H labeling at the α position and because the experiments are not carried out in a sensitivity-enhanced manner. Hence, a significantly larger number of scans is required with typical values ranging from 24 to 48 scans.
7. Decide on the ν_{CPMG} values to use. ν_{CPMG} values that can be used are $\frac{2n}{T_{\text{CPMG}}}$ where n is an integer starting from 1. ν_{CPMG} values up to 1000 Hz can be used. Experiments are performed for ~ 10 – 15 ν_{CPMG} values ranging from $\frac{2}{T_{\text{CPMG}}}$ to 1000 Hz with two or three repeat values for error estimation. Lower ν_{CPMG} values are sampled more closely.
8. Record ^1H - ^{13}C correlation maps for each of the ν_{CPMG} values and a reference plane without the T_{CPMG} delay in an interleaved manner. The usual experimental time varies from 24 to 48 h (*see Note 5*).

3.10 $^{13}\text{C}\alpha$ CPMG Experiment

The $^{13}\text{C}\alpha$ chemical shift provides very valuable structural information. As the $\text{C}\alpha$ chemical shifts are not well separated from the $\text{C}\beta$ chemical shifts, a sample that is ^{13}C enriched at only the $\text{C}\alpha$ sites is used. In this experiment, dispersion profiles are obtained by analyzing the ^1H - $^{13}\text{C}\alpha$ correlation maps.

1. Place the ^1H RF carrier on the water resonance, the ^{13}C carrier at 55 ppm in the $\text{C}\alpha$ region, and the ^{15}N carrier at 119 ppm in the middle of the amide region. Values of delays are available in [11].
2. Set powers and pulsewidths of the ^1H hard pulses. The ^1H CW decoupling field during the CPMG period is applied at a B_1 field strength of ~ 14.5 kHz. Purge trim CW pulses to dephase water are applied at 16 kHz. Proton to carbon and carbon to

proton transfers are carried out using DIPSI-2 sequences with ^1H and ^{13}C fields of 8 kHz at 18.8 T. ^1H decoupling during ^{13}C t_1 evolution is performed using a 5 kHz (18.8 T) WALTZ-16 sequence.

3. Set the power and pulsewidth for the ^{15}N hard pulse.
4. Set power and pulsewidth for the ^{13}C hard pulse and CPMG pulses. The CPMG pulses are applied at lower field strength with a 180° pulsewidth of $\sim 31 \mu\text{s}$. ^{13}C decoupling during acquisitions is performed using a 2.5 kHz (18.8 T) WALTZ-16 sequence. The decoupling field strength is scaled with the spectrometer frequency. Proton to carbon and carbon to proton transfers are carried out using the DIPSI-2 sequence [31] with proton and carbon fields of 8 kHz at 18.8 T. The decoupling field strength is scaled with magnetic field strength.
5. Set the T_{CPMG} delay (*see Note 4*). T_{CPMG} values can be quite low due to the short $\text{C}\alpha$ T_2 values and usually lie between 15 and 30 ms. Further due to heating effects of the ^1H CW field it is prudent to set $T_{\text{CPMG}} \leq 30$ ms.
6. Measure the heating due to the ^1H CW field and compensate for it by reducing the VT setting. Alternatively *see Note 2*.
7. Set the number of scans in multiples of the phase cycle to obtain good S/N ratio. Due to the method of sample production the $^{13}\text{C}\alpha$ enrichment is at most 50%. Further the experiment is not carried out in a sensitivity-enhanced manner. Hence, a significantly larger number of scans are required with typical values ranging between 16 and 48 scans.
8. Decide on the ν_{CPMG} values as described in **step 7** of Subheading 3.5. Note that although ν_{CPMG} values up to 1500 Hz can be used with current probes values up to 1000 Hz are preferred.
9. Run the experiment as described in **step 8** of Subheading 3.9.

3.11 Amide ^1H - ^{15}N ZQ/DQ CPMG Experiment

In addition to ^1H and ^{15}N single quantum transitions one can also excite zero quantum (ZQ) and double quantum (DQ) transitions in the coupled amide $^1\text{H}/^{15}\text{N}$ spin system. Relaxation dispersion curves for the ZQ and DQ transitions can also be recorded using CPMG experiments [32]. During ZQ/DQ CPMG experiments, ZQ/DQ coherence is generated and refocusing pulses are applied simultaneously on both the ^{15}N and ^1H nuclei during the T_{CPMG} period. In the case of exchange the change in the frequency of the ZQ transition between the minor and major states in ppm units is $|\Delta\varpi_{\text{N}} - \Delta\varpi_{\text{H}}|$ while the change in the frequency of the DQ transition is $|\Delta\varpi_{\text{N}} + \Delta\varpi_{\text{H}}|$. Hence if $|\Delta\varpi_{\text{N}}| > 0$ and $|\Delta\varpi_{\text{H}}| > 0$ then either the ZQ or the DQ transitions will give larger dispersions than the SQ experiments making them more sensitive to the underlying exchange process in CPMG experiments. Additionally, ZQ/

DQ CPMG experiments also provide information about the relative signs of $\Delta\varpi_{\text{N}}$ and $\Delta\varpi_{\text{H}}$. If the size of the DQ dispersion is larger than the ZQ dispersion, then $\Delta\varpi_{\text{N}}$ and $\Delta\varpi_{\text{H}}$ have the same sign and the opposite is true when $\Delta\varpi_{\text{N}}$ and $\Delta\varpi_{\text{H}}$ have opposite signs. Relaxation dispersion curves are obtained by analyzing ^1H - ^{15}N correlation maps and the experiment is performed with $^{15}\text{N}/^2\text{H}$ enriched proteins dissolved in 90% H_2O /10% D_2O to back exchange the amide protons.

1. Set the ^1H and ^{15}N carrier frequencies as described in **step 1** of Subheading 3.5. Values of delays are available in [32].
2. Set powers and pulsewidths of the ^1H hard, CPMG pulses and water selective pulses as described in **step 2** of Subheading 3.7. A water selective EPURP1 pulse is used to bring the water magnetization from the Z axis to the transverse plane.
3. Set powers and pulsewidths of ^{15}N hard and CPMG pulses as described in **step 3** of Subheading 3.5. Here also magnetization is transferred to the TROSY component of ^1H magnetization at the end of the pulse sequence and data are acquired without ^{15}N decoupling.
4. Set the T_{CPMG} delay as in **step 5** of Subheading 3.5 and check the S/N in both the ZQ and DQ experiments. It is best to use the same T_{CPMG} delay for both.
5. Set the number of scans in multiples of the phase cycle to obtain good signal to noise. Typical values are 16 or 24 scans.
6. Decide on the ν_{CPMG} values to use as described in **step 7** of Subheading 3.5. ν_{CPMG} values that can be used are $\frac{2n}{T_{\text{CPMG}}}$ where n is an integer starting from 1. ν_{CPMG} values up to 1000 Hz can be used with current probes.
7. Record ^1H - ^{15}N correlation maps for each of the ν_{CPMG} values and a reference plane without the T_{CPMG} delay for both the ZQ and DQ components in an interleaved manner. The usual experimental time varies from 24 to 48 h (*see Note 5*).

3.12 Extracting Relaxation Dispersion Curves from the Data

The NMR data are typically processed using standard programs like NMRPipe [33] and visualized using programs like SPARKY [34]. Peaks intensities in the different 2D planes can be extracted using programs like FUDA [35], PINT [36], and MUNIN [37]. In all three of these programs one can impose that only the peak intensity varies from plane to plane, while the line shape and peak position remain the same in all the planes reducing the number of fitting parameters and improving the precision of the extracted intensities. The programs are also capable of dealing with partially overlapped peaks. The quality of fits can be gauged in FUDA and PINT by visualizing the spectra along with the fitted peaks to decide if the fit is reliable. Once the peaks have been fit to the spectra the dispersion data can be calculated from the intensities. Errors in the $R_{2,\text{eff}}$

values are estimated on the basis of repeat measurements. Generally, a minimum error between 0.3 and 0.5 s⁻¹ is assumed to account for various systematic errors.

3.13 Analyzing Relaxation Dispersion Curves

Initially, amide ¹⁵N/¹H dispersion data are recorded and analyzed. Quite often complete kinetic studies can be performed with only amide ¹⁵N/¹H dispersion data [38]. In the first step larger dispersions where the change in the $R_{2,\text{eff}}$ is greater than 5 or 10 s⁻¹ are analyzed. A two-state exchange model is fit to the dispersion data collected at two different field strengths in a site-specific manner. The programs CATIA [39] and CHEMEX [40] perform the fits by numerically propagating the Bloch-McConnell equations [41]. From the quality of the fits and the site-specific distribution of the exchange rate k_{ex} and the minor state population p_b one can use various statistical tools to judge if there is a single global two-state process or there are multiple processes. If there is a single global process then similar k_{ex} and p_b values will be obtained from almost all the site-specific fits [42] and one can proceed to fit a single process to all the data. If the site-specific k_{ex} and p_b values differ from site to site one can consider fitting a global three-state model [43] or look to see if the k_{ex} and p_b values cluster and that residues close on the structure have similar k_{ex} and p_b in which case one can fit multiple two-state models to sets of residues [44]. If the exchange parameters are not ideal for a CPMG investigation, one can either change the sample conditions like temperature or use other NMR experiments like CEST to study slower processes [45, 46] or $R_{1\rho}$ experiments to study faster process [47]. If k_{ex} is fast and p_b is poorly defined one can use the shifts in peak positions between HSQC and HMQC spectra to define p_b [25]. After the experimental conditions have been established using ¹⁵N/¹H data, other ¹³C/¹H CPMG experiments can be performed. Analysis of the relaxation dispersion data provides only the absolute value of the change in chemical shift between the ground and excited states. The sign information can be obtained comparing peak positions in HSQC spectra recorded at different fields or between HSQC and HMQC spectra recorded at the same field [48]. $R_{1\rho}$ experiments can also provide sign information [49]. Once the sign information is available the chemical shifts of the minor state resonances can be calculated and the spectrum of the minor state can be reconstructed.

3.14 Atomic Resolution Structure of the Minor State

By comparing the minor state chemical shifts with the unfolded state chemical shifts [50] parts of the molecule that are unfolded in the minor state can be identified [51]. In fact, identification of local or complete unfolding can be done with reasonable confidence using just ¹H/¹⁵N data. A preliminary analysis of the minor state backbone chemical shifts using programs like TALOS [52, 53] can provide information about changes in secondary structure between the major and minor states [54–57]. In favorable cases it is possible

to obtain the three-dimensional structures of proteins from just chemical shifts [58, 59] and the structure of a minor state of a protein has been obtained by using this strategy [55]. Using the minor state secondary or three-dimensional structural information mutants can be designed to shift the population toward the minor state conformation and test if the CPMG-derived parameters are meaningful [55]. The quality of structures can significantly be improved by using residual dipolar couplings (RDCs) and residual chemical shift anisotropy shifts (RCSAs) [60, 61]. Minor state RDCs and RCSA shifts can be obtained by performing CPMG experiments in the presence of alignment media [19, 62, 63] and these restraints can also be used to obtain structures of the minor state [54, 56, 57]. Here, we have described CPMG experiments to study the dynamics at protein backbone sites. CPMG experiments have also been developed to study dynamics at side-chain sites [8, 64, 65] and provide information regarding the conformation of the sidechains in the minor state [66, 67].

4 Notes

1. Using a thermocouple is more convenient as it does not require shimming and recording spectra to calculate the temperature as in the case of the standard sample.
2. Alternatively, some newer NMR spectrometers are equipped with a so-called NMR thermometer [68, 69], which measures the temperature inside the NMR sample during the experiment using a second locking agent. If such a system is available and if a second locking agent can be added without interfering with the chemical exchange process under investigation, then the NMR thermometer provides a convenient way of keeping the temperature constant during all experiments.
3. ^{15}N pulsewidths can change when a CPMG pulse train is applied particularly when ^{15}N CPMG experiments are performed using a cryogenically cooled probe. The problem does not occur with ^1H or ^{13}C CPMG experiments. To test if there is a change in the ^{15}N pulsewidths due to the CPMG pulse train, the ^{15}N (CPMG) pulse is calibrated using a ^{15}N CPMG sequence where a gradient, an extra pulse to be calibrated, and a strong gradient are applied immediately after the CPMG pulse train. At this point in the sequence a density element proportional to either N_z , $2N_zH_z$, or $N_z(1-2H_z)$ is present depending on the type of CPMG experiment and the ^{15}N pulsewidth can therefore be calibrated in the usual manner and this is performed for different ν_{CPMG} values. If the ^{15}N pulsewidth depends on ν_{CPMG} , CPMG pulse sequences that

keep the load on the probe constant for various ν_{CPMG} values should be used in order to keep the pulsewidth constant.

4. Increasing the T_{CPMG} delay allows for a greater number of exchange events to occur leading to a greater change in signal intensity as a function of ν_{CPMG} and also allowing for a larger range of ν_{CPMG} values as the lowest possible ν_{CPMG} is $1/T_{\text{CPMG}}$ (or $2/T_{\text{CPMG}}$ based on sequence). However, large T_{CPMG} delays can lead to a large reduction in the S/N due to relaxation losses making it difficult to quantitate the change in intensities. As a compromise, it is common to choose a T_{CPMG} delay where the peak intensity has reduced by half compared with the reference spectrum that has no T_{CPMG} delay.
5. As a large number of pulses are applied during the T_{CPMG} delay, a long interscan delay (d1) has to be used to avoid damaging the probe. Typical d1 values lie between 2.1 and 2.5 s.

Acknowledgments

DFH and PV are grateful to Prof. Lewis E Kay (University of Toronto), Dr. Ranjith Muhandiram, Dr. Algirdas Velyvis, Dr. Dmitry Korzhnev, Dr. Vitali Tugarinov, Dr. Philipp Neudecker, Dr. Patrik Lundstrom, and other members of the Kay lab who exposed them to many of the methods presented here. DFH acknowledges the Biotechnology and Biological Sciences Research Council UK (BBSRC) for financial support. PV is supported by generous startup grant from TCIS/TIFH.

References

1. Karplus M (2000) Aspects of protein reaction dynamics: deviations from simple behavior. *J Phys Chem B* 104:11–27
2. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975) Dynamics of ligand binding to myoglobin. *Biochemistry* 14: 5355–5373
3. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254:1598–1603
4. Palmer AG (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640
5. Sekhar A, Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA* 110:12867–12874
6. Chandler D (1987) *Introduction to modern statistical mechanics*. Oxford University Press, Oxford
7. Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2006) *Protein NMR spectroscopy, principles and practice*, 2nd edn. Academic, New York
8. Mulder FA, Skrynnikov NR, Hon B, Dahlquist FW, Kay LE (2001) Measurement of slow (μs) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* 123: 967–975
9. Palmer AG 3rd, Grey MJ, Wang C (2005) Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. *Meth Enzymol* 394:430–465
10. Korzhnev DM, Kay LE (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* 41: 442–451

11. Hansen DF, Vallurupalli P, Lundstrom P, Neu-decker P, Kay LE (2008) Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J Am Chem Soc* 130:2667–2675
12. Lundstrom P, Vallurupalli P, Hansen DF, Kay LE (2009) Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nat Protoc* 4: 1641–1648
13. McCoy MA, Mueller L (1992) Selective shaped pulse decoupling in NMR - Homonuclear [^{13}C] carbonyl decoupling. *J Am Chem Soc* 114: 2108–2112
14. Geen H, Freeman R (1991) Band-selective radiofrequency pulses. *J Magn Reson* 93: 93–141
15. Morris GA, Freeman R (1979) Enhancement of nuclear magnetic-resonance signals by polarization transfer. *J Am Chem Soc* 101:760–762
16. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114: 10663–10665
17. Ikura M, Kay LE, Krinks M, Bax A (1991) Triple-resonance multidimensional NMR study of calmodulin complexed with the binding domain of skeletal muscle myosin light-chain kinase: indication of a conformational change in the central helix. *Biochemistry* 30: 5498–5504
18. Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665
19. Vallurupalli P, Hansen DF, Stollar E, Meirovitch E, Kay LE (2007) Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci USA* 104: 18473–18477
20. Kupce E, Freeman R (1996) Optimized adiabatic pulses for wideband spin inversion. *J Magn Reson A* 118:299–303
21. Hansen DF, Vallurupalli P, Kay LE (2008) An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B* 112: 5898–5904
22. Jiang B, Yu BH, Zhang X, Liu ML, Yang DW (2015) A ^{15}N CPMG relaxation dispersion experiment more resistant to resonance offset and pulse imperfection. *J Magn Reson* 257:1–7
23. Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broadband decoupling: WALTZ-16. *J Magn Reson* 52: 335–338
24. Ishima R, Torchia DA (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J Biomol NMR* 25:243–248
25. Vallurupalli P, Bouvignies G, Kay LE (2011) Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J Phys Chem B* 115:14891–14900
26. Lundstrom P, Hansen DF, Kay LE (2008) Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: comparison between uniformly and selectively ^{13}C labeled samples. *J Biomol NMR* 42:35–47
27. Lundstrom P, Hansen DF, Vallurupalli P, Kay LE (2009) Accurate measurement of alpha proton chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy. *J Am Chem Soc* 131:1915–1926
28. Kupce E, Freeman R (1995) Stretched adiabatic pulses for broadband spin inversion. *J Magn Reson A* 117:246–256
29. Kupce E, Wagner G (1996) Multisite band-selective decoupling in proteins. *J Magn Reson B* 110:309–312
30. Kupce E, Freeman R (1995) Adiabatic pulses for wide-band inversion and broad-band decoupling. *J Magn Reson A* 115:273–276
31. Brown LR, Sanctuary BC (1991) Hetero-Tocsy experiments with WALTZ and DIPSI mixing sequences. *J Magn Reson* 91:413–421
32. Orekhov VY, Korzhnev DM, Kay LE (2004) Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J Am Chem Soc* 126:1886–1891
33. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe - a multidimensional spectral processing system based on Unix pipes. *J Biomol NMR* 6:277–293
34. Goddard TD, Kneller DG SPARKY 3 University of California, San Francisco
35. Kristensen S, Hansen D (2005) FuDA: a function and data fitting and analysis package
36. Ahlner A, Carlsson M, Jonsson BH, Lundstrom P (2013) PINT: a software for integration of peak volumes and extraction of relaxation rates. *J Biomol NMR* 56:191–202
37. Korzhneva DM, Ibraghimov IV, Billeter M, Orekhov VY (2001) MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. *J Biomol NMR* 21:263–268
38. Sekhar A, Vallurupalli P, Kay LE (2012) Folding of the four-helix bundle FF domain from a compact on-pathway intermediate state is

- governed predominantly by water motion. *Proc Natl Acad Sci USA* 109:19268–19273
39. Hansen DF CATIA. <http://www.ucl.ac.uk/hansen-lab>
 40. Bouvignies G (2012) Chemex. <https://github.com/gbouvignies/chemex/releases>
 41. McConnell HM (1958) Reaction rates by nuclear magnetic resonance. *J Chem Phys* 28: 430–431
 42. Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE (2001) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* 8:932–935
 43. Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430:586–590
 44. Vallurupalli P, Kay LE (2006) Complementarity of ensemble and single-molecule measures of protein motion: a relaxation dispersion NMR study of an enzyme complex. *Proc Natl Acad Sci USA* 103:11910–11915
 45. Vallurupalli P, Bouvignies G, Kay LE (2012) Studying “invisible” excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* 134:8148–8161
 46. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM (2011) Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature* 480: 268–272
 47. Palmer AG, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* 106:1700–1719
 48. Skrynnikov NR, Dahlquist FW, Kay LE (2002) Reconstructing NMR spectra of “invisible” excited protein states using HSQC and HMQC experiments. *J Am Chem Soc* 124: 12352–12360
 49. Auer R, Hansen DF, Neudecker P, Korzhnev DM, Muhandiram DR, Konrat R, Kay LE (2010) Measurement of signs of chemical shift differences between ground and excited protein states: a comparison between H(S/M) QC and $R_{1\rho}$ methods. *J Biomol NMR* 46: 205–216
 50. Tamiola K, Acar B, Mulder FA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132:18000–18003
 51. Sanchez-Medina C, Sekhar A, Vallurupalli P, Cerminara M, Munoz V, Kay LE (2014) Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR. *J Am Chem Soc* 136:7444–7451
 52. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44: 213–223
 53. Shen Y, Bax A (2015) Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Meth Mol Biol* 1260:17–32
 54. Vallurupalli P, Hansen DF, Kay LE (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci USA* 105: 11766–11771
 55. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, Vernon RM, Dahlquist FW, Baker D, Kay LE (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477: 111–114
 56. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein-folding intermediate at atomic resolution. *Science* 329:1312–1316
 57. Neudecker P, Robustelli P, Cavalli A, Walsh P, Lundstrom P, Zarrine-Afsar A, Sharpe S, Vendruscolo M, Kay LE (2012) Structure of an intermediate state in protein folding and aggregation. *Science* 336:362–366
 58. Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685–4690
 59. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615–9620
 60. Prestegard JH, al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* 33(4):371–424
 61. Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. *Meth Enzymol* 339:127–174
 62. Hansen DF, Vallurupalli P, Kay LE (2008) Quantifying two-bond $^1\text{H}^{\text{N}}\text{-}^{13}\text{C}^{\text{O}}$ and one-bond $^1\text{H}^{\alpha}\text{-}^{13}\text{C}^{\alpha}$ dipolar couplings of invisible protein states by spin-state selective relaxation dispersion NMR spectroscopy. *J Am Chem Soc* 130:8397–8405
 63. Vallurupalli P, Hansen DF, Kay LE (2008) Probing structure in invisible protein states with anisotropic NMR chemical shifts. *J Am Chem Soc* 130:2734–2735

64. Skrynnikov NR, Mulder FA, Hon B, Dahlquist FW, Kay LE (2001) Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion NMR measurements: application to methionine residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* 123:4556–4566
65. Lundstrom P, Lin H, Kay LE (2009) Measuring $^{13}\text{C}^{\beta}$ chemical shifts of invisible excited states in proteins by relaxation dispersion NMR spectroscopy. *J Biomol NMR* 44:139–155
66. Hansen DF, Neudecker P, Vallurupalli P, Mulder FA, Kay LE (2010) Determination of Leu side-chain conformations in excited protein states by NMR relaxation dispersion. *J Am Chem Soc* 132:42–43
67. Hansen DF, Kay LE (2011) Determining valine side-chain rotamer conformations in proteins from methyl ^{13}C chemical shifts: application to the 360 kDa half-proteasome. *J Am Chem Soc* 133:8272–8281
68. Hiller S, Arthanari H, Wagner G (2009) The T-lock: automated compensation of radio-frequency induced sample heating. *J Biomol NMR* 44:69–76
69. Ammann C, Meier P, Merbach AE (1982) A simple multi-nuclear NMR thermometer. *J Magn Reson* 46:319–321

Probing the Atomic Structure of Transient Protein Contacts by Paramagnetic Relaxation Enhancement Solution NMR

Vincenzo Venditti and Nicolas L. Fawzi

Abstract

Important biological processes, including enzyme catalysis, signaling, and protein folding, proceed through lowly populated (<5%) states that elude structural characterization by conventional techniques. Here, we describe the steps required for visualization of these sparsely populated conformations and transient protein-protein interactions using paramagnetic relaxation enhancement solution NMR. We describe experimental design, sample preparation, data acquisition and processing, and the basics of data analysis of structural ensembles.

Key words Transient interactions, Protein-protein interactions, NMR spectroscopy, Site-directed spin-labeling, Encounter complex, Lowly populated states, Dark states

1 Introduction

The history of Paramagnetic Relaxation Enhancement (PRE) in solution NMR applications dates back to the early work of Solomon and Bloembergen who established that, in paramagnetic solutions, the dipolar coupling between a nucleus and an unpaired electron results in an increase in the nuclear spin relaxation rates [1, 2]. The PRE is proportional to the population-averaged $\langle r^{-6} \rangle$ distance between the paramagnetic center and the nucleus. Owing to the large magnitude of the electron magnetic moment, the PRE can extend for distances up to ~ 35 Å from the paramagnetic center, making it an important source of long-range distance restraints in structure calculation protocols. Due to this very strong dipolar interaction and very steep distance dependence, PREs are also uniquely able to detect and characterize lowly populated (<5%) conformational states [3]. Although they are sampled only transiently, these lowly populated states underlie a range of crucial biological processes, acting as encounter complexes between protein interaction partners and as intermediates in enzyme catalysis, protein folding, and aggregation [4, 5]. In order to be detected by

PRE experiments, the minor (invisible to conventional NMR techniques) state must undergo rapid exchange with the major (NMR visible) state, and the distance between the paramagnetic center and the nuclei of interest must be shorter in the minor state than in the major state [3]. When these conditions are satisfied, because a nucleus in close proximity to a paramagnetic center exhibits an extremely large PRE ($>1000\text{ s}^{-1}$), the contribution from the close-proximity state will dominate the overall PRE measured for that nucleus, even if it constitutes a minor percentage of the total population. Therefore, PRE NMR for transient states takes advantage of the strong dipolar interaction and r^{-6} (i.e., steep) distance dependence. Currently, PRE is the technique of choice to provide direct structural information on lowly populated conformational states and transient encounter complexes.

Despite its ability to provide unique structural information, for about 50 years the use of paramagnetic NMR had been restricted primarily to metalloproteins, where a paramagnetic center can be easily introduced by replacing the natural metal cation with a paramagnetic one [6, 7]. PRE has become a routine tool in biomolecular NMR spectroscopy only in the last two decades after the introduction of straightforward biochemical methods for site-specific incorporation of paramagnetic labels in proteins and nucleic acids [8–10] and the development of a robust theoretical framework to account for the intrinsic flexibility of artificially introduced paramagnetic tags in back-calculation of the PRE from atomic-resolution structures [11, 12]. Here, we describe in detail a protocol that takes advantage of the most recent technical advances in paramagnetic NMR spectroscopy to characterize transient *intramolecular* protein contacts (e.g., for proteins with a flexible hinge [13]). The notes describe modifications of the protocol necessary to detect transient *intermolecular* contacts.

2 Materials

Prepare all the solutions using deionized water and analytical grade reagents. Prepare and store all the reagents at room temperature (unless indicated otherwise).

2.1 Preparation of Protein and Labeling Material

1. Protein stock solution: $\sim 0.1\text{ mM}$ (for one NMR sample) of purified protein in 2 mL of any buffer in H_2O . The protein should be uniformly ^{15}N labeled (e.g., grown in M9 minimal media with $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source) (*see Note 1* regarding experiments to detect transient intermolecular contacts). Protein hydrogen atoms may be deuterated at carbon-attached positions for large proteins to improve the quality of spectra. Store at $4\text{ }^\circ\text{C}$. The protein should contain a single surface cysteine for spin-labeling. We recommend purifying

and storing the protein stock in a buffer containing a reducing agent (e.g., 2 mM DTT). Data from several single cysteine variants may be necessary to characterize structure. Care must be taken when selecting a labeling site that is close to the interface being interrogated to ensure that labeling with MTSL does not disrupt binding or create label-dependent binding. Comparing spectra of diamagnetically labeled protein (i.e., the MTSL-labeled protein with the unpaired electron reduced by ascorbate addition, or the protein labeled with a diamagnetic analogue of the paramagnetic tag) to the wild-type protein (see below) can be used to ensure the site is appropriate.

2. MTSL labeling agent: (1-Oxyl-2,2,5,5-tetramethyl- Δ^3 -pyrroline-3-methyl) Methanethiosulfonate (Toronto Research Chemicals, O875000) (*see Note 2*). We typically use 10 mg vial size to have a fresh label stock. Stock solution: 100 mM solution in ethanol. Store at -20°C (*see Note 3*).
3. Dithiothreitol (DTT) stock solution: 0.5 M in protein buffer (*see Note 4*). Store at -20°C .
4. Desalting buffer: 20 mM Tris-HCl (pH 7.4–8.0) (*see Note 5*).
5. Ascorbate stock solution: 100 mM in NMR buffer, readjust pH after dissolution.

2.2 NMR Sample Preparation

1. NMR tube: wash a conventional NMR tube (do not use Shigemi—*see Note 6*) with concentrated nitric acid (*see Note 7*). Then rinse the tube with 200 mL of water. Completely dry the tube before usage.
2. NMR buffer: prepare 100 mL (or more) of the working NMR buffer (i.e., the buffer that results in the best quality ^1H - ^{15}N HSQC spectra of the protein of interest). If possible, we recommend using a slightly acidic buffer (i.e., pH 6.5 instead of 7.4) (*see Note 8*). Do not add reducing agents to the NMR buffer (*see Note 9*). Purify the buffer from any residual ion using Chelex-100 (Sigma). We suggest filtering the NMR buffer with a vacuum filtration unit on whose membrane a thin layer of Chelex-100 has been applied. Rinse the Chelex-100 with 300 mL of water (twice) before filtering the NMR buffer.

2.3 NMR Data Acquisition

1. Pulse sequence: any pulse sequence for measuring ^1H transverse relaxation rate constants can be used to collect PRE data. We suggest the sequences used by Clore and coworkers [14].
2. Data processing: any processing software can be used for post-acquisition processing of the NMR spectra. Processing parameters are strongly dependent on the NMR pulse program.

2.4 Fitting Structural Models to PRE Data

1. $^1\text{H}/^{15}\text{N}$ spectral assignments of protein of interest in the experimental conditions used to acquire the PRE data.
2. XPLOR-NIH software and processing scripts. XPLOR-NIH software is available at this site for academic download and use: <http://nmr.cit.nih.gov/xplor-nih> [15, 16]. Scripts for back calculation of PREs from 3D structural models, and for refinement of protein 3D structures against PRE data are available at <https://spin.niddk.nih.gov/clore/> on the Software page at the Spin Label Build link.

3 Methods

3.1 Site-Specific Labeling Reaction

1. Prepare the $^1\text{H}/^{15}\text{N}$ labeled protein for labeling by adding DTT to a final concentration of 10 mM to ensure any intermolecular disulfide bonds are reduced. To enable completion of the following two desalting steps without need to concentrate the sample, keep the protein sample volume to 15 mL or less. Incubate for at least 1 h at room temperature or at 4 °C overnight.
2. Equilibrate the HiPrep 26/10 desalting column with the labeling buffer (without reducing agent).
3. To remove the DTT, pass 7 mL (or less) of the protein/DTT mixture through the equilibrated HiPrep 26/10 Desalting column. Use a flow rate of 7 mL/min. Collect the fractions containing the protein and discard all the fractions containing DTT. If a UV detector is connected to the chromatographic system in use, the protein fractions can be identified by inspecting the chromatogram. Otherwise, gel electrophoresis can be used to locate the protein fractions. It is important to completely separate the protein from DTT. DTT has a small absorbance at 280 nm and can be detected in the UV chromatogram. In any case, due to the small molecular size, DTT elutes from a 26/10 desalting column at volume >25–30 mL. In order to facilitate the following steps of the labeling protocol, the recollected protein sample should be <15 mL.
4. Immediately add the dissolved MTSL label stock to the tube containing the desalted protein. The label should be added in 5× (or higher) excess to the protein and should have a final concentration ≥ 1 mM.
5. Incubate the reaction at room temperature in the dark for 2 h.
6. During the reaction, equilibrate the desalting column with the experimental buffer (without reducing agent).
7. To end the reactions, pass the protein/MTSL mixture through the HiPrep 26/10 desalting column at 7 mL/min. Collect all the fractions containing the protein and discard the later

fractions containing the unreacted MTSL (*see step 3* above). As in the case of DTT, MTSL has a weak UV absorbance at 280 nm. Therefore, the MTSL peak can be detected if an UV detector is connected to the chromatographic apparatus. Note that small contamination from unreacted MTSL may be tolerated because it will be further eliminated during the buffer exchange step (*see step 8*).

8. Concentrate and buffer exchange the protein into the NMR buffer of interest. We suggest using Amicon Ultra-15 centrifugal filter units (EMD Millipore) of the appropriate molecular weight cutoff.
9. To confirm labeling efficiency, total mass determined by LC/MS for a sample after labeling can be compared to a sample before labeling. MTSL addition at a cysteine position (forming the “R1” sidechain using the terminology of Hubbell and coworkers [9]) adds 184 Da.

3.2 NMR Sample Preparation

1. In a 1.6 mL tube, prepare a 500 μL sample of up to 200 μM (*see Note 10*) of the MTSL labeled protein in NMR buffer including up to 10% D_2O for NMR lock (*see Note 11* regarding estimating protein concentration). Transfer to a standard NMR tube.
2. Prepare a 500 μL sample with half of the concentration which will be used to check for the presence of intermolecular interactions.

3.3 NMR Data Collection

1. Record a two-dimensional, $^1\text{H}/^{15}\text{N}$ correlation spectrum (HSQC or TROSY) and compare the signal intensities (or signal-to-noise ratios) to unlabeled reference spectra. Confirm that no significant chemical shift differences are present.
2. Using the high concentration sample, set up a $^1\text{H}_\text{N}$ - R_2 NMR experiment. For the quantitative interpretation of PRE data, we strongly recommend measurement of $^1\text{H}_\text{N}$ - R_2 using a two-time-point approach as opposed to attempting to measure PREs by quantifying intensity attenuations in two-dimensional correlation data—*see* discussion in ref. 14. Adjust the length of the relaxation delay for the second time-point (T_b) such that T_b is $1.15/({}^1\text{H}_\text{N}\text{-}R_{2,\text{dia}} + \Gamma_2)$, where Γ_2 is the PRE value that should be measured with greatest accuracy (*see* details in Iwahara et al. [14]). A value of 15 ms is common for anticipated Γ_2 values up to 50 s^{-1} . In practice, choice of T_b with 50% signal intensity compared to intensity at T_a , the reference delay ($= 0 \text{ ms}$) time point, is a good starting place. It is recommended to check the lower concentration sample at this point to ensure that a significant concentration dependence of $^1\text{H}_\text{N}$ - $R_{2,\text{para}}$ is not already observed in the first increments, which would suggest contribution of intermolecular interactions.

3. Record the full $^1\text{H}_\text{N}$ - R_2 two-time-point experiment with sufficient signal to noise (typically using 32 scans or more), as described previously [14].
4. Record the full $^1\text{H}_\text{N}$ - R_2 two-time-point experiment for the sample at half concentration, if desired.
5. To make the diamagnetic reference sample, add a concentrated ascorbate stock solution to a final concentration of 2 mM, incubate at room temperature for 1 h to reduce the nitroxide radical.
6. Record the full $^1\text{H}_\text{N}$ - R_2 two-time-point experiment on the reduced sample with sufficient signal to noise. The delay T_b must be kept the same due to the contribution of $^3\text{J}_{\text{HN/HA}}$ scalar coupling to the evolution of the NMR signal. We recommend using the exact same experimental parameters, including number of transients, as for the paramagnetic sample.

3.4 NMR Data Processing

1. Process the interleaved NMR relaxation spectra with the appropriate spectral parameters using nmrPipe [17] or another equivalent approach. The [series.com](https://spin.niddk.nih.gov/bax/software/NMRPipe/index.html) script of the nmrPipe distribution is a convenient approach (<https://spin.niddk.nih.gov/bax/software/NMRPipe/index.html>) for processing data from NMR relaxation experiments. Alternatively, any other software package that performs similar tasks may be used.
2. Transfer known assignments to measure peak intensities in the paramagnetic and diamagnetic sample data. The [ipap.com](https://spin.niddk.nih.gov/bax/software/NMRPipe/index.html) script of the nmrPipe distribution is convenient for this process.
3. Measure the peak intensities in the two time point spectra for the paramagnetic and diamagnetic conditions.
4. Measure the standard deviation of the noise (σ), for example using the statistics function in nmrDraw.
5. Compute the values of Γ_2 , the paramagnetic relaxation enhancement values, at each H/N position from the intensities using Eq. 5 in Iwahara et al. [14]

$$\Gamma_2 = R_{2,\text{para}} - R_{2,\text{dia}} = \frac{1}{T_b - T_a} \ln \frac{I_{\text{dia}}(T_b)I_{\text{para}}(T_a)}{I_{\text{dia}}(T_a)I_{\text{para}}(T_b)}$$

where $I_{\text{dia}}(T_a)$ and $I_{\text{para}}(T_a)$ are the peak intensities in the diamagnetic and paramagnetic spectra, respectively, at time point T_a , and $I_{\text{dia}}(T_b)$ and $I_{\text{para}}(T_b)$ are the peak intensities in the diamagnetic and paramagnetic spectra, respectively, at time point T_b .

6. Compute the values of $\sigma(\Gamma_2)$, the uncertainty, or standard deviation of the paramagnetic relaxation enhancements, at each H/N position from the intensities using Eq. 6 in Iwahara et al. [14]

$$\sigma(\Gamma_2) = \frac{1}{T_b - T_a} \sqrt{\left\{ \frac{\sigma_{\text{dia}}}{I_{\text{dia}}(T_a)} \right\}^2 \left\{ \frac{\sigma_{\text{dia}}}{I_{\text{dia}}(T_b)} \right\}^2 \left\{ \frac{\sigma_{\text{para}}}{I_{\text{para}}(T_a)} \right\}^2 \left\{ \frac{\sigma_{\text{para}}}{I_{\text{para}}(T_b)} \right\}^2}$$

where σ_{dia} and σ_{para} are the standard deviation of the noise measured for the diamagnetic and paramagnetic spectra, respectively (*see step 3*).

7. Verify that PREs are large for backbone H/N groups within 5–10 Å of the paramagnetic center. Also, remove unreliable PREs from the list before further analysis. A convenient approach to judge reliability of experimental PREs is to plot the experimental PREs versus $\log(I_{\text{dia}}(T_a)/I_{\text{para}}(T_a))$. A linear correlation is expected. Outliers must be removed at this point from the list of experimental PREs [3].

3.5 Structural Fit of Tag Positions to Intra-domain Data

1. Create a list of residues within the same domain as the tag position (i.e., that are not expected to undergo backbone motion relative to the label site).
2. For these residues, create a file `fittag.tbl` specifying the Γ_2 and $\sigma(\Gamma_2)$ at each position using the following format, where `res id` is the residue number and the last two values are Γ_2 and $\sigma(\Gamma_2)$:

```
assign (name NS1) (resid 32 and name HN) 15.4939 2.3295
```

More details and an example data file are described in the Spin Label Build README file.

3. Using XPLOR-NIH, follow the procedure described in the Spin Label Build README file to: (a) set the protein rotational correlation time in `fit.py` and `calc.py`, (b) rename the tagged residue name to “CYSP” in the `pdb` file, (c) automatically generate a `psf` file and add any necessary atoms to the `pdb` file (`xplor -py buildtag.py`), and (d) build a five-label conformer model (`xplor make5conf.inp`—*see Note 12*).
4. Edit the `fit.py` file to replace the single instance of `data.tbl` with your file name `fittag.tbl` from above (*see step 2*).
5. Calculate a structural ensemble of the five tag positions that best reproduce the observed intra-domain PRE data, as described in the Spin Label Build README file (`xplor -py fit.py > fit.out`).

3.6 Calculate and Analyze Predicted PRE Data for the Entire Structure

1. Calculate the PREs for all the backbone H/N groups using the reference pdb file including the optimized five-conformers representation of the paramagnetic tag (`xplor -py calc.py > calc.out`).
2. Extract the predicted PREs from the output file using the script `GrabPREs.awk` provided by the Spin Label Build package (`GrabPREs.awk > GrabPREs.out`).
3. Graph the predicted (i.e., calculated) and observed (i.e., experimental) PREs vs. residue number. Because the experimental values were used to optimize the spin label positions for back calculation, the predicted and observed PREs should match within experimental error for the intra-domain PREs (i.e., all the experimental PREs listed in the `fittag.tbl` file). For all other PREs, deviations of the experimental values from those predicted from the reference pdb file suggest transient protein conformational change. For example, higher than predicted PREs suggest transient approach of the tag to these positions.
4. Repeat the protocol for each tag position of interest to validate the observed transient conformational change. For example, given elevated PREs observed near region A when the label is placed in region B, a sample with the label placed in region B should show elevated PREs in region A, confirming that the system samples a closer approach of the two regions than is represented by the static structure in the reference pdb file.
5. The inter-domain PREs from several label positions can also be simultaneously used to generate a dynamic ensemble of structures quantitatively consistent with all experimental data (*see* the approach of Tang et al. [13]). This approach requires advanced use of XPLOR-NIH with scripts available on the software website. The most faithful representation of the solution structural ensemble can be achieved by combining PREs with structural restraints from complementary solution-phase techniques, such as small and wide angle X-ray scattering (SAXS/WAXS) data and NMR residual dipolar couplings [18].

4 Notes

1. This protocol describes the use of PRE NMR to observe transient intramolecular contacts. A parallel approach can be used to detect transient intermolecular contacts. The changes to the protocol are briefly described here for clarity. For intermolecular contacts, NMR samples constitute both an NMR silent (e.g., ^{14}N , natural abundance)/paramagnetically labeled species with an NMR visible (^{15}N isotopically labeled) species. Preparation of the paramagnetically labeled species is

unchanged from this protocol. By positioning the label at several locations, intermolecular PREs can be used to define the surfaces mediating intermolecular contacts. However, quantitative interpretation of intermolecular PREs requires accurate positioning of the label on the protein surface, which is best accomplished by the intramolecular PRE approach described in this protocol (e.g., measurement of intramolecular PREs on a sample simultaneously and uniformly MTSL-labeled and ^{15}N -labeled using the protocol described in Subheading 3 is advised as a precondition for intermolecular PREs). Additional considerations concerning intermolecular PREs can be found in previous review articles [3, 14].

2. Several choices for spin label exist. The most flexible and best characterized is the stabilized nitroxide radical MTSL which attaches via a single disulfide bond. A variety of MTSL derivatives with substituent groups [19] and two disulfide bond sites [20] exist if a more rigid label is desired. An acetylated diamagnetic analogue of MTSL, (1-Acetoxy-2,2,5,5-tetramethyl- δ -3-pyrroline-3-methyl) Methanesulfonate (Toronto Research Chemicals, A167880) is available to aid characterization of the diamagnetic state. If the hydrophobic character of the spin-label results in label-dependent contact formation or structural changes, a functionalized form of EDTA chelated with a paramagnetic metal ion such as Mn^{2+} provides a potential alternative [12]. In this case, a diamagnetic metal chelate (e.g., Ca^{2+} or Mg^{2+}) can be used as a diamagnetic control. For proteins where a single surface-free cysteine cannot be engineered, spin labels can also be attached at non-natural amino acids using bio-orthogonal labeling [21] though special protocols must be used for labeling as well as protein expression in minimal media for isotopic labeling [22]. To determine if the combination of spin label and position is appropriate for the system, the NMR spectra, titration, or NMR dynamical observables of the diamagnetic form (with ascorbate or with the acetylated MTSL analogue) can be compared to the wild-type protein.
3. The stock of MTSL can be aliquoted and kept at $-20\text{ }^{\circ}\text{C}$ for a week or flash frozen and stored at $-80\text{ }^{\circ}\text{C}$. Because MTSL is used in excess, slight precipitation of the compound in the stock is not of high concern.
4. Any reducing agent can be chosen including TCEP. However, we prefer DTT to β -mercaptoethanol (β ME) because DTT is a stronger reducing agent and β ME can form adducts at cysteine positions. Although some studies report that TCEP will not interfere with disulfide bond formation by methanethiolsulfonate labeling reagents, this has not been our experience.

5. Spin-labeling can be performed in any sulfhydryl-free buffer including phosphate and Good's buffers. However, the spin-labeling reaction via disulfide linkage to a cysteine residue requires deprotonation of the cysteine side-chain and is faster at basic pH. Therefore, we recommend running the reaction at $\text{pH} > 7.0$. We have had good experience with $\text{pH} 8.0$.
6. Plugs in Shigemi tubes are optimized to match the magnetic susceptibility of pure water. Paramagnetic centers substantially alter the magnetic susceptibility of the NMR sample. The use of Shigemi tubes with paramagnetic samples may render the shimming step challenging.
7. NMR tubes—even the brand-new ones—may contain trace paramagnetic heavy metal contamination. Rinsing the tubes with nitric acid followed by standard cleaning with water and ethanol ensures removal of these contaminants that may introduce unwanted paramagnetic ions in the sample.
8. Detection of PREs using $^1\text{H}_\text{N}$ positions is convenient due to the large gyromagnetic ratio of hydrogen and the excellent spectral dispersion of backbone amide positions, which results in single residue resolution. However, base-catalyzed exchange of H_N positions with water hydrogens is rapid at high temperatures and high pH. These water exchange events are therefore an additional relaxation mechanism for proton transverse magnetization, resulting in an increase in the background $^1\text{H}_\text{N}$ - R_2 . In addition, slight changes in pH can have a large effect on rates under these conditions. Exchange with water may also cause a reduction in the intensity or, in the most extreme cases, the complete disappearance of NMR cross-peaks, especially at H_N positions that are exposed to the solvent and not involved in intramolecular hydrogen bonds. Therefore, wherever possible, we recommend running PRE experiments at acidic pH (e.g., $\text{pH} 6.5$ or lower—the minimum exchange is at $\text{pH} 3\text{--}4$) and low temperature. We also recommend avoiding Tris buffer that has a high temperature dependence on pH. Predicted exchange rates for disordered regions lacking stable hydrogen bonds can be made using the program Sphere from the Roder group (<http://landing.foxchase.org/research/labs/roder/sphere/sphere.html>).
9. There are commercially available spin-labels that bind the protein using nonreducible covalent linkers. Reducing agents should not be used in PRE experiments, even if using this class of spin-label compounds. Indeed, reducing agents can reduce the paramagnetic center and hence quench the PRE.
10. If the concentration of the spin-labeled species is too high, so called solvent PRE [23] effects can dominate the observed paramagnetic relaxation enhancements. These PREs arise

from random intermolecular collisions rather than physiologically relevant transient interactions. Typical advice is therefore to limit samples to 300 μM or less, though this value will be system dependent. Proteins that self-assemble or favorably interact with the paramagnetic tag will show intermolecular PREs at lower concentrations [24].

11. The labels are attached by a disulfide bond that adds a small contribution ($\sim 125 \text{ M}^{-1} \text{ cm}^{-1}$) to the absorbance at 280 nm, A_{280} . Delocalization of the radical in MTSL adds approximately $700 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm (note the compound is yellow) based on our empirical tests. This additional absorbance can be significant contribution to A_{280} in proteins without tryptophan residues. In our experience, acetylated diamagnetic control labels are colorless and do not absorb at 280 nm, aside from the disulfide linker.
12. The optimal number of conformations for the spin-label depends on numerous experimental factors, such as the chosen paramagnetic tag, the labeling site, the temperature, the pH, and others. When determining the ensemble of label positions that best reproduce the experimental data, as a general rule we perform several optimization runs on simulated systems with increasing number of spin-label conformers (typically ranging from 1 to 5). For each system the agreement between experimental and back-calculated PREs is evaluated by χ^2 statistic or by the Q -factor [11, 14], which is analogous to the crystallographic R -factor:

$$Q = \sqrt{\frac{\sum_i (\Gamma_2^{\text{obs}}(i) - \Gamma_2^{\text{calc}}(i))^2}{\sum_i \Gamma_2^{\text{obs}}(i)^2}}$$

In general, increasing the number of spin-label conformers results in a lower Q -factor, indicative of better agreement between experimental and back-calculated PREs. However, when the number of conformers is sufficient to fully account for spin-label flexibility, further increasing the ensemble size will not cause a significant reduction of Q -factor.

Acknowledgments

Research reported in this publication was supported in part by the National Institute Of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Number R01GM118530 (to N.L.F) and by startup funding from Iowa State University (to V.V.).

References

- Bloembergen N, Morgan OL (1961) Proton relaxation times in paramagnetic solutions effects of electron spin relaxation. *J Chem Phys* 34:842–850
- Solomon I (1955) Relaxation processes in a system of two spins. *Phys Rev* 99:559–565
- Anthis NJ, Clore GM (2015) Visualizing transient dark states by NMR spectroscopy. *Q Rev Biophys* 48(1):35–116. doi:[10.1017/S0033583514000122](https://doi.org/10.1017/S0033583514000122)
- Clore GM (2011) Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. *Protein Sci* 20(2):229–246. doi:[10.1002/pro.576](https://doi.org/10.1002/pro.576)
- Sekhar A, Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA* 110(32):12867–12874. doi: [10.1073/pnas.1305688110](https://doi.org/10.1073/pnas.1305688110)
- Bertini I, Luchinat C, Piccioli M (2001) Paramagnetic probes in metalloproteins. *Meth Enzymol* 339:314–340
- Cheng H, Markley JL (1995) NMR spectroscopic studies of paramagnetic proteins: iron-sulfur proteins. *Annu Rev Biophys Biomol Struct* 24:209–237. doi:[10.1146/annurev.bb.24.060195.001233](https://doi.org/10.1146/annurev.bb.24.060195.001233)
- Donaldson LW, Skrynnikov NR, Choy WY, Muhandiram DR, Sarkar B, Forman-Kay JD, Kay LE (2001) Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. *J Am Chem Soc* 123(40):9843–9847
- Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J, Hubbell WL, Rosevear PR (2000) Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci* 9(2):302–309. doi:[10.1110/ps.9.2.302](https://doi.org/10.1110/ps.9.2.302)
- Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39(18):5355–5365
- Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126(18):5879–5896. doi:[10.1021/ja031580d](https://doi.org/10.1021/ja031580d)
- Tang C, Iwahara J, Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* 444(7117):383–386. doi:[10.1038/nature05201](https://doi.org/10.1038/nature05201)
- Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449:1078–1082. doi: [10.1038/nature06232](https://doi.org/10.1038/nature06232)
- Iwahara J, Tang C, Clore GM (2007) Practical aspects of ¹H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson* 184(2):185–195. doi: [10.1016/j.jmr.2006.10.003](https://doi.org/10.1016/j.jmr.2006.10.003)
- Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Res Spectrosc* 48(1):47–62. doi:[10.1016/j.pnmrs.2005.10.001](https://doi.org/10.1016/j.pnmrs.2005.10.001)
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73. doi:[10.1016/S1090-7807\(02\)00014-9](https://doi.org/10.1016/S1090-7807(02)00014-9)
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293
- Venditti V, Egnér TK, Clore GM (2016) Hybrid approaches to structural characterization of conformational ensembles of complex macromolecular systems combining NMR residual dipolar couplings and solution X-ray scattering. *Chem Rev* 116(11):6305–6322. doi:[10.1021/acs.chemrev.5b00592](https://doi.org/10.1021/acs.chemrev.5b00592)
- Fawzi NL, Fleissner MR, Anthis NJ, Kalai T, Hideg K, Hubbell WL, Clore GM (2011) A rigid disulfide-linked nitroxide side chain simplifies the quantitative analysis of PRE data. *J Biomol NMR* 51(1–2):105–114. doi:[10.1007/s10858-011-9545-x](https://doi.org/10.1007/s10858-011-9545-x)
- Fleissner MR, Bridges MD, Brooks EK, Cascio D, Kalai T, Hideg K, Hubbell WL (2011) Structure and dynamics of a conformationally constrained nitroxide side chain and applications in EPR spectroscopy. *Proc Natl Acad Sci USA* 108(39):16241–16246. doi: [10.1073/pnas.1111420108](https://doi.org/10.1073/pnas.1111420108)
- Fleissner MR, Brustad EM, Kalai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peucker S, Schultz PG, Hubbell WL (2009) Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc Natl Acad Sci USA* 106(51):21637–21642. doi: [10.1073/pnas.0912009106](https://doi.org/10.1073/pnas.0912009106)

22. Venditti V, Fawzi NL, Clore GM (2012) An efficient protocol for incorporation of an unnatural amino acid in perdeuterated recombinant proteins using glucose-based media. *J Biomol NMR* 52(3):191–195.
doi:[10.1007/s10858-012-9606-9](https://doi.org/10.1007/s10858-012-9606-9)
23. Bernini A, Venditti V, Spiga O, Niccolai N (2009) Probing protein surface accessibility with solvent and paramagnetic molecules. *Prog Nucl Magn Res Spectrosc* 54:278–289
24. Venditti V, Fawzi NL, Clore GM (2011) Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. *J Biomol NMR* 51(3):319–328.
doi:[10.1007/s10858-011-9559-4](https://doi.org/10.1007/s10858-011-9559-4)

Chapter 13

From Raw Data to Protein Backbone Chemical Shifts Using NMRfx Processing and NMRViewJ Analysis

Bruce A. Johnson

Abstract

Assignment of the chemical shifts of the backbone atoms (HN, N, CA, CB, and C) of proteins is often a prerequisite to using NMR information in the study of proteins. These chemical shifts and their perturbations are the basis for the analysis of protein dynamics, ligand binding, and backbone conformation. They are generally assigned prior to full side-chain assignments and the determination of the complete three-dimensional molecular structure. This chapter describes the use of two software packages, NMRfx Processor and NMRViewJ, in going from raw NMR data to backbone assignments. The step-by-step procedure describes processing of the data and the use of manual and automated features of the RunAbout tool in NMRViewJ to perform the assignments.

Key words Nuclear magnetic resonance (NMR), Chemical shifts, Backbone assignments, NMRViewJ, NMRfx Processor

1 Introduction

NMR Spectroscopy of proteins derives much of its utility from the ability to have spectroscopic information that can be assigned to specific atoms in the structure. Monitoring the chemical shifts, relaxation rates, and NOE interactions of specific atoms allows the measurement of biologically important properties such as ligand binding, conformation, molecular dynamics, and the full molecular structure. Significant information can be obtained by the spectroscopic analysis of just the backbone atoms, so assigning the chemical shifts of these atoms is often a key step in the NMR analysis of a protein. Molecular dynamics [1] and ligand binding experiments [2] are often focused on monitoring the relaxation rate and chemical shift perturbations of the amide proton and nitrogen. Information about secondary structure is generally derived from the shifts of these atoms, and the CA, CB, and C atoms [3]. Furthermore, the assignment of the backbone atoms is also generally performed prior to full assignment of the side-chain assignments. Given the

fundamental importance of obtaining backbone chemical shifts, this chapter will focus on using two software packages to process and analyze NMR data with the goal of obtaining assignments for the HN, N, CA, CB, and C atoms of a protein.

The techniques described here make use of a set of three-dimensional, proton-detected, NMR spectra collected on protein that is isotopically labeled with ^{15}N and ^{13}C (and optionally ^2H for larger proteins). Protocols for sample preparation and data collection will not be described in this chapter. A variety of combinations of experiments can be used for the assignment. The methods of this chapter will be illustrated with data from ubiquitin, and this sample data has what could be considered a full set of backbone data. Three pairs of experiments are used that give inter-residue ($i - 1$) and intra-residue (i) connectivities to the amide proton. One pair (HNCO and HN(CA)CO) connects the ^{13}C spins, the second pair (HN(CO)CA and HNCA) connects the $^{13}\text{C}^\alpha$ spins, and the final pair (HNCACB and CBCACONH) connects the $^{13}\text{C}^\beta$ (and redundantly for the HNCACB experiment, the $^{13}\text{C}^\alpha$) spins. For this analysis, we will rely on one of the experiments as a reference experiment. Here, we will use the HNCO experiment as it is simple (giving rise to only a single peak per residue) and the most sensitive of the set. It is also possible to use a ^1H , ^{15}N HSQC experiment as the reference experiment, but the additional dimension (^{13}C) of the HNCO experiment has the advantage of making it more likely to find a resolved peak for each residue.

A wide variety of applications are available for NMR processing [4, 5], analysis [6–10], and assignment [11–15]. In this chapter, we will use two programs developed in the author's group. The protocols described below do the processing using NMRFX Processor [16], a recently developed NMR data processing program. NMRFX Processor is written in the Java programming language so it will run on all major desktop operating systems (macOS, Windows, Linux). NMRFX Processor operates in GUI mode (as described below) or command line mode. In both the cases, the processing is done by executing a Python script that consists of a series of processing operations. In the following methods, we will make use of a subset of the available operations. Each operation can be configured with a variety of optional parameters. We will use only a subset of these parameters as the default values will mostly suffice for the datasets used here. Consult the actual documentation for full details on all operations and their parameters.

The visualization and spectrum assignment will be done with NMRViewJ [17, 18], a well-known and highly used program for macromolecular spectral analysis. Like NMRFX Processor, it is written in Java and runs on all major desktop operating systems. Together, these two programs will allow one to do all the analysis beginning with processing of the raw NMR data to final assignment of the spectra.

2 Tools and Data

1. Download and install NMRfx Processor. It is currently available (for free) at <http://www.nmrfx.org> or can be used pre-installed within the NMRBox virtual environment, available at <http://nmrbox.org>. To download and install do the following.
 - (a) Proceed to <http://www.nmrfx.org> and navigate to the Processor page.
 - (b) Click on the Downloads Link.
 - (c) Register (link at top-right of page) to get a username and password.
 - (d) Login to the site.
 - (e) Links are present for macOS, Windows (32bit and 64bit) and Linux (32bit and 64bit). Choose the link appropriate to your computer operating system.
 - (f) On macOS, the downloaded file will be a .dmg file. Just double-click on the downloaded file to open it and, in the folder that appears, drag the NMRfx Processor icon onto the Application folder icon.
 - (g) On Windows, the downloaded file will be an executable installer. Double-click on the installer and proceed through the installation steps in the installer.
 - (h) On Linux, the downloaded file will be an executable shell script. Open a terminal window, navigate to the location of the downloaded file and type “sh filename,” where filename is the downloaded file, and then follow the installation steps in the installer.
2. Download and install NMRViewJ or use the version available in the NMRBox virtual environment. The protocol is the same as above, but choose the NMRViewJ page. At present, NMRViewJ requires an activation license to run. A free license for unsupported use is available by clicking on the *Demonstration and Unsupported Use License* link on the Download page. Copy this license and paste it into the License Manager window that appears when NMRViewJ starts up (also available via the License... menu entry under the NMRViewJ Help menu).
3. Copy the appropriate backbone-assignment NMR experiments to a convenient location on your computer. Typically, these will include experiments such as HSQC, HNC0, HNCA, HNCACB, HN(CO)CA, HN(CO)CACB, and HN(CA)CO [19, 20]. The examples and figures in this chapter use sample spectra from the Ubiquitin NMR Resource (*see Note 1* for information on obtaining these files).

3 Methods

3.1 Processing the 2D HSQC Spectrum

While we are not going to use the HSQC as our primary reference experiment during the assignment process, it is a standard experiment and often useful to have so we will start describing the processing protocol using it as an example.

1. Startup NMRfX Processor. On macOS and Windows this is done by double-clicking on the icon for the installed program (or choosing it from the Start menu on Windows). On Linux, you may need to open a terminal window and type in the name of the program: `NMRfXProcessor` (assuming it is installed on your operating systems PATH).
2. Open up the $^1\text{H}, ^{15}\text{N}$ HSQC experiment (the sample experiment folder is named *15Nhsqc.fid*). You can open a file by simply dragging the file onto the graphics display window of NMRfX Processor or by using the file browser that appears after clicking the Open button in the toolbar. The actual raw NMR data is contained in a file named *fid* (Agilent and Bruker 1D) or *ser* (Bruker 2D or higher). If you are dragging the file onto the spectrum display window, you can drag either this file, or the folder that contains it. If you are opening the file with the file browser, you must navigate to and open the *fid* or *ser* file itself. Once the file is opened you will see the first (or only for 1D files) FID of the file displayed in the graphics window, and a separate Processor window will appear (Fig. 1).
3. Set up a processing script (*see Note 2*) for the file. This can be done by manually adding a sequence of processing operations for each dimension, but for common experiments this can be done by choosing **Auto Generate** from the Scripts menu in the Processor window. This command will examine the parameter files associated with the dataset and create a script containing an appropriate sequence of processing operations. The processing operations for the first dimension will be applied and the graphics window will now display a spectrum of the first row of data (Fig. 2). The view of the spectrum can be altered with the tools in the toolbar at the top of the spectrum window (*see Note 3*). The operations for the first dimension will be displayed in the Operations tab of the Processor Window (Fig. 3). We will adjust the operations after we have got the spectrum close to being in phase.
4. Display the phase controls by clicking the **Phasing** checkbox at the lower right corner of the graphics window. Adjust the zero order (PH0) slider to bring at least one peak (preferably near one side of the spectrum) into phase so that the peak has a symmetrical shape with positive intensity for the entire peak. The phase sliders have a limited range, but if you reach the end

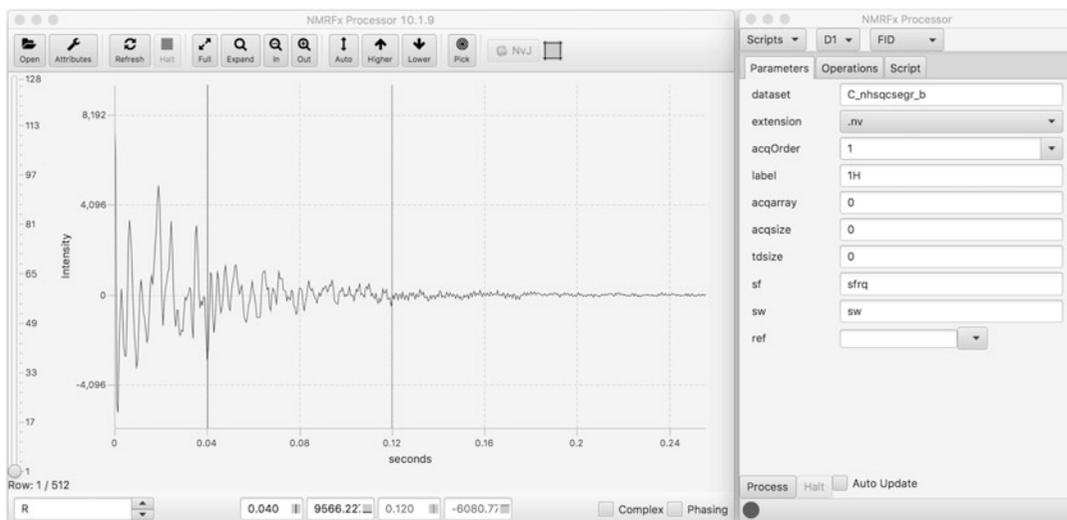


Fig. 1 NMRfX Processor showing the main spectrum window and the Processor window. The first FID of the ^1H , ^{15}N -HSQC experiment is shown. The spinbox at *bottom left* indicates that the first (hyper-complex Real) vector corresponding to the first time-increment is shown. The slider at *left* allows the user to scroll through time increments. The toolbar at *top* allows access to opening files, the Attributes Dialog, Processor control window, PDF file export, and interactive adjustment of the view. The Processor control window at *right* allows for setting up processing scripts, processing multiple files at once (the Scanner), and has a console with access to interactive Python commands

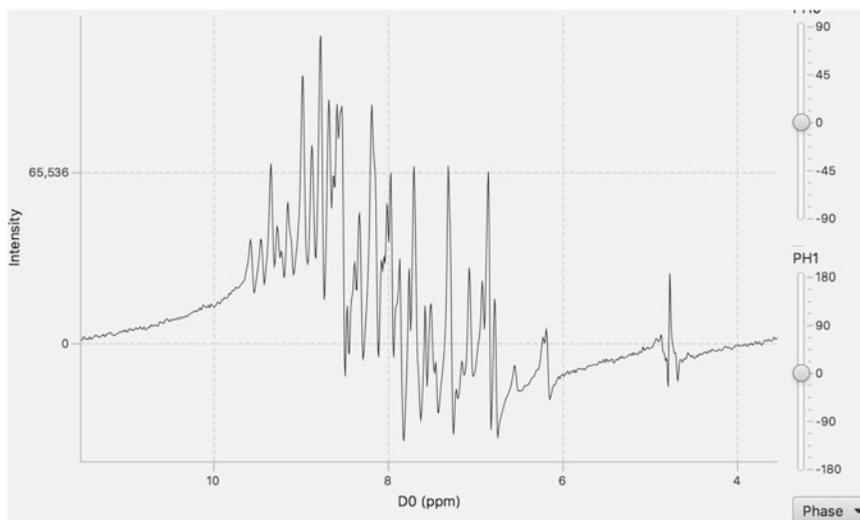


Fig. 2 The first time increment of the ^1H , ^{15}N -HSQC dataset after adding SB (sine bell apodization), ZF (zero filling), and FT (Fourier Transform) operations, but without phasing. The phasing controls have been displayed and are ready to be used to adjust the spectrum phases

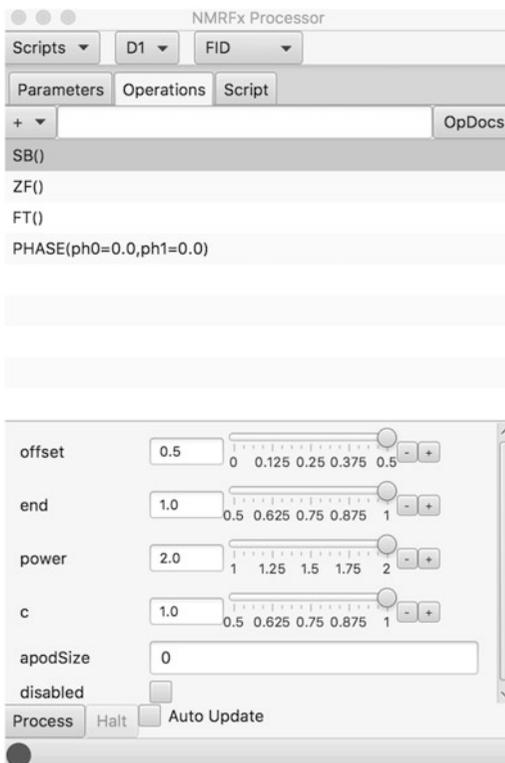


Fig. 3 The Processor window showing the operations used in Fig. 2. The **SB** operation is currently selected so the parameters appropriate to that command are adjustable using the controls at *bottom*. Slider controls can have their range adjusted with the +/- buttons at *right*. Documentation for each operation can be displayed with the **OpDocs** button and information for each parameter can be displayed by letting the mouse hover over the parameter name

and release the mouse button the range will update to a new range centered around the current value. If the entire spectrum is in phase, there is no need for further adjustment. However, some spectra will require adjustment of the first-order phase value which adds a phase correction that varies linearly across the spectrum. Position a vertical cross hair (click the left mouse button if one is not displayed at the position of a peak that is in phase). Now set a pivot point by choosing **Set Pivot** from the **Phase** menu (near lower right). You can now adjust the PH1 slider to bring the remaining peaks in phase, without altering the phase at the position of the pivot. Note that whenever you release the mouse button on the phasing sliders the new phases will update the **PHASE** operation in the operation list.

5. It will be important for the subsequent analysis of data that all axes are properly labeled and referenced. Switch to the Parameter tab of the Processor window and set the label field to *HN*.

The exact name is less important than that you use a consistent naming scheme across all spectra. Ideally, the reference position is set by careful calibration of the instrument or use of an internal standard such as DSS [21]. If the parameter files contain correct information the ref field can be left empty. The ubiquitin sample spectra used here needs adjusting. We will use the less desirable protocol of using the position of the water as a reference marker. Place the cross hair at the position of the water (the distorted signal at the right side) in the spectrum. From the menu to the right of the ref field choose **PPM at Crosshair** and select **H2O**. In this example, the ref field will now be populated with a value of 7.530 (the approximate chemical shift at the center of the spectrum).

6. A common method of data collection of data involving an HSQC module allows preservation of two orthogonal components of magnetization and hence leads to sensitivity-enhanced spectra. These PEP (preservation of equivalent pathways) [19] methods typically require formation of a new pair of data rows from the co-addition and co-subtraction of two original pairs of data rows. This technique requires use of a processing operation to combine the data rows in the appropriate manner (*see Note 4* about using the spectrum display to see if you need this). Many modern datasets will have parameters set that will allow NMRFX processor to add the operation automatically when the processing script is initially generated (*see Note 5*). If the operation needs to be added explicitly (as with many of the ubiquitin sample spectra demonstrated here) switch to the Operations tab, click the + menu, and choose **TDCOMB** (for time domain data combination) from the **FID** submenu. Set the **coef** parameter to **echo-antiecho** (a common name for this type of data collection). When you set the parameter, you should see the spectrum double in height as the first data row is now the subtraction of the original first row and an anti-phase second row (*see Note 4*).
7. The default sine bell apodization (SB), zero-filling (ZF), and Fourier Transform (FT) operations have reasonable default parameters for the sample datasets. You may wish to adjust them for some spectra. For example, using smaller values of the offset or power parameters in the SB command can be considered to trade off spectrum quality in terms of resolution, sensitivity, and truncation wiggles.
8. Some of the spectra may have significant residual water signals whose broad, and possibly out-of-phase, edge could extend into the area of the amide protons. Adding a computational solvent suppression operation may improve the spectral quality. This HSQC experiment may not require adding this solvent suppression, but we will use it to illustrate the protocol.

Choose **TDSS** from the operation menus **TD-Solvent** sub-menu. By default, this will suppress the signal at the center of the spectrum, but this HSQC spectrum is collected with the carrier away from the water position, so we need to adjust the suppression position. Adjust the **TDSS shift** parameter with the slider to move the suppression position so that you see the water signal reduced in intensity. The position is expressed as a fractional value with 0.0 at the center (a value of about 0.339 is appropriate for the sample HSQC spectrum).

9. The backbone assignment experiments used here all involve direct detection via the amide proton so there are seldom any relevant signals upfield of the water and you may wish to discard regions of the spectrum to minimize the size of the processed file. The sample HSQC spectrum already is collected with the carrier in the center of the amides, and thereby is reasonably efficient in space use, but we can (as an example) reduce it even more. Position the black crosshair at about 10.5 ppm and the red crosshair (*see Note 6*) at about 5.5 ppm. Right click the mouse button in the spectrum area to display a pop-up menu and choose **Add Extract Region**. A new operation will be added to the operation list that will extract the specified region during processing.
10. Now adjust parameters for the second (^{15}N) dimension. Use the second menu at the top of the Processor window to change the operation dimension from D1 to D2. The operation list will now be updated to show the operations appropriate to the second dimension. **NMRfx Processor** will load data values from the file that correspond to the first time-domain signal in the second dimension, apply the processing operations, and display the resulting spectrum. This works in the sample HSQC spectrum, but it may not load the correct data for all data types (*see Note 7*). The operations used here (**SB**, **ZF**, **FT**, **PHASE**) are similar to that in the first dimension, but some parameters may differ. For example, the **SB** operation has a *c* value of 0.5. This indicates that the first point of the time-domain signal should be multiplied by 0.5. This is commonly used in the indirect dimensions to account for an artifact related to the use of the discrete Fourier Transform [22]. You can interactively adjust the *c* value and see that when the value is greater than 0.5 the spectrum has a shift of the baseline above the zero point.
11. Adjust the label and referencing information for the second dimension with the controls in the **Parameters** tab. The values derived from the parameter file for the sample HSQC in this dimension are not correct (*see Note 8*), so you need to change the **label** to *N*, the **sf** value to *sfreq2*, and the **ref** value to *N* (*see Note 9*).

12. The edges of the first row ^{15}N spectrum have artifacts that can lead to streaks in the final dataset. We can treat these in the same manner as the solvent suppression used in the first dimension. Just add a TDSS operation and slide the shift all the way to the edge (0.5).
13. Now we can apply the full processing to this dataset (*see Note 10*). Just click the **Process** button near the lower left corner of the Processor window. When processing is done (it only takes about 200 ms for this dataset), you will see the processed dataset replace the one-dimensional spectrum in the graphics window. You may need to adjust the vertical height (Auto, Higher, and Lower icons in the toolbar) to see the crosspeaks clearly without displaying significant noise signals.
14. Now the data looks like a good HSQC spectrum with one major exception. ^1H , ^{15}N HSQC spectra often have a pattern of a subtle slope in the number of peaks from lower left to the upper right. This spectrum is opposite that, suggesting that the N-dimension needs to be reversed. We can do this in two different ways. Changing the sign of imaginary values (of the complex NMR data points) before the Fourier Transform is done, will reverse the spectrum. To do this go to the Operations tab, select the **FT** operation, and turn on **negateImag** in the property region below the operation list. Alternatively, you can add an explicit **REVERSE** operation (from the **Transform** submenu). After doing one of these choices, click the **Process** button again to apply the new script.
15. In **step 4**, above, we interactively set the phase values for the first dimension while looking at the first row of the spectrum, but we may need to adjust them further by examining the fully processed spectrum. This is particularly true for spectra where low intensity makes it hard to see signals in the first row. Also, it may be difficult to phase the indirect dimensions before processing. It is useful, when looking at the phasing of the spectrum, to have the display of negative contours turned on. This is the default state, but if they are not on, this can be activated in the Attributes dialog. Click the on checkbox in the **Negative** region of the **File** tab and then click the **Refresh** button in the toolbar. Checking the phases of the processed spectrum is normally done by observing the phase of slices through the dataset, adjusting the phase with the phasing tool, and then reprocessing the data. Alternatively, we can use an autophase operation in the script, as described in the next step. To visually examine and adjust the phasing do the following:
 - (a) Turn on the Slices checkbox at the bottom of the spectrum window. Now, when you move the crosshairs the software will display a slice through the dataset at the

position of the black crosshairs (note: if the **Phasing** checkbox is active you will only see one crosshair as described in **step b**, below). The horizontal slice represents a vector parallel to the first dimension of the dataset, and the vertical slice represents a slice parallel to the second dimension of the dataset. You can adjust the color, vertical scale, and display offset position of slices in the **Slice** tab of the **Attributes** window.

- (b) Turn on the phasing tool with the **Phasing** checkbox. Only the horizontal (first dimension) slice will remain indicating that it is the dimension being phased. Position the crosshair to show a slice through a well-resolved peak that is located to one side or the other of the spectrum (Fig. 4). Make sure both the horizontal and vertical crosshairs intersect the peak. The vertical crosshair will be used

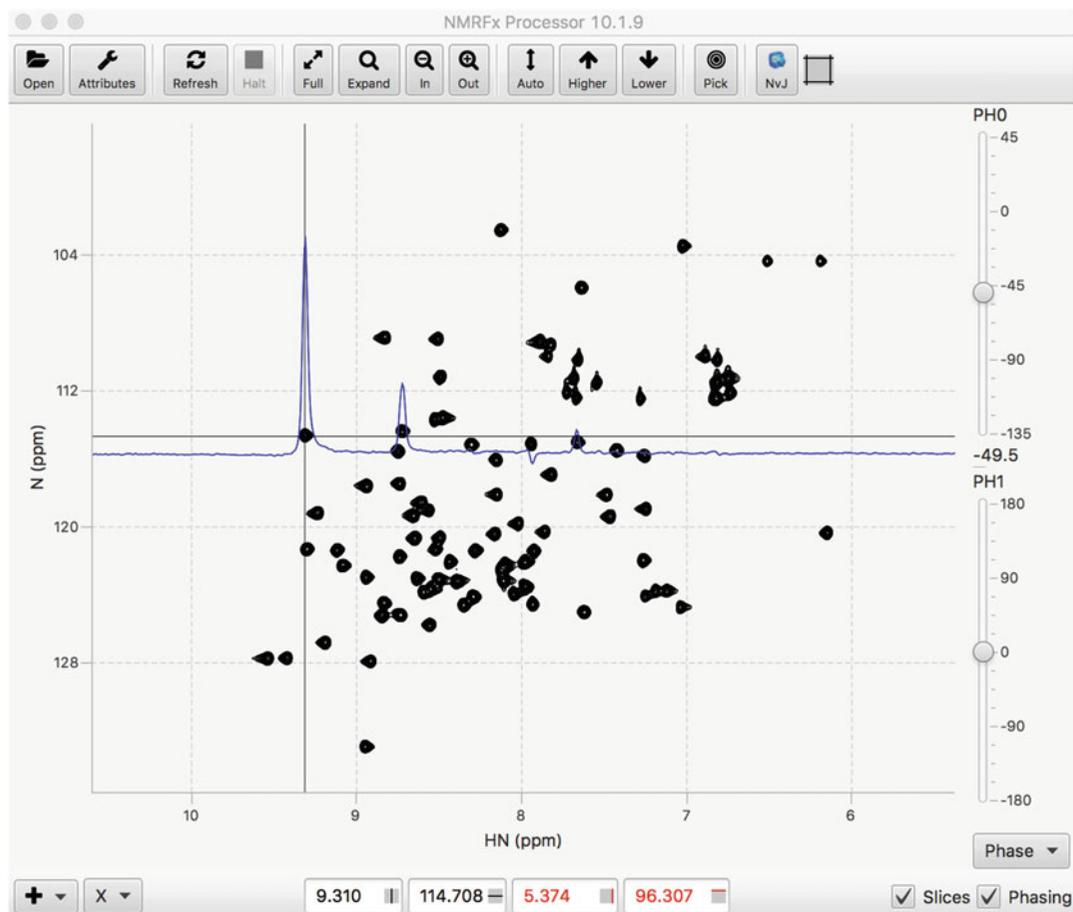


Fig. 4 The 2D contour plot of the ^1H , ^{15}N -HSQC spectrum with slices turned on. Since the Phasing controls are activated only the horizontal slice is shown. The crosshairs have been positioned on a peak near the *left edge*, an appropriate place to set the pivot position

to set the location of the pivot (if first-order phasing is needed). Phasing is done the same as with **step 6** above (setting pivot, and adjusting the PH0 and PH1 sliders), but you can (after setting the pivot) reposition the slice to intersect other displayed peaks (typically you will set the pivot on a peak at one side, adjust PH0, then move so you get a slice intersecting a peak near the other side, and adjust PH1) (*see Note 11* for idiosyncrasies of phasing this HSQC Spectrum). Once you have phase values set, click the **Process** button to reprocess the dataset with the new phase values.

- (c) Now change the operation dimension from D1 to D2. If the **Slice** and **Phasing** checkboxes are selected, you will now see a single vertical slice for dimension 2 (N). Adjust the phases as above, but note that you now set the pivot position with the horizontal crosshair. This particular dataset requires minimal phasing in the second dimension. Be aware that some experiments require particular phase values in the indirect dimensions. For example, experiments whose initial indirect delay time is equal to one-half the dwell time may require a -90° zero order phase value and a 180° first-order phase value (the **Phase** menu has options to set phases to $-90,180$). Once the second dimension phase is set, reprocess the dataset again by clicking the **Process** button.
16. It is also possible to automatically set the phases within the processing script. This is done with an operation that examines the fully processed spectrum, calculates new phase values, and rephases (without reprocessing) the spectrum using the calculated values. To do this, first change the Dimension selector from **D1** or **D2** to **D_ALL**. Operations entered here act on the whole dataset at once, rather than on individual vectors. Add a **DPHASE** operation from the Dataset menu. The **DPHASE** operation has a variety of parameters. The two most important are **dim** and **firstOrder**. The **dim** parameter specifies which dimension of the dataset should be phased. The default value of 0 automatically phases all dimensions. If the **firstOrder** parameter is off (the default) only the zero-order phase value will be calculated. The **DPHASE** operation may not give optimal results with poor signal to noise spectra or if there are significant regions of artifacts so it is good to compare the spectrum with and without the use of this command. The **DPHASE** operation works after the dataset is initially processed so any **PHASE** commands included in the operations for each dimension are applied first. It is appropriate to set approximate phases in the **PHASE** command, especially if a dimension requires specific phasing such as the $-90,180$ mode mentioned above.

17. Some spectra may benefit from correcting the baseline to fix distortions that are present. Baseline correction can be done either during the normal operations of each dimension, or after the spectrum is fully processed. Correction during the normal processing of each dimension is most useful when there are clear regions that do not contain signals for all vectors of the dataset and you manually define them. Automatic identification of baseline regions is most useful when you do it after full processing so that there are fewer peaks in any given vector. Because the baseline region identification can be done in automatic or manual mode and there are multiple ways to actually correct the baseline, the baseline correction is divided into two operations: region identification and the actual baseline correction.
 - (a) To manually correct the spectrum set up baseline regions. This can be done when initially setting up the processing (around **steps 7 and 8** above) or after the full processing. In either case, we will be adding new operations so the processing will have to be executed again. To do this while viewing the first processed row, make sure that the selector at top right of Processor window is set to FID (this displays the FID with any processing operations applied).
 - i. Choose the desired dimension (D1, D2...) with the dimension selector.
 - ii. Place the black and red vertical crosshairs on the edges of a region without any signals.
 - iii. Right click with the mouse and choose **Add Baseline Region** from the menu. You will see that region of the spectrum displayed with an orange overlay and a new **REGION** operation appearing in the operation list.
 - iv. Repeat this to add any additional regions (*see Fig. 5*).
 - v. Now add a new baseline correction operation. The two most useful are **BCPOLY**, which fits a polynomial curve of specified order to the baseline regions, and **BCWHIT**, which fits a smoothed line to the baseline regions. When you add this operation, you will immediately see the baseline correction applied to this row of data.
 - vi. Click Process now to reprocess the whole dataset (or after setting additional operations if you are doing this around **steps 7 and 8** above).
 - vii. You can repeat this for additional dimensions.

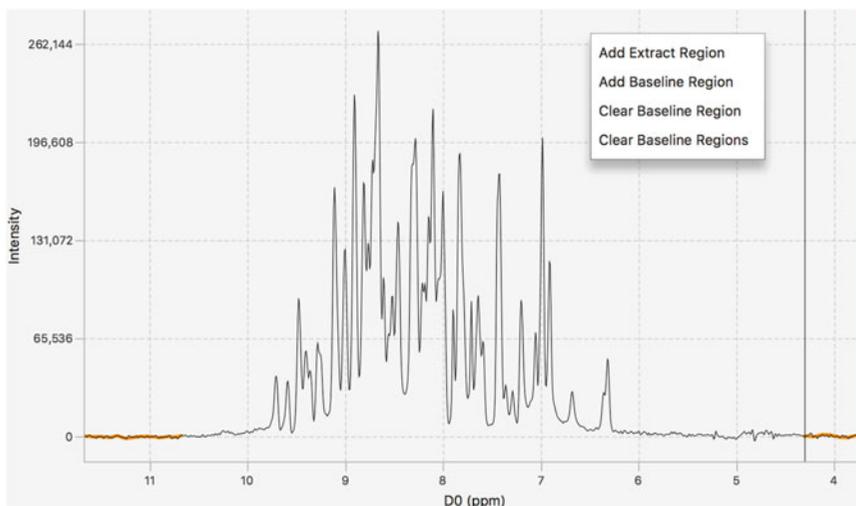


Fig. 5 The phased, first row of the ^1H , ^{15}N -HSQC spectrum being set up for baseline correction. Three baseline regions have been manually defined and are shown with an *orange overlay*. The crosshairs are in the position where the pop-up menu was used to add the last baseline region

- (b) Use the automated method as follows
 - i. Choose the desired postprocessing dimension such as P1 or P2 (these operations will be applied after the previous dimension operations are done so typically apply after the data is Fourier transformed).
 - ii. Add an AUTOREGIONS command from the operations Baseline sub-menu.
 - iii. Add a baseline correction operation (as in **step v** above).
 - iv. You can repeat this to add these operations to additional dimensions.
 - v. Process the whole dataset.
18. You should now have a fully processed HSQC spectrum with good phases and baselines. The processed dataset will, by default, be placed in the same folder as the original FID file. You may wish to copy (using operating system commands) the processed spectrum into a separate folder where you will accumulate all the datasets used here.
19. The actual processing script (built of reference commands and processing operations) can be seen in the Script tab. This script is saved as a file named process.py in the folder containing the FID file each time processing is completed without an error. You may wish to copy this file to a new name in that folder, or to some other folder in case you accidentally overwrite the process.py file. You can save and open processing scripts using the **Scripts** menu at top left of Processor. Here is the processing script generated above for the HSQC dataset.

```

import os
from pyproc import *
FID('. /fidfiles/15Nhsqc.fid')
CREATE('. /fidfiles/15Nhsqc.fid/C_nhsqcsegr_b.nv')
acqOrder('1')
skip(0,0)
label('HN','N')
acqarray(0,0)
acqsize(0,0)
tdsize(0,0)
sf('sfrq','dfrq2')
sw('sw','sw1')
ref(7.5302,'N')
DIM(1)
TDCOMB(coef='echo-antiecho')
TDSS(shift='0.339f')
SB()
ZF()
FT()
PHASE(ph0=-46.8,ph1=-3.8,dimag=False)
EXTRACT(start=132,end=772,mode='region',disabled=True)
DIM(2)
TDSS(shift='0.5f')
SB(c=0.5)
ZF()
FT(negateImag=True)
PHASE(ph0=0.0,ph1=0.0,dimag=False)
REGIONS(regions=[0.004,0.051,0.867,0.996])
BCPOLY()
DIM()
DPHASE(dim=1,firstOrder=True)
DIM(1)
AUTOREGIONS()
BCWHIT()
run()

```

20. Setting up processing scripts is most easily done in the graphical display of NMRfX Processor. But the generated scripts (or manually written ones) can be executed from the command line. This is especially useful if you wish to make small changes in the scripts and reprocess them or apply the processing to a new dataset using a script developed with a previous dataset. You can apply the processing to a new dataset (with compatible parameters) by simply changing the files referenced in the FID (the location of the raw FID) and CREATE (the location of the

processed spectrum output) operations. To process from the command line, do the following.

- (a) Locate the `nvjp` (`nvjp.bat` for windows) shell script (*see Note 12*).
- (b) Execute the shell script with the processing script as an argument.

```
nvjp process.py
```

3.2 Processing the 3D Backbone Spectra

Processing the 3D spectra proceeds in a manner similar to that described above for the 2D HSQC spectrum, but requires the script have operations for all three dimensions, and some additional steps in checking the phasing. In the following steps, we will abbreviate those that are essentially the same as described above. We should note that there is growing use of nonuniform sampling in NMR data collection [23–26]. The sample data collected here have been collected with conventional, uniform sampling, and so will not require any special processing steps. But NMRFX Processor supports several processing schemes for NUS type data (*see Note 13*) and generally only requires the addition of one operation to the processing scheme.

1. Open the HNCO experiment (the sample experiment folder is named *hmco3d.fid*). This will be used as a reference experiment for the backbone assignment described below (you can use the HSQC spectrum, but the third dimension of the HNCO allows resolution of more peaks in overlapped spectra). Because of this, use particular care in setting up this experiments referencing etc.
2. Set up a processing script using the **Auto Generate** command from the **Scripts** menu. Correct and add any additional operations.
 - (a) This dataset needs a **TDCOMB** operation with the **coef** parameter set to *echo-antiecho*. Note that the **TDCOMB** operation has a **dim** parameter that must be set correctly for 3D (and higher) datasets. The PEP technique is used in the nitrogen dimension of this spectrum. Since this is dimension 3, the **dim** parameter should be set to 3.
 - (b) Set the reference using the crosshair on water and using the **PPM at Crosshair - H2O** menu item in the ref section of the Parameters tab.
 - (c) Add a **TDSS** operation to remove residual water.
 - (d) Phase the first dimension.
 - (e) Add an **EXTRACT** operation (using cursors at left edge, and at 5.5 ppm).

3. Set Labels for the first, second, and third dimensions to *HN*, *C*, and *N*.
4. Now click the Process button to initiate processing of this dataset. When processing is complete (in about 3 s) you will see a contour plot with the first dataset dimension on the *x*-axis and the second dimension on the *y*-axis. By default, this will be a plot of the center plane of the 3D spectrum. Display the **Spectrum Attributes** window (**Attributes** button on toolbar) and switch to the View tab. Here, you will see the plot limits and dataset dimensions for the *X* and *Y* dimensions, and the indices of the displayed planes for the *Z*-axis (64 is the center of the sample HNC0 dataset). You can move up and down planes in the spectrum by using the up/down arrow keys on your keyboard (after clicking once with the mouse in the spectrum to make sure it is the target of the key presses). Display all the planes by clicking the **Z** menu item and choose **Full** and then click the **Refresh** button in the toolbar. You can show the third (*N*) dimension of the dataset on the *y*-axis by clicking the numerical menu at far right of the **Y** row and choosing 3. (Note: there are a variety of key shortcuts for navigating through the spectrum, see full documentation).
5. Check the phasing in all three dimensions. This can be done after processing all dimensions as described in this step, or by only processing two dimensions as described in the next step. Typically, you will want to be on a single plane (rather than the full range of planes we set above). Checking phases in the three-dimensional experiment works the same as with the two-dimensional experiment as described in **step 15** above with one important caveat. To check the phases for the third dimension you need to set the view to display the third dimension on the *y*-axis (as described in the previous step) and make sure the operations dimension menu (top of the Processor window) is set to **D3**. You can set the zero-order phase value and pivot on a peak on one plane, and scan through the dataset to find a slice through a peak at the opposite side of spectrum to observe when the first-order phase value is set.
6. An alternative way to set the phases is to process the dataset in only two dimensions at a time (^1H - ^{13}C and ^1H - ^{15}N). This allows you to observe the peaks in a single plane, rather than having to step through them to find appropriate planes for setting the zero and first-order values. The **Parameters** tab display for each dimension has a setting labeled **skip**. If you turn this on, that dimension will be skipped in the processing. Try this by going to the third dimension (**D3** menu setting) and turn on skip. Now process. You will get a 2D display with all the peaks. You may need to click the full button and adjust the contour levels to see the peaks. Adjust the phasing just as

you would with a 2D dataset. Now turn off the skip value for **D3** and turn it on for **D2**. Reprocess and check the phases for the third dimension (which will be displayed on the *y*-axis). Be sure the dimension selector is set to **D3** when you adjust the *y*-axis phases.

7. You can also adjust the phases automatically using the **DPHASE** operation in the **D_ALL** dimension. You can use one **DPHASE** operation with the **dim** parameter set to 0, or multiple **DPHASE** operations with the **dim** parameter set to 1, 2, or 3. The latter technique would allow you to autophase only a subset of the dimensions.
8. Repeat the above steps for each of the remaining backbone experiments. Tested processing scripts are available for download at the software website: <http://www.nmrfx.org> (login may be necessary to see the scripts). You can load a processing script by using the **Open...** item in the Scripts menu. When you load a script the FID and CREATE commands will automatically be updated to refer to the currently open data file.
9. After processing the experimental data, copy the resulting datasets to a common folder for easy access (as in **step 18** in Subheading 3.1).

3.3 Setup an NMRViewJ Project

Having completed the processing of the datasets we can now switch to visualization and analysis using NMRViewJ. When doing anything more than simple display of spectra, it is a good idea to set up an NMRViewJ project. The project stores related information in a set of files within a hierarchy of folders. All molecular information (protein or nucleic acid sequence), peak lists, chemical shift assignments, relaxation data etc. are stored in a BMRB (BioMagResBank) STAR file [27]. The state of open spectrum windows and saved spectrum windows are stored in a series of canvas files (an NMRViewJ specific format). The state of the project should be regularly saved as one works on it. A history feature in the Project Browser allows the user to return to the exact state of the project at any saved point.

1. Start up NMRViewJ following the same instructions as given above for NMRfX Processor (but using the appropriate executable program name).
2. Create the new project by choosing the **File>Projects>New** menu item (*see Note 14*). A file browser will be displayed so that you can choose a location and name for the project. Navigate to the desired location for the project, enter a name for the project, and click **Save**. A new project folder will be created with a set of subfolders for storing the project information.

3. The spectrum files that were processed in the above steps can be copied directly to the *datasets* folder within the project, or you can add a link in the project to a different folder. To do this, choose **Attributes** from the **File>Projects** menu to display an attributes window. Click the folder icon in this window and choose any existing folder of datasets outside the project folder (for example, the one you copied files to after processing in previous steps). Whenever you open up this project all the files in the projects own datasets folder, and in any folders you explicitly added, will be loaded.
4. You can load datasets manually (from the Datasets menu) or tell NMRViewJ to load all the project datasets (added in the previous step) right now by clicking the **Load** button in the **Project Attributes** window. You can now close the Project Attributes window.
5. Check to see what datasets have been opened by displaying the Datasets Table by selecting the **Datasets>Datasets Table** menu item. This will list all the datasets and show their location on disk (their file path). By choosing various Modes you can also display appearance and reference information.
6. Get in the habit of regularly saving the state of the project. We have created a project and added datasets so now is a good time to save. Just choose the **File>Projects>Save** menu item.

3.4 Check and Configure Datasets

Before proceeding with any assignment tools it is a good idea to open and view all the datasets and check the reference information. We can view them individually, or as a whole group. Here, we will display all the datasets together in a single window.

1. Display the Datasets Table (**Datasets>Datasets Table** menu item).
2. Check the dataset labels (*see Note 15*) by setting the Dataset Table Mode selector to Referencing-1. The Label field for all the datasets should be *HN*. Now repeat this for the second dimension. The labels for the two experiments that detect carbonyls (hnc0 and hncaco) should be *C* and those for the other 3D experiments should be *CA*, and the hsqc should be *N*. Now check the third dimension. All the 3D experiments should have a label of *N*. If you need to change labels, just double-click in the entry in the Label column of the row for the particular datasets you wish to change. Type in a new label, and choose (while that table row is still selected) **File>WritePar** from the **Dataset Table** menu. This latter command will write out a parameter file for this dataset that contains referencing information (including the new label value). By saving to a parameter file the reference values will be persisted between NMRViewJ sessions.

3. Now display all the 3D experiments in one window. Select all the 3D experiments in the table (click on the top experiment and then shift click on the last 3D experiment). Click the Draw button and change the title field (this is optional) to something like “HNall3D.” Leave *N Spectra* at 6 and *N Rows* at 2 (these can be used to determine how the spectra are arranged within the window that will appear. Now click **Create**. A new window will appear with a 2×3 arrangement of spectra (*see Note 16* about grid arrangements). We would like to see the HN axis on the *X*-axis and *N*-axis on the *Y*-axis. To do this we will use a keyboard shortcut in the spectrum window (you can do this from the Attributes window, but the shortcut is faster to do). Click the mouse in any spectrum window (so it is active) and type “VyN” and hit the Return key. This is short for View command, set *y*-axis to display the **N** dimension of the dataset. Use of an upper case (rather than lower case) “V” implies that this command should apply to all six spectra in the window rather than just the active window.
4. Adjust the vertical scale so there is a good display of peaks with a minimum amount of noise. You can use the up/down arrows in the toolbar to adjust the contour level. If you hold down the Shift key while clicking the contour level, the change will apply to all six spectra. Otherwise, it will apply to the active spectrum (the spectrum you last clicked a mouse in).
5. Now we will fine-tune the referencing so all spectra are referenced the same. If you have processed with the sample scripts (or their reference values), you will find that the referencing should be very good, so consider this an example of what to do if it is not. Let us assume that we have the reference correct for the HNC0 experiment and align the remaining spectra with it. Choose a smaller region containing a few peaks with the black and red crosshairs and click the Expand icon in the tool bar. Now choose a peak that is recognizable in all spectra and align the black crosshairs on it in the HNC0 spectrum (Fig. 6). Now in each of the other spectra:
 - (a) Align the red crosshairs on the corresponding peak.
 - (b) Right click with the mouse to bring up a menu.
 - (c) Choose **Reference>Shift Spectrum**.
 - (d) Answer *yes* to the question that appears about saving parameter files (*see Note 17*).
 - (e) Scan the black crosshair around various peaks and make sure they are well aligned in all spectra.
6. Zoom the view out so that we can see the entire region of each spectrum by clicking the **Full** icon in the toolbar. We might want to refer to this grid of windows periodically during our analysis, so let us save it to an NMRViewJ canvas file in the

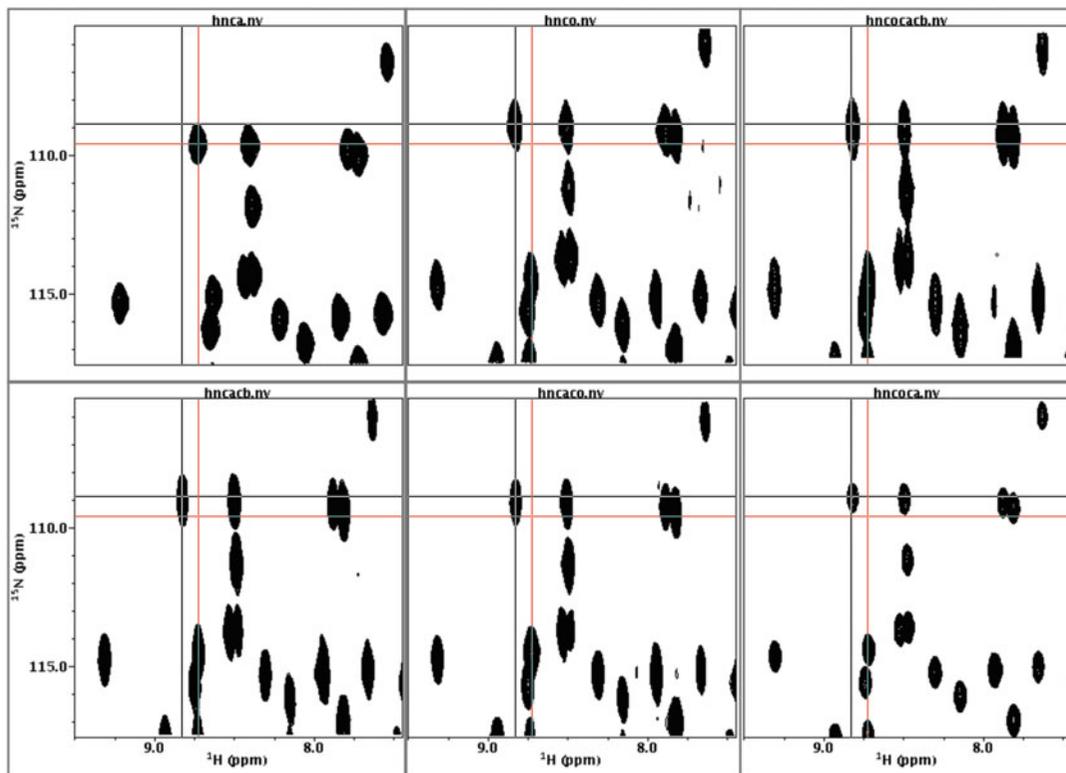


Fig. 6 A grid of spectra showing the triple-resonance backbone experiments with the ^1H dimension on the X-axis and the ^{15}N dimension on the Y-axis. This view is being used to adjust the referencing of these two dimensions (the spectra were deliberately mis-referenced for illustration). The *black crosshair* has been centered on a peak in the hnco experiment and the *red crosshair* centered on the corresponding peak in the hnca experiment. The spectrum menu **Shift>Reference** command can now be used in the hnca spectrum window so the position at the *red crosshair* is adjusted to correspond to that of the *black crosshair*, thereby bring the hnca experiment into alignment with the hnco experiment. An automated command is also available in RunAbout to align spectra based on all the peaks

project. Just click on the **Favorite** icon in the toolbar. You will be prompted for a name (which will default to the name we used when we created the top-level window, HNall3D). Now click OK to save it. You can reload a file at any time by choosing the **Favorite** menu item in the Windows menu. Just choose the desired window and click **Open**.

- Change the view of the window so that the second dimension of the dataset is on the y -axis (by typing `Vy2 Return`) in any of the spectra so we can view the ^1H - ^{13}C view of the spectra. The HNCACB experiment has positive and negative contours (CB peaks are negative), but the display only shows positive peaks. Open the Spectrum Attributes panel (click on Attributes icon on toolbar) and click once on the HNCACB spectrum to make it the active spectrum. The File tab of this window has a row for each dataset shown on the spectrum (only one is shown

at present) and a column labeled “-.” Click the checkbox in that column to turn on the display of negative contours (which will be in red as indicated by the color in the next column). Click the Refresh icon to redraw the spectrum.

8. Check referencing in the Carbon dimension as well. Here, it is good to show the spectra in pairs (like the HNCO and HNCACO). Some or all of the peaks in the inter-residue (HNCO) experiment will also show up in the intra-residue experiment. You can look for these peak pairs to check the alignment.
9. When NMRViewJ assigns a dataset to a window (as happens regularly in various tools such as the RunAbout assignment tool), it checks stored parameters for that dataset. These parameters include the default contour level, whether or not to draw positive and negative contours, and the colors used for the contour display. It is important to set up these defaults so that the spectrum display looks good without needing to manually intervene in adjusting the display each time you assign the dataset to a window. Check each of the six spectra and make sure that the contour levels look reasonable and the contour modes are set (we did this for the HNCAB here). Once you have a good display of each spectrum open the Attributes panel, and after activating each spectrum (click in the spectrum), click on the Disk Icon in the File tab. This will save the parameter file for the corresponding dataset with the chosen display attributes.

3.5 Pick Peaks in All Spectra

We now want to identify the locations of peaks in the spectrum. The algorithm used for locating peaks is quite simple but relatively robust and rapid. Peaks are considered points of local maxima (that is, any point that has a higher intensity than all adjacent points). When NMRViewJ locates peaks, it also performs the following steps: identify the peak bounds (the width of the peak at the level of the intensity threshold); estimate the half-height peak width; determine whether the peak is on the edge of the spectrum or adjacent to other peaks; calculate the center position by interpolating the intensities of the adjacent data points. When working in a graphical spectrum display the peak picker only considers points that are higher than the current contour limit (that is, what you see, is what it picks). Since the HNCO experiment will be a reference spectrum for the RunAbout analysis, we want to be particularly careful in peak picking.

1. Display the HNCO experiment in its own window so we can get a good look at it. You can do this several ways. The easiest is to display the Dataset Table, select the row for the HNCO experiment, and click Draw. Alternatively, if you have the HCa3D grid of windows displayed, click once on the HNCO

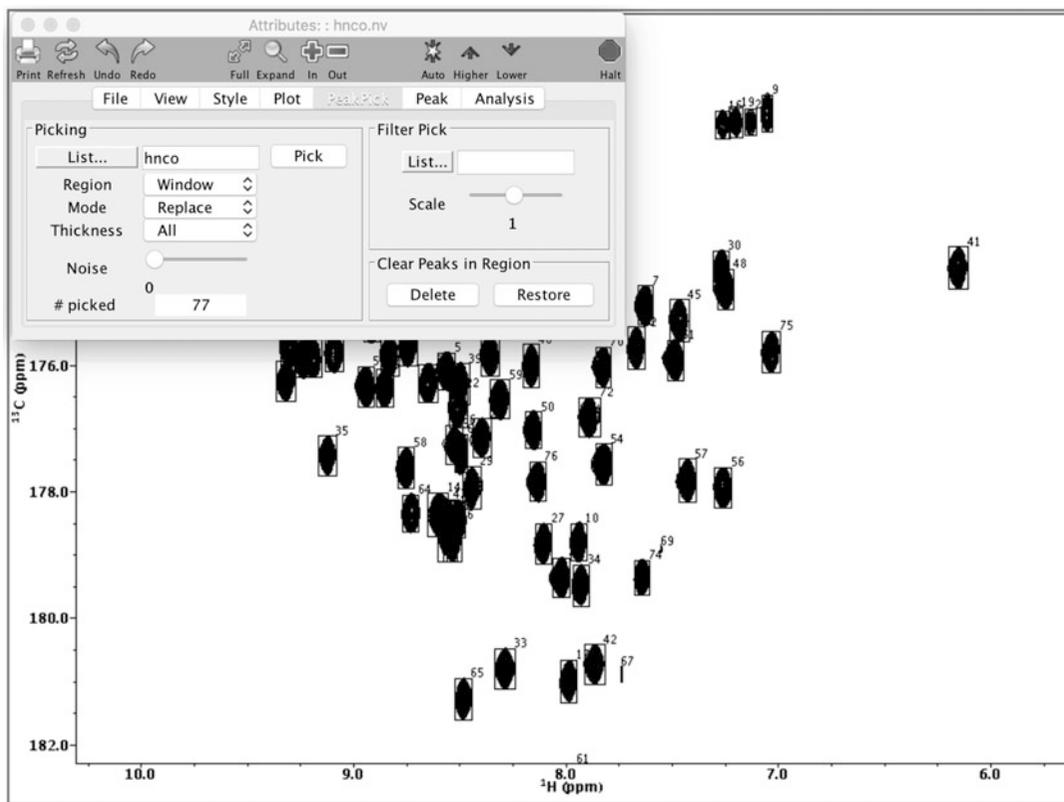


Fig. 7 The **PeakPick** tool being used to perform peak-picking of the HNC0 experiment. Controls are available to set the name of the peak list (it defaults to the dataset name), what regions are to be picked, whether a pre-existing peak list is to be replaced or appended to, how peaks are to be picked across multiple planes, and whether streaks of noise are to be avoided. *Labeled boxes* are displayed around each picked peak in the spectrum. These can be interactively resized, repositioned, or deleted

spectrum and type **sr** (a keyboard short cut for “spectrum replicate”). This will create a new spectrum window with that spectrum displayed in it.

2. Display the **PeakPick** tab of the **Attributes** window and click the **Pick** button (Fig. 7). Boxes will appear around all the peaks and the “# picked” entry will indicate the number of peaks found (for the sample HNC0 experiment, with a contour level of 1.5, you should get 77 peaks).
3. Check the peaks. You may want to delete particularly weak ones that the picker found. For example, in the displayed spectrum three peaks are quite weak and may be artifacts (numbered 61, 67, and 69). You can delete these in several ways, including:
 - (a) Change the cursor to peak delete mode (using the cursor icon in the toolbar). Now you can click the left mouse button while the cursor (the active point is the intersection of the “crossbones” in the cursor symbol) is centered over the peak.

- (b) Show a table of peaks (**Peak>Show Peak Table**), sort by intensity by clicking on the *intensity* column. Now select the three weak peaks and click the X button. The deleted peaks will remain, but be shown in red.
4. You can also systematically examine each peak and look for ones that require deletion or special attention. This technique is a powerful way to ensure that you have rapidly examined every peak in the spectrum (*see Note 18*).
 - (a) Show the Peak Inspector (**Peak>Show Peak Inspector**).
 - (b) Make sure the inspector's List value is set to the HNCO peak list.
 - (c) Go to the Peak tab of the Attributes window.
 - (d) Set **Show Mode** to *Expand* (after showing one peak in the next step switch this to *ExpandFixed* so the relative size of the window is preserved).
 - (e) Now, step through peaks in the list using the arrow buttons in the Peak Inspector. As you move to each peak you will see an expansion around that peak in the spectrum window.
 - (f) You can resize or adjust the position of the peak box using the selection cursor (*see Note 19*), or delete the peak (using the X button in the Peak Inspector).
5. Peaks deleted as above are not actually removed from memory. They are simply marked as being deleted and will not be displayed or used in any analysis. To permanently remove them go to the peak inspector and choose **Edit>Compress** or **Edit>Compress & Degap** (the latter renumbers the list so there are no gaps in the numbering after peaks are removed).
6. Remember, you should periodically Save the project state! It is a good idea to do this before any actions like deleting peaks or compressing the peak lists.
7. When examining the peak list, you might want to not only delete peaks, but add them. For example, if you go down in contour levels you will find a variety of side-chain peaks and artifacts which we will ignore here, but there is also a peak at about $^1\text{H} = 10.07$, $^{13}\text{C} = 179.4$, $^{15}\text{N} = 121.4$ ppm. This peak shows up in the HSQC spectrum and all the other 3D experiments so is likely a valid, but weak, signal. Change the cursor to PeakAdd mode, and click the top of the cursor peak shape on the peak. This will add a new peak at that position.
8. Since we have a nice spectrum with picked peaks of this HNCO dataset, let us save it as another favorite (just click the Favorite icon and give the window the name hnc).

Table 1

Peak picking results for the six backbone experiments used as examples for the assignment process. The experiments were picked at the specified threshold levels and the specified numbers of peaks were found

Dataset	Contour level	Number of peaks
HNCO	1.5	78
HNCA	0.15	160
HNCACO	0.25	156
HNCOCA	0.28	128
HNCACB	0.75	302
HNCOACB	0.586	74

9. Now repeat this for the remaining 3D spectra. Using the manual intervention for the HNCO list described above, and simple peak picking for the remaining sample 3D spectra the results shown in Table 1 were obtained.
 - (a) Display spectrum.
 - (b) Adjust contour level to values in Table.
 - (c) Save Dataset Preference with new contour level.
 - (d) Pick peaks.
 - (e) Save spectrum window (Favorite).
10. Remember, it is always a good idea to save the project whenever you do something significant like picking peaks.

3.6 Load Protein Sequence

In order to assign the spectra to specific atoms in the molecular structure, we need to define what that structure is. This is done by reading in the amino acid or nucleic acid sequence. Once loaded in, the sequence is saved in the project (within the BMRB STAR file) so this step is done just once. Sequences can be defined either with a text file format, or within a graphical input window. To use the GUI, do the following:

1. Choose **Read/Write Topology>Sequence GUI** from the **Molecule** menu.
2. Enter a name for the **Molecule** (for example, *ubiq*). It is a good idea to keep the name simple (no spaces or special characters) and short.
3. Enter a value for the **Polymer** and **CoordSet** fields. Typically, just use A (as in chain A) for simple (single chain) structures.
4. Turn on **One Letter Code**.

- Paste (into the large box) the protein sequence using the one-letter amino-acid code (spaces and line-breaks are ignored). To input the ubiquitin sequence, paste in:

```
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQ
QRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRGG
```

- Click the Create Button.

Reading the sequence from a file allows greater flexibility. For example, this mode allows you to specify the sequence number for one or more residues. The format is a simple text file with one line for each amino acid. Each line consists of the three-letter amino acid code (ala, gly, ser etc.) followed by an optional residue number. For example, the start of ubiquitin is:

```
met 1
gln
ile
phe
val
```

It is possible, with additional annotations in the text file, to define complex multi-chain structures consisting of both protein and nucleic acids and small-molecule ligands. See the full documentation for details.

Once you have read in the sequence, verify that it has been created successfully. There are several built-in tools to display the molecular structure information.

- Select the **Assign>Atoms** menu item to display the molecular structure in a table of atoms. You should see a table with one line for each atom in the structure. The Filter menu (or explicit filter entries like *.CA,CB in the field next to it) can be used to limit the display to particular categories of atoms (Protons, Carbons, Backbone Atoms, etc.). This table will ultimately be populated with the assigned chemical shifts and used to show various reference values (for example, average shifts from the BMRB [27]).
- Select the **Assign>Sequence** menu item to see a display of the amino-acid sequence in single-letter mode. This tool is also used for various plots of chemical shift indices.
- Select the **Molecule>Analysis>Viewer** menu item to activate a simple 3D molecular viewer (this may require NMRViewJ be started with a supported activation license).

3.7 Set up the RunAbout Assignment Tool

The RunAbout tool provides a combination of visualization, automation, and manual adjustment tools to aid in the assignment of backbone and side-chain resonances of proteins. These tools build

on concepts of linking related peaks into groups or clusters and then assembling them into fragments that can be assigned into the sequence as described and used in a variety of programs such as [28]. The overall protocol used here is more complex than a fully automated method, but allows the user to monitor and correct the individual steps of automation. Open up the tool (**Analysis>Run-About**) and configure a set of necessary parameters. The description below applies to the ubiquitin sample data, but different combinations of experiments can be used.

1. Click **Setup Peak Lists** . . . in the Parameters tab.
2. Use the menus at the left side of the PeakLists window to select the types of experiments used for the reference list (hnco) and the six backbone experiments (hnco, hncaco, hncacb, hncocacb, hnca, hncoca). Note that the hnco experiment is used as both the reference list, and as one of the assignment experiments. Now, use the menus in the middle of the window to assign an actual peak list name, to the peak list type (in our example the type names and peak list names are identical) (Fig. 8).
3. Each experiment type has a set of patterns that describe the nature of detected signals and it is important to get this set right in order for the analysis to work. Each experiment type (as set up in the previous step) has a predefined pattern for each dimension. You can examine, and modify, the pattern by clicking the “i” (for information) button in each row. Clicking this button will display a new dialog with a row for each dimension of the experiment. You can edit the pattern and the tolerance used for matching peaks in each dimension. The patterns consist of a residue symbol (i or j), a dot character as a separator, and an atom name symbol. The residue and atom portions can have multiple fields separated by a comma character. Examples will help to clarify this:
 - (a) i.h The amide proton of the current residue (i).
 - (b) i-1.ca The ca atom of the previous residue.
 - (c) i,j-i-1.ca,cb The ca or cb atom of the current residue, or the previous residue.

Atoms can have additional qualifiers indicating constraints on the intensity. For example, the pattern “i,j-i-1.cb{ $\$intensity < 0$ },ca{ $\$intensity > 0.0$ }” is used in the hncacb experiment and indicates that the atom can only be a cb if the intensity is negative, and can only be a ca if the intensity is positive. Peaks in this list can be observed for both the current residue (i) and the previous residue (i – 1).

- (a) The “i” (for information) symbol will have a red background if something is not correct in the setup for that peak list. Details will be shown when you display the

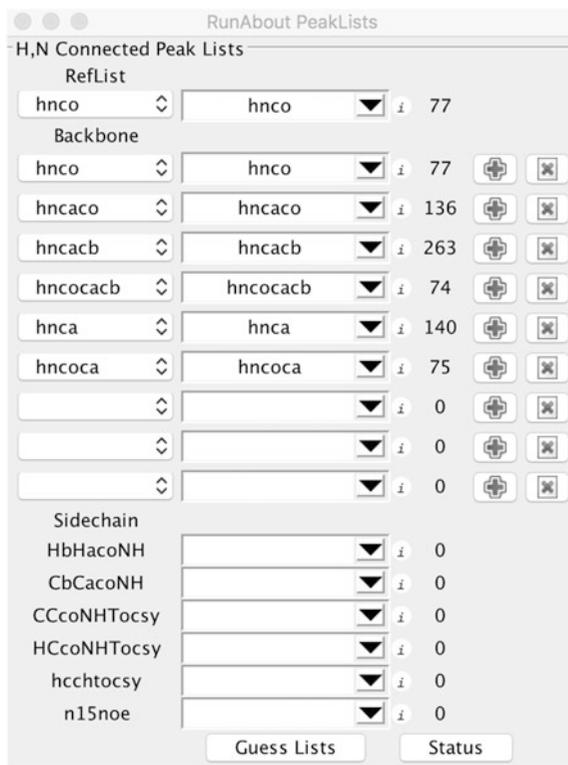


Fig. 8 The RunAbout **PeakLists** control window. This window is used to set up which of the available peaks are to be used for assignment. The actual peak lists are in the *center column*. Menus at *left* allow defining the type of experiment represented by the actual name. The “i” button at *right* displays a window with controls for adjusting the patterns that define the atom types expected for each dimension of the peak as well as tolerance used for finding overlapping peaks. This chapter focusses on backbone experiments, but the lower section of the window shows that experiments can also be configured for doing side-chain assignments

dialog for that peak list. Generally, the problem is that the tolerances (which we have not explicitly set yet are considered too small).

- (b) The number to the right of the “i” indicates the number of peaks in the spectrum.
 - (c) The “+” and “X” buttons allow you to add or remove lists after RunAbout analysis has been started (see documentation).
4. Tolerances can be set individually for each peak list or, more conveniently, for all lists at once. In the main Parameter tab is a section for setting tolerances (Fig. 9). You can either set explicit values for the H, N, and C dimensions (enter values and click

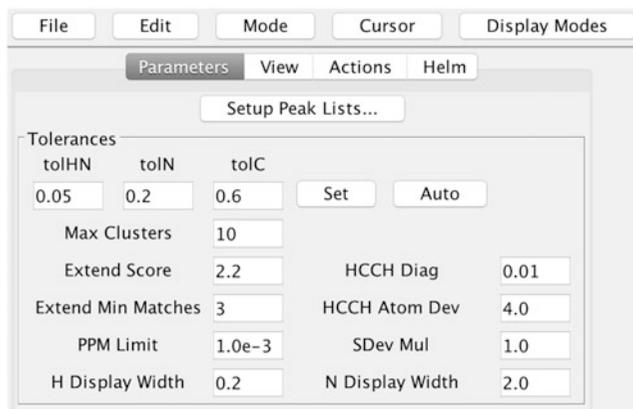


Fig. 9 The **Parameter** tab in RunAbout gives access to the PeakLists control shown in Fig. 8, and has controls for manually or automatically assigning the tolerances to all peak lists used. Various other parameters are used in operations such as automatic linking of clusters into fragments

the **Set** button) or allow RunAbout to automatically calculate them (click the **Auto** button). Automatic calculation of the tolerances is done by calculating the median line width in each dimension of each peak list and multiplying that value by a scale factor. Let us do that for the sample data. You can examine the resulting values by going back to the Peak List display and bringing up the pattern/tolerance display for each list.

- It is possible to determine the expected number of peaks for each residue by examination of the patterns. For this ubiquitin sample data, each residue (non-proline, non-glycine) should have a total of 11 peaks (including HNCO:1, HN(CA)CO:2, HN(CO)CA:1, HNCA:2, CBCACONH:1, HNCACB:4).

3.8 Examine the Peaks in a Coordinated Manner

RunAbout is set up to operate in a series of modes as you move through the process of assigning your protein. The first mode “Edit Peaks” is designed to let you rapidly examine your datasets and peak lists. In particular, before you continue with the process of assignment you want to know whether your spectra are peak picked at appropriate levels and whether they are referenced to a common value so that peaks representing the same information align as closely as possible between the different spectra.

The number of spectra displayed in the main window will depend on what datasets are available. Typically, there will be one row of spectra for each atom type (C, CA, CB). The actual atom types available are determined from the peak list patterns specified. The left half of the spectra are for those experiments that (again, according to the peak list patterns) give rise to $i - 1$ (inter-residue) connectivities (like HNCOCA, HNCOCACB, or HNCO). The right half of the spectra are for those experiments that (again,

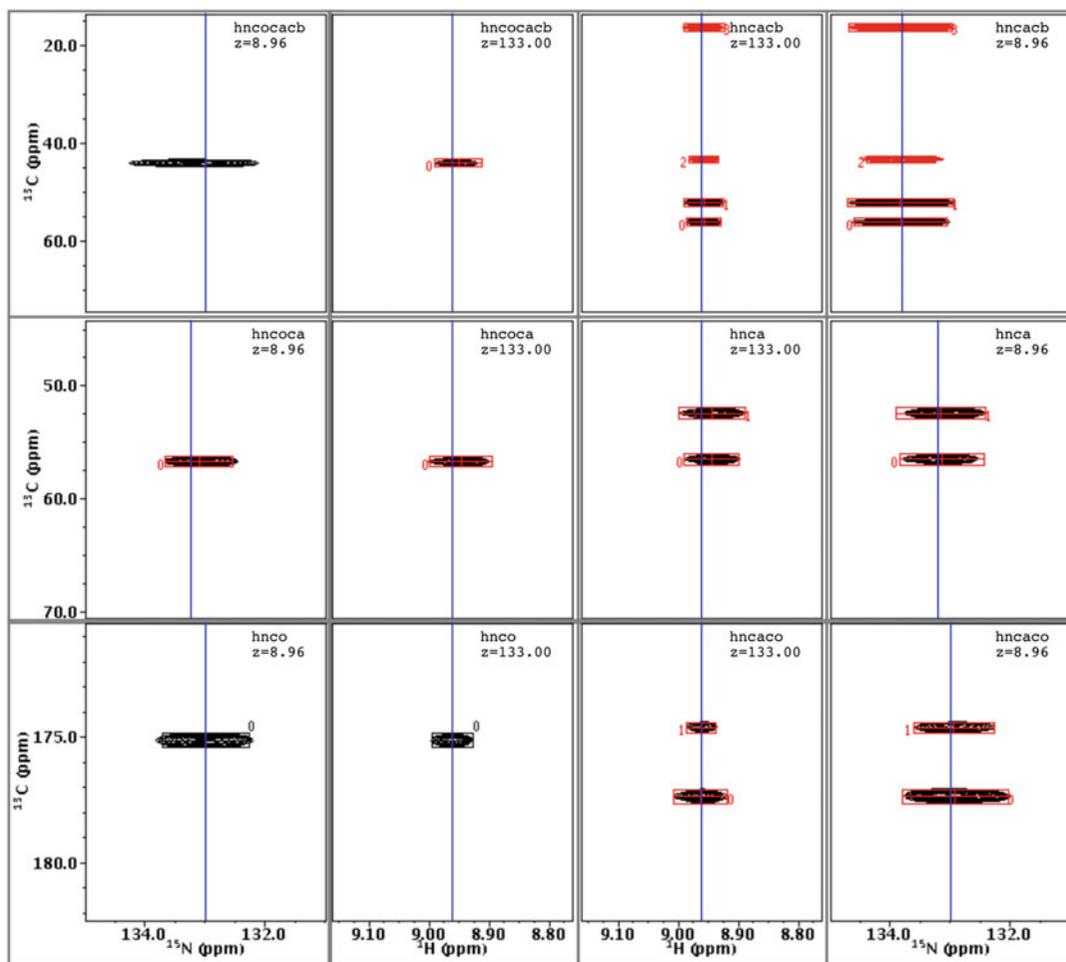


Fig. 10 The grid of spectra when RunAbout is in **Peak Edit** mode. The different spectra are labeled and each spectrum is represented in two orthogonal views. The outer pair of windows show the ^{15}N , ^{13}C axes (*centered* on the reference peaks ^{15}N shift and at the plane of the peaks ^1H shift) and the inner pair of windows show the ^1H , ^{13}C axes (*centered* on the reference peaks ^1H shift and at the plane of the peaks ^{15}N shift). Showing both views is useful for detecting nearly overlapping peaks, but the outer pairs can be turned off to increase the space available to the inner pair. The *blue vertical line* is drawn at the ^{15}N and ^1H shifts of the corresponding reference peak. Well referenced spectra should have all peaks from the same residue displayed along these lines

according to the peak list patterns) give rise to *i* (intra-residue) connectivities (like HNCA, HNCACB, or HNCACO).

1. Set the RunAbout **Mode** to *Edit Peaks*.
2. The spectrum display region (right side of window) of RunAbout will now show a grid of windows (Fig. 10). This coordinated view of all six (in this case) datasets allows you to view the regions of the datasets that have chemical shifts corresponding to the proton and nitrogen of the amide group for each residue.

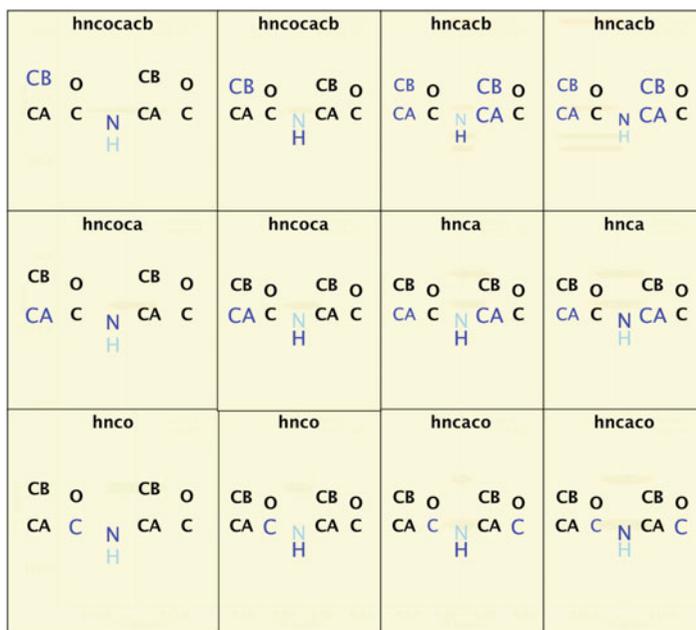


Fig. 11 The RunAbout spectrum display with labels displayed. These symbolically show what dataset and atom types will be shown in each spectrum. The labels can be turned on and off from the View tab. The pattern of labeling is determined from the pattern configured for each peak list. The atoms on the *X* and *Y*-axis are shown in *blue* and the atom on the *Z*-axis plane is shown in *green*. Weaker, inter-residue peaks in the primarily intra-residue experiments are shown in a *smaller font*

Two orthogonal views (^1H - ^{13}C and ^{15}N - ^{13}C) are shown for each of the six experiments. The blue vertical line is at the amide proton shift for the first view and amide nitrogen shift of the second view. Clicking the **Show Labels** button in the View tab of RunAbout will overlay the spectra with annotations that show what spectral information is displayed in each window (Fig. 11).

- Go to the Helm tab. Here you will find up/down arrows that allow you to navigate through the list of peaks in the reference (HNCO) spectrum. You will probably find it useful to step through most, if not all, of the peaks in the reference list and observe the corresponding spectra. If you notice many peaks that are not picked, or many artifacts that are picked, you may well want to go back and redo the peak picking at a more appropriate contour level. If there are many reference peaks that are picked at positions of artifacts or noise you may want to re-pick the whole spectrum at a different contour level. But you can easily delete a few peaks by clicking the Delete (Skull and Crossbones) button to the right of the up/down peak navigation buttons.

Also, note how well the peaks are aligned. In each spectrum, a vertical blue line is drawn at the chemical shift of the reference peak. The peaks that are likely coming from the same amide proton should line up right on that line. If not, you may need to adjust the referencing of the spectra. Getting your spectra well aligned at this point, will make things work better in later stages, so spend some time getting this right. Tools available for alignment are described in the next section.

4. Peak lists can be manually or automatically aligned.
 - (a) To align peaks manually step through the peaks in the Helm until you find a set of peaks you want to use for alignment. Choose a spectrum window in the main Run-About spectrum display and move the black vertical cursor and adjust it till it is centered over the peak that you wish to be aligned to the reference line. Now click the Manual button in the Alignment section of the “Actions” tab. This will adjust the chemical shift referencing of both the dataset and peak list displayed in that window so that the position at the crosshair is shifted to the reference line position. You will need to repeat this process for each spectrum that you wish to align, and for both the proton and nitrogen views of each spectrum.
 - (b) The manual alignment procedure assumes that you can use a single peak to choose best alignment. The automatic alignment method makes use of information from all peaks in each list. In the Actions tab click the “Auto” button in the Alignment section. This will take each peak list and compare it to the reference list. A comparison score is calculated by using a bipartite matching algorithm to come up with the optimum matching between reference peaks and the other list. The offset (for HN and N) between the two lists is iterated till it converges to a minimum. The lists should not be dramatically different in alignment before starting so you may want to use the manual procedure first to get peaks reasonably close. The amount that lists are shifted will be displayed in the NMRViewJ console window.

3.9 Filter and Trim Peak Lists

Two methods are available to do a bulk elimination of peaks that are artifacts. These rely on criteria that may not hold for your spectra so it is a great idea to save the project state first. Then try the analysis and verify (perhaps by scanning through with the Peak Edit mode described above).

1. The Filter action compares each of your peak lists to the selected reference peak list and removes peaks that do not exhibit amide ^1H and ^{15}N frequencies found in the reference

list. For the experiments handled by RunAbout, all spectra should look like HNCO spectra when projected onto an ^1H - ^{15}N plane, so choose the *hnco* peak list as your “reference peak list” unless you have a special reason for not doing so. When filtering, RunAbout looks for reference peaks that are within two times the tolerance value set for the peak list. This increased tolerance is to minimize the probability of erroneously removing peaks that are valid, but not well aligned. To perform this action, go to the **Actions** tab, set the **Filter Peak List** to the *hnco* experiment, and click the **Filter** button. The NMRViewJ console window will display information about how many peaks were marked for deletion.

2. The Trim action removes peaks based on their intensity. For a given protein and an experiment type, one expects a certain number of cross peaks. This number is calculated based on the peak list patterns and number of amino acids. If you picked your peaks conservatively, you should probably have more than the expected number of peaks, even after filtering. The “Trim” function allows you to cut out the less intense peaks, thus the ones most likely to not be valid, based on the number of expected peaks. Here, you should provide a “trim number,” which RunAbout equates, just for this purpose, with the number of residues in your protein; to be safe, give an estimate that is ~10% larger than the number of residues in your protein. Ubiquitin has 76 residues, so here the trim number is set to 84. Click **Trim** to perform the action. Trimming has a significant risk of removing weak, but valid peaks, so use this with care, and save the project first!
3. After doing the above commands, click the **Compress/Degap** button in the **Actions** tab to ensure that peaks marked for deletion are permanently removed and gaps in the peak numbering are removed.

3.10 Cluster Peaks

Each of the backbone 3D experiments detects magnetization from the amide protons so there will be multiple peaks from the experiments that share the same amide resonance. The clustering step groups together peaks from your peak lists that have the same amide H and N frequencies and thereby probably come from the same atom. Clustering is thus the basis by which sequential assignment is made: clustering determines which resonances in different peak lists belong to the same residues, and also begins distinguishing intra- and inter-residue peaks.

Peaks are clustered together based on the tolerance values set up for each list. Additionally, the clustering is constrained so that every cluster must have one, and only one, peak from the reference list. Clusters will be assigned an identification number that corresponds to the index of the reference peak of that cluster.

1. Click the **Cluster Peaks** button in the Actions tab.
2. The NMRViewJ console will display the results of the clustering with information on how many clusters there are of specified size. For example, the sample data will yield something like:
 - 3 cluster(s) with 1 peaks
 - 1 cluster(s) with 2 peaks
 - 1 cluster(s) with 3 peaks
 - 3 cluster(s) with 8 peaks
 - 9 cluster(s) with 9 peaks
 - 12 cluster(s) with 10 peaks
 - 35 cluster(s) with 11 peaks
 - 8 cluster(s) with 12 peaks
 - 1 cluster(s) with 13 peaks
 - 1 cluster(s) with 14 peaks
 - 1 cluster(s) with 18 peaks
3. Some spurious clusters will probably be created in this process. The ones most easily identified as being misleading are those with just one or two peaks in the cluster—so-called lonely clusters. You can easily get rid of these by setting a “lonely Limit,” which corresponds to the minimum number of peaks in a legitimate cluster, and clicking **Purge Lonely**. Here, we set the lonely limit to 2 so that all clusters with just one peak are eliminated, and all those with two or more are kept. After clicking **Purge Lonely**, the console will display information on how many peaks were removed and another summary of the cluster statistics.

3.11 Examine and Edit Clusters

Now that you have grouped your peaks into clusters, you should check whether they look reasonable. Having well-defined clusters is important for the subsequent steps within RunAbout and also important if you want to export clusters to be used in an external auto-assignment program. Enter into *Edit Clusters (HN-C)* mode in RunAbout’s top toolbar by clicking **Mode>Edit Clusters (HN-C)**. Clicking this will refresh the spectra in the display, but the spectra corresponding to each position in the panel are the same as in the *Edit Peaks* mode. However, you will observe that the peak labels are all preceded with “c” in the “Edit Clusters” modes, indicating that the labels correspond to the cluster number.

Switch to the RunAbout Helm tab to see information about each cluster and controls for navigating and editing the cluster list. You can see immediately, from the entry field between the **Split** button and the up-arrow button, that you are working with cluster

#0. The cluster numbering is linked to the peak numbering in your reference spectrum (probably your HNC0), and the reference peak itself is displayed in the **Reference Spectrum** panel at the bottom left of the Runabout window. The up- and down-arrows enable moving between clusters. The **Clusters** display, in Edit Clusters mode, provides a list of all the clusters, and draws a box around the one you are looking at now. You can move directly to any cluster by clicking its number in the **Clusters** panel. The **Peaks in Cluster** panel shows the peaks from different peak lists that are linked into this cluster; “hnca.1” is peak #1 from the hnca peak list, for example. The **Current Cluster** section provides some handy numerical data about your active cluster in the “Cluster1” section: cluster number, number of peaks in the cluster, and the HN and N frequencies. **Cluster2** refers to the peaks in the right-hand panel in the “Peaks in Cluster” panel. Right now, this is empty, but will be populated and useful when we work with moving peaks between two clusters.

Examine clusters systematically to ensure that you have the best use of your data. Two strategies are useful. You can start at cluster 0, and step through all clusters, examining and editing them as you go. Or, you can examine clusters in groups by criteria based on how many peaks (relative to the expected number) are present. Here, we will do the latter:

1. Choose *correct* from the menu above the **Clusters** list in the **Helm**. You will see a list of about 35 clusters that have the correct (11) number of peaks. Move through these by clicking the Cluster navigation arrows in the **Edit Cluster** region of the GUI. Starting with the correct clusters is a good idea as it gives you a sense for what pattern is desirable. But do remember, the ubiquitin data used here is very high quality. You may be missing a greater number of peaks in your own protein and there will be fewer clusters with the “correct” number.
2. A useful aid in examining clusters is the **Cluster Detail** window (Fig. 12). Display this by clicking the blue “i” button in the Helm. For each peak list, it displays information for each peak in that cluster. Peak numbers or peak chemical shifts (depending on settings at top of window) are shown under the type of atom that peak has been typed as (ca, cb, c, and intra- or inter-residue). Next to each peak number is a vertical bar, showing the relative intensity of that peak. The bar is blue for positive peaks and red for negative peaks. If there are more than the appropriate number of peaks for a given peak list, the weaker peaks will be shown in subsequent rows. The detail window shows a yellow box where there is an expected peak for each atom type in each list. For example, an hncacb experiment with a “i,i-1.ca,cb” pattern would have four boxes. Empty boxes are something you will want to look at as they indicate either missing peaks, or peaks that were incorrectly assigned by the software to atom name and

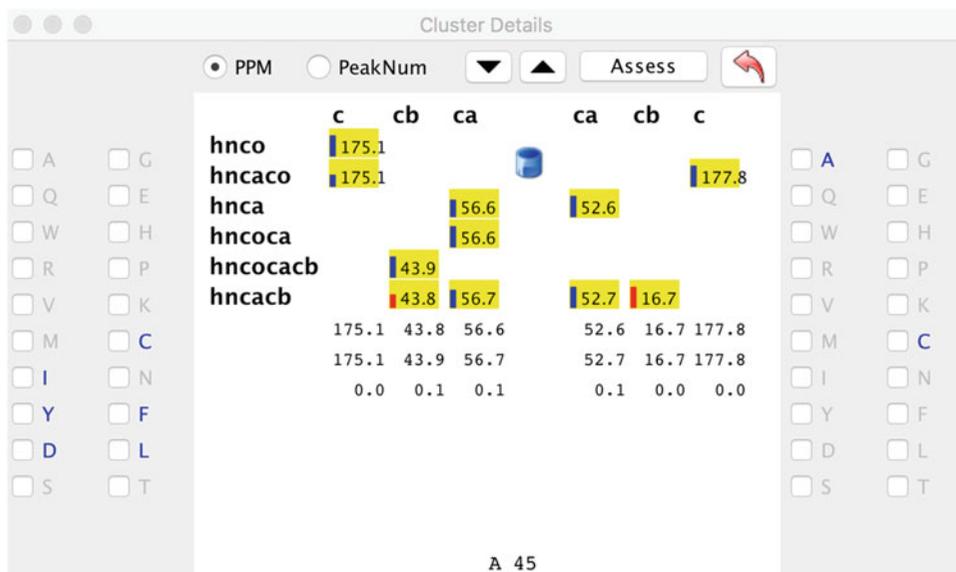


Fig. 12 The **Cluster Detail** window showing data for the first (0) cluster. All the *yellow boxes* are filled in with chemical shift values indicating that this cluster has all the expected peaks. The shift values within each column have very nearly the same values indicating good referencing and likely correct choice of peaks for the different possible types. The inter-residue ($i - 1$) carbon shifts (CA of 56.6 and CB of 43.8) are consistent with six possible amino acids (YDCFL). The intra-residue shifts are only consistent with A and C. Note that C will often show up as a possible assignment type as currently no distinction is made between reduced and oxidized cysteine CB atoms, and they thereby have a large standard deviation in the database allowing them to be consistent with many peaks. It is possible to set the expected shift and range for each cysteine to specific values for either the oxidized or reduced state in the Atom assignment table and this will minimize the number of clusters consistent with cysteine

type (inter or intra-residue). The cluster detail window also shows the chemical shift range for each atom type. Three numbers appear at the bottom of each atom, the minimum shift, the maximum shift, and the delta between those two. If the delta is greater than 0.5, it will be highlighted in a magenta color. A large delta is not necessarily bad, it may just mean that a peak has been typed as “ $i,i-1$ ” (could be either the intra-residue or inter-residue peak), and will still match to the sequence properly. But if you like to have all your “ducks in a row,” you should pay attention to this. Peaks that are probably “inter-residue” peaks but have not been typed that way are shown in magenta. The Cluster detail window has additional information including the possible residue types for the current cluster. Review the NMRViewJ documentation for more information. One important additional feature is the list across the bottom of the window. This lists some of the possible sequence positions that the cluster could fit at (based on the carbon shifts and expected values for different residue types). If there are no entries in the list it means this cluster will not fit anywhere and you will want to review it carefully.

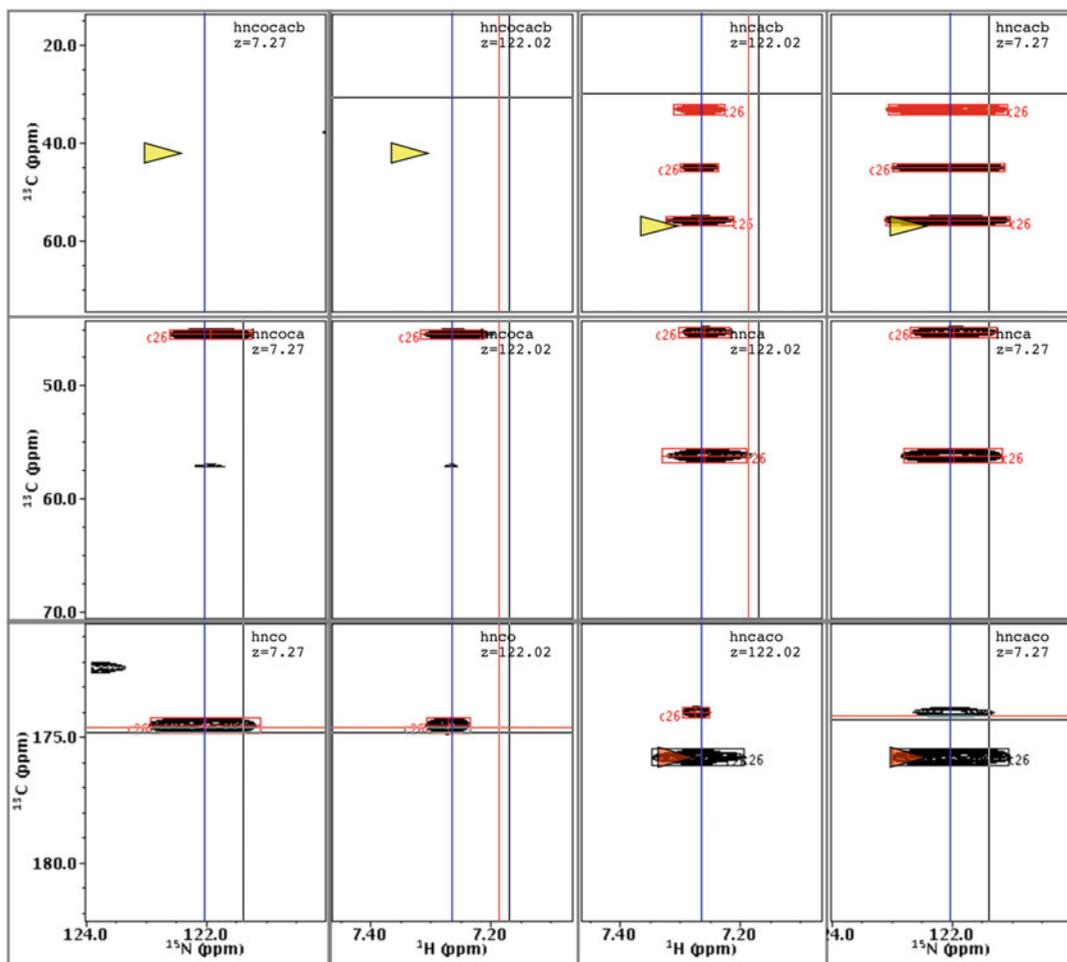


Fig. 13 The grid of spectra in **Edit Cluster** mode for a cluster with fewer peaks than expected for the hncocacb and hncacb experiments, and an extra peak for the hncoca experiment. The *arrows* are only hints as to where extra peaks might be removed or missing peaks added

3. Continue examining all the “correct” clusters. Most of these clusters will look good, but a few have some artifacts. For example, cluster 27 (if your numbering is the same) is unusual. It coincidentally has the correct total number of peaks (11), but is missing some peaks for some experiments, and has extra peaks for another (Fig. 13). By examining the spectra and Cluster Details window you will see that the hncocacb experiment does not have any peaks and the hncaco and hncoca have extra peaks. There are several ways to remove the extra peaks.
 - (a) Click the Trim button in the Helm. This removes the weakest peaks in any experiments that have extra peaks. The peaks that will be removed have “x” characters next to them in the **Peaks in Cluster** list and magenta arrows pointing at them in the spectrum display.

- (b) Remove them by zooming into each spectrum. Click the mouse in a spectrum window (in RunAbout) and hit the z key. With the mouse cursor in Selector mode click on the peak box to be deleted (this selects it) and click the “d” key (this is a special key binding that works in the zoomed RunAbout window). Hit z again to zoom back to the full grid of windows.
 - (c) Identify the peak label in the Cluster detail window and click and drag it to the center of the window (under the trash can icon). This will delete it. This cluster is missing a peak in the hncocacb because the residue is probably a glycine so we will ignore this (in fact, there is a negative peak that we have not shown at the correct position).
4. Check out additional clusters that have extra peaks, by changing the Clusters menu to *extra*. Proceed through each cluster and remove artifactual peaks. The fastest way to do this is with the Trim method described in a above, but some care is advised to examine what peaks will be removed. (Note that there is a **Trim All** button in the **Actions** tab that will do this trimming automatically (use caution, and save project first).
5. Often, the extra peaks in a cluster are due to artifacts, but sometimes a cluster may actually represent the peaks from two separate residues. This can happen if the amide H and N shifts for the two clusters are overlapping. In this case, it is likely that you will have more than just one or two extra peaks and you will want to split the cluster into two. You can perform a split by clicking the Split button in the Helm tab. This will create a new cluster by finding the largest gap between the existing peaks in the cluster and dividing the peaks between the original cluster and the new cluster. A new HNCO (or other reference) peak will be created at the center of the new cluster HN/N shifts. You can interactively move peaks between the two clusters by clicking on their entries in the cluster lists or zooming in the spectrum window (click “z”), selecting a peak, and clicking “s” to swap it between the clusters. Once you have set up the new cluster, hit the **Relink** button or the changes will be lost.
6. Now examine clusters that are missing peaks by changing the Clusters menu to *missing*. There are a variety of reasons that peaks may be missing. Two peaks may be overlapped so that they are indistinguishable, peaks may be simply too weak to be observable (especially true for the weaker inter-residue peaks found in experiments that primarily show intra-residue peaks like the HNCA), they may be obscured by an artifact such as a streak from residual water, or no peak is expected (CB peaks for glycine).

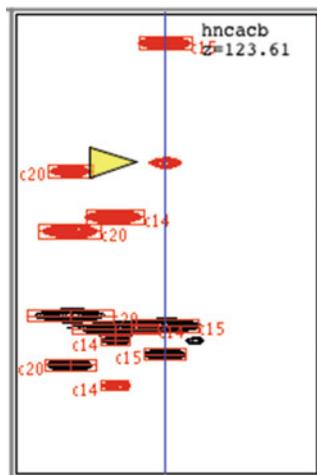


Fig. 14 The spectrum area in RunAbout **Edit Cluster** mode for the hncacb experiment. This cluster is missing a peak and the *yellow arrow* points to a likely place to add one. Often the *yellow arrow* will point to an area where no signal intensity is visible above the contour level, but the algorithm used will have found a local maximum below the visible level. In this displayed case, there is signal above the contour level, but it was not picked as a peak because this peak exists as a shoulder of an adjacent peak. Use the “z” key binding to zoom the spectrum so it occupies the entire RunAbout spectrum area. Several new key bindings are then available. Use “A” to add a peak at the exact position of the mouse cursor

Some reasons listed above will preclude actually adding the missing peak, but if it appears that there are missing peaks that can be added it is a good idea to do this. The yellow arrows in the spectra are suggestions (based on the intensities in the spectrum) as to where a peak might be. The zoomed-in window (see step b below) for the hncacb experiment of cluster 15 has a missing a peak, and the yellow arrow points to a reasonable spot to add one (Fig. 14). It turns out the position indicated has signal intensity above the contour threshold, but there is another, more intense peak, that overlaps this one so only a single peak was picked.

- (a) Lower the contour threshold in the window till you see if a reasonable peak exists (in cases where it is not already visible as it is here).
 - (b) Click **z** to zoom into the window.
 - (c) Place the mouse cursor over the area where you wish to add a peak.
 - (d) Click the **A** button to add the peak (using a capital A will force the pick to occur at exactly the cursor position, otherwise the peak picker may scan to a nearby maxima).
7. Sometimes, a cluster can have too few peaks because some peaks were clustered in with an artifactual peak in the reference list. Clicking the **Neighbor** button will display the cluster

nearest to the current cluster (and each time you click it you will get another close cluster). You can move peaks (except the reference peak) from the second cluster to the first (as described in **step 5** above) and then delete the remaining artifactual peaks (click the red X button to delete an entire cluster).

8. The experiments that involve intra-residue magnetization transfer, such as the HNCA and HNCACB, will generally also transfer magnetization via a weaker inter-residue connectivity through the carbonyl carbon. Properly distinguishing which peaks are the intra-residue and which are the inter-residue peaks is useful in performing the assignment. This typing is done during the initial cluster analysis, but can also explicitly be done with a more exhaustive analysis after clustering. You can do this by clicking the **Assess** clusters button in the **Actions** tab. Save your project and click this button now. This applies the algorithm to all clusters. It is also possible to do a single cluster (perhaps after doing some adjustments to the cluster) from the **Assess** button in the Cluster Detail window.
9. Having checked clusters by status (correct, missing peaks, extra peaks), it is a good idea to do one more pass through all clusters. At each cluster, have the Cluster Detail open and check that the shift range for carbon chemical shifts for each atom is small and that the cluster fits into at least one sequence position.

3.12 Link Clusters into Fragments

Clusters in HN detected experiments represent a set of peaks that have a common amide proton and nitrogen frequency and have carbon shifts from the previous and current residues. The next step of assignment is to find overlaps of the current residue shifts of one cluster with the shifts of another cluster. Matching is done by looking for inter-residue shifts of one cluster that match intra-residue shifts of another, or vice versa. In this way, a series of clusters can be linked together into overlapping clusters that form an entire fragment of the amino acid sequence.

There are several ways to link together clusters within the RunAbout tool.

1. Manually link clusters based on an examination of the cluster-cluster overlap information.
2. Automatically link clusters based on criteria similar to those used in the manual linking method.
3. Using an automated internal or external tool to identify overlaps and sequence assignments and then importing that information back into RunAbout for confirmation and use in completing the assignments.

Each of these methods will be described below.

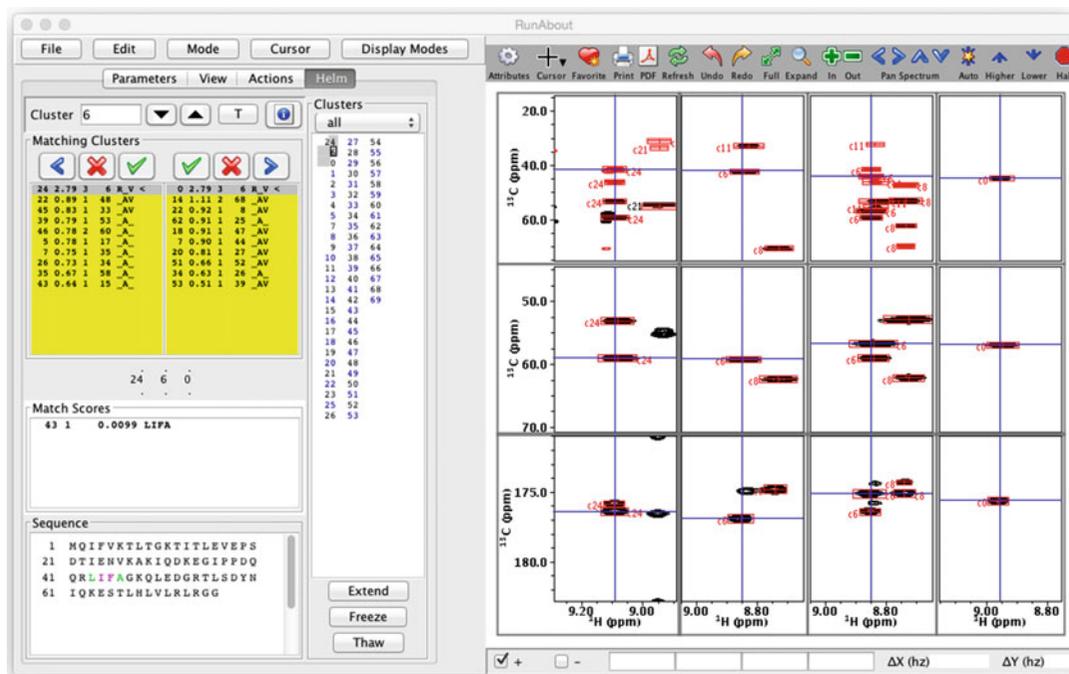


Fig. 15 RunAbout in **Link Edit** mode for cluster number 6. The Matching Clusters region shows two lists for clusters that have one or more overlapping carbons with the current cluster. The *left-most list* showing cluster matches with overlaps to this cluster inter-residue ($i - 1$) peaks and the *right-most list* showing clusters with carbon matches to the intra-residue (i) peaks of this cluster. The Clusters region shows a list of all clusters. The link cluster links 24-6 and 6-0 have been confirmed so both lists are in *yellow*. These three clusters (24-6-0) are listed together in the list of all Clusters, and the sequence region from residues 43 to 46 is *highlighted*. Short fragments can match in multiple places and the list of places they match, and a probability-based score is shown in the Match Scores section (here there is only one possible match)

3.12.1 Manual Linking

To manually link clusters first put RunAbout into the **Edit Links** mode (Fig. 15). The Helm area will change to show information about potentially matching clusters and the spectrum area will show the spectral regions of these clusters. As you step through clusters the two lists of information in the Matching Clusters area will be updated with information about possible matches. The list on the left shows matches from the current cluster to the cluster of the preceding residue ($i - 1$: i), and the list on the right shows matches from the current cluster to the cluster of a succeeding residue (i : $i + 1$). The fields on each line in the list box represent the following information:

1. The number of the matching cluster.
2. The score of the match. Higher values are better and the maximum value is equal to the number of potentially matching

carbons, so it would be as high as 3.0 for a match involving CA, CB, and C.

3. Number of carbons with shifts that match within a specified tolerance.
4. The cluster that the matching cluster best matches.
5. A three-character representation, where **R** indicates that the match to the other cluster is reciprocal (meaning the other cluster best matches to this cluster), **A** indicates that the matching cluster is Available (has not already been linked to another cluster) and finally **V** indicates that a fragment formed by linking the two clusters would have chemical shifts consistent with some location in the sequence and is, therefore, a Viable match.

The spectrum display windows are similar to that in the other (Peak Edit and **Cluster Edit**) modes, but the outer columns of spectra represent different clusters. The center two columns represent the inter-residue ($i - 1$) and intra-residue (i) regions of the current cluster. The leftmost column represents the intra-residue (i) regions of a possible previous cluster (and therefore should have similar carbon peak positions as the column to its right). The rightmost column represents the inter-residue ($i - 1$) regions of a possible succeeding cluster (and therefore should have similar carbon peak positions as the column to its left). Selecting a row in either of the two list boxes will update the corresponding (leftmost or rightmost) spectrum columns so that you can visually check for peak alignments.

Figure 15 shows an example that should make the above clear. RunAbout is set on cluster 6. The middle pair of columns in the spectrum display show the peaks involved in this cluster. The left list box in the Matching Clusters area shows a list of ten clusters (24, 22, 45 ...) that are possible clusters to represent the residue previous to the residue giving rise to peaks in cluster 6. They are listed in order of their matching score and it can be seen that the first one, cluster 24, clearly has the best score (2.79). The best match of cluster 24 is to cluster 6 (the third column) so the match is reciprocal, it has not been matched to another cluster so it is available, and a fragment formed by cluster 24-6 would match into the sequence so is viable, hence the “**RAV**” value. The right list box shows the values for potential cluster matches for the next residue and here cluster 0 is clearly best. Examination of the spectra shows excellent visible alignment (along the horizontal blue lines) of the peaks for the CB, CA, and C carbons. Based on the good match scores and visible alignment we can manually confirm that these cluster pairs (24-6 and 6-0) should be confirmed. This can be done by clicking the two green checkbox buttons above each list. The confirmation is done for whichever entry in the list box is

currently selected. The background of each list box will turn yellow when the cluster pair is confirmed. Linking two or more clusters forms a fragment and the carbon chemical shifts of that fragment will be matched to the expected values for the given amino-acid sequence. The Sequence region of the Helm, in Link Edit mode, shows the proteins amino acid sequence and possible locations for the fragment involving the current cluster will be displayed in green (with the residue pair for the current cluster shown in magenta). As fragments grow in length (by confirming cluster matches at each end) the number of possible sequence positions will be reduced till there is usually only one possible place.

Having confirmed the matches for a particular cluster, you can now move onto another cluster. The blue arrows above each list box will take you to the cluster that is linked to the current cluster. Extending the fragment in each direction allows you to grow the fragment until you reach a point where there is no match (because of missing peaks or the presence of a proline residue), or the match is ambiguous. When a fragment can no longer be extended in either direction, move to a new cluster and begin the process again.

Linking clusters to form fragments does not actually update the assignment fields of the involved peaks, or update the chemical shifts of the involved atoms. This is done by “Freezing” the fragment by pushing the **Freeze** button at the bottom of the **Clusters** display. Freezing updates these data values and thereby marks the residues involved as being unavailable for assignment of other clusters. Doing this, therefore, removes possible assignment locations for clusters whose previous assignment possibilities were ambiguous.

3.12.2 Automatic Linking

The process of manually linking clusters described above relies on assessing cluster matches based on various criteria. The **Extend** button (below the Clusters region) or the **Extend All** button in the **Actions** tab automates this process. The extend protocol checks various criteria (the number of matching carbons, the score, whether the match is reciprocal, etc.) and if it meets the criteria the cluster-cluster pairing is confirmed. The process is continued from the newly linked cluster and this is repeated until no more links can be made. The other end of the fragment is then analyzed in a similar fashion. The criteria for forming a link can be set in the Parameters tab. Try this out on the sample data by going to cluster 6 and clicking the **Extend** button. Clusters will be linked to form a fragment that extends from P38 through to G47. The automatic extension stops at P38 because it is a proline (lacking an HN proton for cluster formation), and at G47 because it is a glycine (the default matching requires three carbon matches and glycine, lacking a CB, can only match 2).

The protocol initiated by the **Extend All** button repeats this process for each cluster. The automated protocol recognizes glycine residues as only having two available carbons and so will continue linking through them. Click on a cluster in each fragment listed in the **Clusters** area to see where that fragment matches into the sequence. Nearly complete assignment of the protein has been obtained. As with the manual method above, you need to explicitly freeze the fragments to update assignment information. The **Freeze All** button, located in the **Actions** tab, will freeze all fragments.

3.12.3 Automatic Assignment Tools

Clusters formed within the RunAbout tool are also an input source for automatic assignment algorithms. The cluster data can be exported in several formats using menu items in the **File** menu of the RunAbout tool. Here, we will describe using an assignment tool within RunAbout. This tool, NetMatch, can be run as a standalone tool or internally to NMRViewJ. In either case, it can use formatted cluster data exported from RunAbout.

The NetMatch tool (short for Network Matching, as it matches up networks of related peaks) uses two algorithms to come up with an assignment of the clusters to sequence positions. This tool is relatively new, and the details of the method unpublished so we give a brief overview of it here.

First, it uses a maximally weighted bipartite matching algorithm to derive matches between clusters and residues [29]. The weights for the matching are based on how similar the chemical shifts of the atoms of the cluster are as compared to the expected values for each pair of residues in the sequence. The more similar the actual and expected chemical shifts are, the higher the weight for that match. The algorithm comes up with the best possible matching, that is the one where the sum of the weights for each match in the solution is highest.

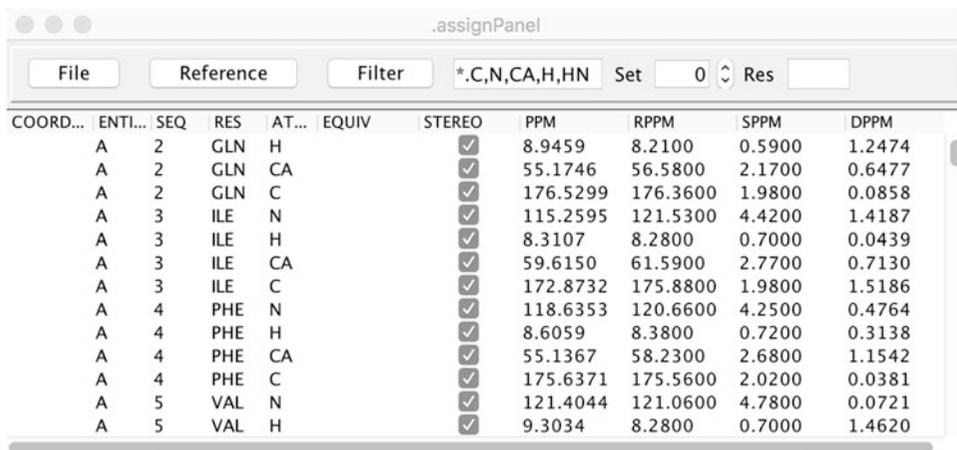
Because there is a significant range of expected chemical shifts and these overlap for multiple amino acid types, and because the same pair of residues can occur at multiple positions in the sequence this is, unfortunately, not sufficient to unambiguously assign the clusters to the sequence. To resolve the ambiguity, it is necessary to also include information about the consistency of the chemical shifts that derive from overlapping clusters. That is, the chemical shift of each carbon atom might be represented in two adjacent clusters. If the inter-residue shifts ($i - 1$) of one cluster represent the same atoms as the intra-residue shifts (i) of the previous cluster, the two sets of chemical shifts should agree within some tolerance as is also used in the above manual protocol. The sequence matching score and a score measuring cluster-cluster agreement can be combined into a single global score as done, for example, in FLYA [30].

Finding the best matching of clusters to sequence positions is a problem of combinatorial optimization, where the goal is to find the matching that has the largest global score. Many algorithms are available for performing this optimization. The NetMatch tool uses a Genetic Algorithm as previously used in other assignment tools [15] and as implemented in the open source Jenetics package [31]. Using a preexisting package like Jenetics allows us to use a variety of variants of the Genetic Algorithm built into the package. The matching problem is represented as a sequence of unique integers. The position in this sequence represents one residue-pair in the protein sequence, the integer value at that position represents a particular cluster. The algorithm evolves a starting configuration via a process of mutation (swapping the values at two locations in the sequence of integers), crossover (combining parts of two separate match sequences), and selection (choosing match sequences with better scores).

The NetMatch tool significantly enhances the ability of the genetic algorithm to find an optimal solution by using the bipartite matching algorithm to provide an initial population of good starting matches. Multiple starting matches are generated by repeatedly calling the bipartite matcher. Prior to each call a different match pair is eliminated from the set of possible matches so a different, but relatively good, match sequence will be generated. The actual NetMatch implementation does several additional refinements of the starting configurations that further improve their global score. The starting populations are then evolved toward an optimal solution.

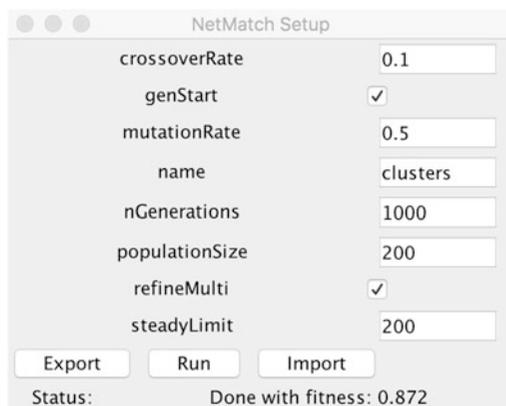
The actual protocol for the tool is described below. Note that this algorithm is under rapid development so refer to the latest documentation as well.

1. Generate, inspect, and clean up clusters as described above in Subheadings 3.1–3.11. The better the quality of the clusters the better the expected quality of the solution.
2. Open up the Atom Assignment Panel (**Assign>Atoms** menu item).
3. Select the **BMRB_Avg** menu item from the **Reference** menu. This will set the expected chemical shift for each atom in the sequence based on statistical values from the BMRB (see Fig. 16). If you have an alternative source of reference shifts (from a related protein, or shift prediction algorithms, for example), you could read those values in using, for example, the **Read File** menu item (see documentation).
4. Display the NetMatch Setup dialog from the RunAbout **File>NetMatch** menu item (Fig. 17).



COORD...	ENTI...	SEQ	RES	AT...	EQUIV	STEREO	PPM	RPPM	SPPM	DPPM
A	2	GLN	H			<input checked="" type="checkbox"/>	8.9459	8.2100	0.5900	1.2474
A	2	GLN	CA			<input checked="" type="checkbox"/>	55.1746	56.5800	2.1700	0.6477
A	2	GLN	C			<input checked="" type="checkbox"/>	176.5299	176.3600	1.9800	0.0858
A	3	ILE	N			<input checked="" type="checkbox"/>	115.2595	121.5300	4.4200	1.4187
A	3	ILE	H			<input checked="" type="checkbox"/>	8.3107	8.2800	0.7000	0.0439
A	3	ILE	CA			<input checked="" type="checkbox"/>	59.6150	61.5900	2.7700	0.7130
A	3	ILE	C			<input checked="" type="checkbox"/>	172.8732	175.8800	1.9800	1.5186
A	4	PHE	N			<input checked="" type="checkbox"/>	118.6353	120.6600	4.2500	0.4764
A	4	PHE	H			<input checked="" type="checkbox"/>	8.6059	8.3800	0.7200	0.3138
A	4	PHE	CA			<input checked="" type="checkbox"/>	55.1367	58.2300	2.6800	1.1542
A	4	PHE	C			<input checked="" type="checkbox"/>	175.6371	175.5600	2.0200	0.0381
A	5	VAL	N			<input checked="" type="checkbox"/>	121.4044	121.0600	4.7800	0.0721
A	5	VAL	H			<input checked="" type="checkbox"/>	9.3034	8.2800	0.7000	1.4620

Fig. 16 The **Atom Assignment** panel. This shows a list of all atoms in the molecular structure. It can be filtered by atom type (here set to shown backbone atoms). The Reference menu can be used to populate the RPPM column with reference chemical shifts and the SPPM with their standard deviations. Actual assigned shifts for each atom would appear in the PPM column



Parameter	Value
crossoverRate	0.1
genStart	<input checked="" type="checkbox"/>
mutationRate	0.5
name	clusters
nGenerations	1000
populationSize	200
refineMulti	<input checked="" type="checkbox"/>
steadyLimit	200

Buttons: Export, Run, Import

Status: Done with fitness: 0.872

Fig. 17 The **NetMatch** dialog window. This is used to configure options for this tool and to export and import data files used and produced in the analysis. See the full documentation for information on this new, and developing, tool

5. Configure any desired parameters. Here we will just use the default values. But note that you can do multiple runs and keep the results separate by giving each one a different name value.
6. Export the data values needed for the NetMatch tool by clicking the **Export** button. Two files will be created inside a *netmatch* subdirectory of the current project (see **Note 20**).
7. Run the NetMatch tool by clicking the Run button. The process will start and the status area will be updated with the elapsed time. The NetMatch tool runs in its own process thread

so you can continue interacting with NMRViewJ while it is running (*see Note 21* about multi-core). When complete the status will be updated with the final fitness value and various data files will be written to the netmatch directory in your project (*see Note 22*).

8. You can import the results of the NetMatch tool into RunAbout by clicking the **Import** button. This examines the information in the finalmatch.txt output file for this run. Any clusters that have a match to a specific residue number will be marked as being restricted in their assignment possibility to that residue number.
9. You can now proceed with Link Edit mode as described above for either the manual or automated modes, but each cluster will have additional information about residue restrictions so the number of possible cluster matches with an RAV status will typically be limited to one. You can view the residue constraints by looking in the Cluster Detail window. The area above the list of amino acid types will display a residue number if that cluster has been constrained to a particular sequence position. Clicking the **Extend All** button will join clusters in to fragments as described in the above sections, but will now have access to the residue constraints on each cluster that is imported from the NetMatch output files.

3.13 Examine Assignments

Assignments done in RunAbout (or other ways) can be examined in several different ways.

1. Check that peak labels involved in assignments have atom values. You can see this on a peak-by-peak basis by looking in the Peak Inspector (from the **Peak>Show Peak Inspector** menu item) or for all peaks in a list by using a Peak Table (from the **Peak>Show Peak Table** menu item). A summary of all the assigned peaks can be seen by showing the Peak Lists table (from the **Peak>Peak Lists Table** menu item). This will show a table of all peak lists and the number of peaks and number of assigned peaks for each one.
2. Check that the Atoms have chemical shift values by opening the Atom Assignment table (from the **Assign>Atoms** menu item) as shown in Fig. 16. You can use the Filter menu (or enter patterns like *.C,N,CA,H,HN) to see a subset of atoms. If you display Reference information (for example, **Reference>BMRB_Avg** menu item), you can compare your assigned shifts to the reference (typically BMRB average and standard deviations) values by looking at the DPPM column. By default, this shows the number of standard deviations your shifts are from the reference value. Click on the column header to sort the table by this value.

3. Check the Resonance Table (from the **Assign>Resonances** menu item). NMRViewJ organizes peak and assignment information by resonance. A given atom resonance can be shared by multiple peak dimensions. The Resonance table will show all the resonances, the assignment for each, how many peaks are shared by the resonance, and the mean and standard deviation of the shifts from each peak. Sorting the table by values in the *ppmdev* column (the standard deviation of shifts shared by that resonance) allows you to observe any resonances that have an unusually large shift range and thereby involve peak dimensions that should not have been combined.
4. Examine the LACS plot (from the **File>Analyze Shifts>LACS Plot** menu item of the Atom Assignment table). This plot is of the $\Delta\delta C\alpha$ vs. $\Delta\delta C\alpha - \Delta\delta C\beta$ chemical shifts. Two separate lines are fit to the values that have $\Delta\delta C\alpha - \Delta\delta C\beta$ values above 0 and below 0. These lines should intersect at the origin of the plot if your data was correctly referenced. The LACS plot for the sample data is shown in Fig. 18. The offset of the intersection with the origin is only 0.04 ppm, indicating that the data was well referenced.

The above protocols illustrate the basic steps in using triple resonance NMR data to determine the chemical shift assignments of a protein. The ubiquitin sample data used here is of very high quality making this procedure relatively straightforward. Other

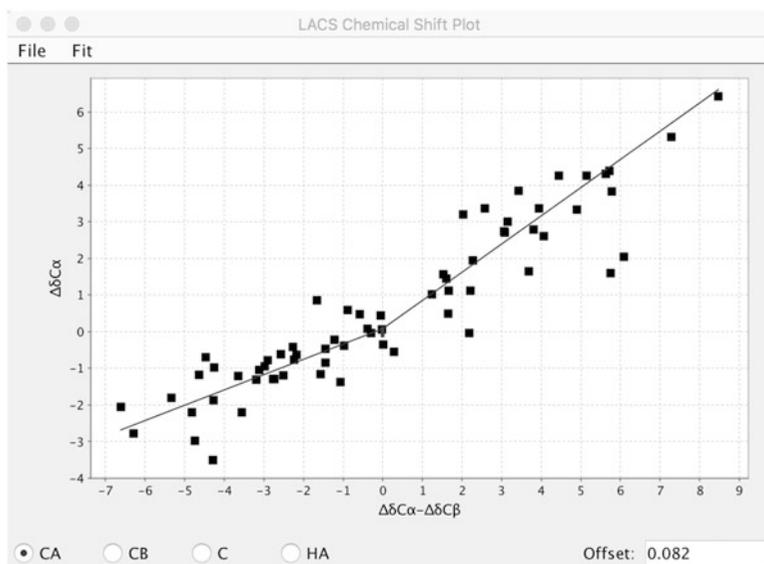


Fig. 18 A LACS plot [35] of the assigned chemical shifts. This is useful for checking if referencing was done correctly as the two sloping *blue lines* should intersect at the origin (0,0). The offset value shows the distance this intersection is from the origin and is thereby a measure of the quality of referencing. Outliers on this plot should be checked for correctness in their assignment

proteins, especially if they are larger, unstable or less soluble, may require greater time and attention to the details of the process, yet the basic protocol described above will still be appropriate.

4 Notes

1. This chapter will be illustrated with sample data on the 76-residue protein, ubiquitin. A wide variety of NMR data on ubiquitin has been collected and made available at The Ubiquitin NMR Resource Page (<http://www.biochem.ucl.ac.uk/bsm/nmr/ubq/>). You can get the appropriate data for this chapter by going to that page and selecting the TRIPLE RESONANCE archive on the left side of the page. On the Triple Resonance page click on the “Raw Data” link to download the files as a compressed tar archive. Uncompress the files on your computer (using the method appropriate to your operating system).
2. NMRFx Processor executes processing commands in the form of a script written in the Python programming language (actually in Jython, the Java implementation of Python [32]). The script consists of a series of commands that set up reference information and then a section of processing commands for each dimension of the dataset. You do not need to understand Python to use the software, but an advanced user may be interested in taking advantage of extended capabilities available through scripting.
3. The NMRFx Processor toolbar has controls for changing the spectrum view, including showing the full spectrum, expanding within the crosshair region, zooming in and out, automatically adjusting the vertical scale, and manually adjusting the vertical scale to make peaks appear higher or lower. The Attributes icon can be clicked to bring up an additional window with controls for adjusting the view. Overall, these tools are designed to be familiar to users of NMRViewJ.
4. Multidimensional NMR data is generally collected with multiple (2 for 2D, 4 for 3D, etc.) FIDs at each combination of indirect time increments. The individual FIDs may need to be combined in various ways depending on the design of the pulse sequence. The control at the lower left corner of the spectrum display window allows you to select which of these are displayed. Comparing the different signals (with operations that include the Fourier Transform) can be useful to confirm what type of combination of them is needed. For example, PEP data will often show one FID with positive signals and the other with negative signals. Proper combination with the TDCOMB command will convert this to one with twice the intensity

(from the difference of the two) and another with near zero intensity (from the sum of the two).

5. When the AutoGenerate command is used in NMRfX processor to generate a processing script, the software will examine the parameter files in order to determine what sort of TDCOMB (if any) is needed. Both Bruker and Agilent have parameters that should be useful in determining this, but not all datasets use the parameters correctly (modifications of the pulse sequence may do magnetization transfer that is not reflected in the parameter values). In this case, you may need to manually choose the correct value for TDCOMB.
6. NMRfX uses two sets of crosshairs displayed in black and red. FIDs and 1D spectra can have vertical crosshairs displayed while contour plots can have both vertical and horizontal crosshairs. If you have a one-button mouse the red crosshair will appear when you click with the mouse a sufficient distance away from the black crosshair. If you have a three-button mouse, then as soon as you click once with the middle mouse button, the software will recognize that you have a middle mouse button. From that point on, it will use the left mouse button for the black crosshair and the middle mouse button for the red crosshair. You can move the crosshairs to an exact position by typing a value into the entry boxes at the bottom of the spectrum window and hitting the Enter key.
7. In FID mode (the default when a dataset is first loaded, and controlled by the control at top right of the Processor window) the software will load and display FIDs from raw FID file. This is straightforward for the directly detected dimension as the data is commonly arranged as a contiguous set of values. Loading signals from the FID file for the indirect dimension (done when the dimension selector at the top of the Processor window is set to a value greater than **DI**) requires assembling a signal from non-contiguous regions of the data file and can fail for some complex arrangements of data. Also, there may not be sufficient signal for the indirect FIDs to be useful for visual examination.
8. This ubiquitin sample HSQC dataset was collected on a Varian NMR spectrometer with at least three frequency channels. The pulse program used for the experiment uses the third RF channel for pulses for the second dimension of the dataset. It is not obvious from parameter values that this is true, so NMRfX Processor assumes the use of the second RF channel for the second dimension and thereby gets erroneous carbon values for the nitrogen dimension.
9. NMRfX can calculate reference values for the indirect dimension by using IUPAC standard ratios [21]. You just need to

correctly indicate the nucleus (C, H, ^2H , N, or P). Doing this requires that you have correct referencing done for the direct (proton) dimension.

10. Once you have processing operations defined for all dimensions you can process the dataset. Processing is generally pretty fast (at least for 1D, 2D, and 3D) datasets so it can be useful to check the processing operations by actually doing the processing. You can switch between the display of the processed dataset and the original FID (with the operations applied for the specified dimension) using the control at the top right of the Processor window.
11. The ^1H , ^{15}N -HSQC dataset used here was apparently collected with a pulse sequence that resulted in some phase errors that cannot be corrected for. So even with setting the zero and first-order phases in the direct dimension, some peaks will show small phase errors.
12. NMRFx Processor can be run on the command line (without the GUI). To do this a shell script is used to run the version of Java that is installed within downloaded program. Check the full documentation for information on this script. You may need to edit the script to set the location of your installation of NMRFx Processor. Also, you may need to set your operating system shells PATH variable so the script can be found, or invoke it by typing the full path to where it is installed.
13. NMRFx Processor currently supports an internal version of IST [33] and NESTA [34] and the ability to call an external implementation of NESTA-NMR [23]. This is a rapidly developing area and expect that new algorithms will be added both internally and as externally callable code.
14. NMRViewJ has a variety of commands that are accessible via menus. In this chapter, we will use a format like **File>Projects>New** to indicate a menu item named **New**, that is found in a sub-menu of the menu **File**. Menus can be found in the main control window and various other windows (like the Peak Inspector or RunAbout).
15. Dataset labels are of great importance in the use of NMRViewJ. For example, automatic correlation of crosshairs happens between spectrum axes that have the same name, and the display of peak boxes uses common labels between the peak and the spectrum to determine what axis of the peak to display on what axis of the spectrum. Accordingly, it is important to choose a consistent scheme for naming dataset axes. These can be changed (via, for example, the Dataset Table), but it is always a good idea to start out a project with consistent naming.

16. Spectra are displayed in spectrum items on a so-called Canvas widget within a display window. A single spectrum item can have multiple, superimposed, datasets. You can also arrange multiple spectrum items within a single canvas. The arrangement of multiple spectra on a canvas can be completely arbitrary, but it is common to display them within an organized grid with a specified number of rows and columns. When drawing multiple spectra from the Dataset Table you will be prompted to select the number of spectrum items and number of rows (the number of columns will be calculated from the two other values). If the datasets in each row or column have the same axis values, you can save space by choosing a minimal arrangement that only displays the axes across the x -axis of the bottom row and y -axis of the left column. Switch between these modes with the **Minimize Border** and **Normal Borders** menu items in the **Canvas>Arrange Spectra** menu. Each of these two choices also has an option for showing the name of the dataset across top of the spectrum as it can be confusing to keep track of which spectrum is which in the grid.
17. NMRViewJ does not change the reference information within the actual dataset. Instead, it saves reference information in a separate parameter file. For example, if you have a dataset named *hsqc.nv* the parameter file would be *hsqc.par* (the file will not exist until you use a command within NMRViewJ to create it). Values in the parameter file take precedence over values in the dataset itself. Remember, if you copy a dataset file to a new location, copy any existing parameter file as well.
18. NMRViewJ provides a very powerful and relatively rapid method for interactively examining peaks. One or more windows can be set (on the “Peak Attributes” dialog) to have their spectral display region adjusted whenever the choice of peak displayed in the “Peak Analysis” dialog changes. The controlled windows have their display region changed so that they are centered on the chemical shift of the corresponding peak. When setting this up it is a good idea to first set the “ShowPeaks” attribute to the Expand mode. With this setting, the width of the display region will be set to a multiple of five times that of the width of the bounding box of the peak. Step through a few peaks in the “Peak Analysis” dialog until you find a peak that has a fairly typical intensity for the particular peak list. Now change the “Show Peaks” to the ExpandFixed mode. With this setting, the display will center on the peak position, but the display regions width will remain constant. Now, as you step through the individual peaks in the list, their relative widths will be visually obvious.
19. The cursor (controlled by the mouse) can be used in several different modes. The default is an arrow pointer that is used to

move the crosshairs. Additional modes are Selection (for selecting and moving objects such as peak boxes), Peak Delete (for clicking on peak boxes to delete the peak object), and Peak Add (for clicking in the spectrum to add a peak object at the position). You can change the cursor mode with the **Cursor** menu in the toolbar found at the top of most spectrum containing windows or with various keyboard shortcuts that start with “c”.

20. NetMatch uses two text files. The first, named *clusters.txt*, contains two sections, Peaks and Atoms. The Peaks section contains a row for each cluster and lists the chemical shifts of the peaks in the cluster. The Atoms section contains a row for each residue and lists the atoms that will be found for that residue (including the intra- and inter-residue atoms). The second file, *clusters_ppm.txt*, lists the expected chemical shifts for each atom type used. The NetMatch algorithm basically works to find the best matching between the Peaks (clusters) and Atoms (residues).
21. The Jenetics Genetic Algorithm library runs some of the algorithm in parallel on multiple-core computers for enhanced performance. You may notice your computer fans speed up as the CPU cores are fully utilized.
22. Two key files are output to the netmatch folder. The *finalscore.txt* file includes information about the chemical shift assigned to each atom and the *finalmatch.txt* file gives information about the cluster that matches to each residue.

Acknowledgments

This work was supported in part by a grant from the National Institute of General Medical Sciences of the National Institutes of Health (P50 GM 103297 to B.A.J.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Palmer AG III (2009) A topical issue: NMR investigations of molecular dynamics. *J Biomol NMR* 45(1–2):1–4. doi:10.1007/s10858-009-9345-8
2. Stark JL, Powers R (2012) Application of NMR and molecular docking in structure-based drug discovery. *Top Curr Chem* 326: 1–34. doi:10.1007/128_2011_213
3. Hafsa NE, Wishart DS (2014) CSI 2.0: a significantly improved version of the Chemical Shift Index. *J Biomol NMR* 60(2–3):131–146. doi:10.1007/s10858-014-9863-x
4. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293
5. Hoch J, Stern A (1985) The Rowland NMR Toolkit. Rowland Institute for Science Technical Memorandum RIS-18t
6. Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6(1): 1–10. doi:10.1007/BF00417486

7. Keller R CARA. <http://cara.nmr-software.org/portal/>
8. Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31(8):1325–1327. doi:10.1093/bioinformatics/btu830
9. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59(4):687–696. doi:10.1002/prot.20449
10. Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG, Vuister GW (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J Biomol NMR* 66(2):111–124. doi:10.1007/s10858-016-0060-y
11. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLoS Comput Biol* 5(3):e1000307. doi:10.1371/journal.pcbi.1000307
12. Grishaev A, Llinas M (2002) CLOUDS, a protocol for deriving a molecular proton density via NMR. *Proc Natl Acad Sci USA* 99(10):6707–6712. doi:10.1073/pnas.082114199
13. Jung YS, Sharma M, Zweckstetter M (2004) Simultaneous assignment and structure determination of protein backbones by using NMR dipolar couplings. *Angew Chem Int Ed Engl* 43(26):3479–3481. doi:10.1002/anie.200353588
14. Vitek O, Bailey-Kellogg C, Craig B, Kuliniewicz P, Vitek J (2005) Reconsidering complete search algorithms for protein backbone NMR assignment. *Bioinformatics* 21(Suppl 2):ii230–ii236. doi:10.1093/bioinformatics/bti1138
15. Bartels C, Billeter M, Guntert P, Wuthrich K (1996) Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J Biomol NMR* 7(3):207–213. doi:10.1007/BF00202037
16. Norris M, Fetler B, Marchant J, Johnson BA (2016) NMRFX Processor: a cross-platform NMR data processing program. *J Biomol NMR* 65(3–4):205–216. doi:10.1007/s10858-016-0049-6
17. Johnson B (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313–352. doi:10.1385/1-59259-809-9:313
18. Johnson BA, Blevins RA (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4(5):603–614
19. Cavanagh J, Fairbrother W, Palmer A, Rance M, Skelton N (2007) *Protein NMR spectroscopy: principles and practice*, 2nd edn. Academic, Amsterdam, The Netherlands
20. Rule GS, Hitchens TK (2006) *Fundamentals of protein NMR spectroscopy*. Focus on structural biology, Vol 5. Springer, Dordrecht
21. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6(2):135–140
22. Otting G, Widmer H, Wagner G, Wuthrich K (1986) Origin of t_1 and t_2 ridges in 2D NMR-spectra and procedures for suppression. *J Magn Reson* 66(1):187–193. doi:10.1016/0022-2364(86)90122-8
23. Sun S, Gill M, Li Y, Huang M, Byrd RA (2015) Efficient and generalized processing of multi-dimensional NUS NMR data: the NESTA algorithm and comparison of regularization terms. *J Biomol NMR* 62(1):105–117. doi:10.1007/s10858-015-9923-x
24. Kazimierczuk K, Orekhov V (2015) Non-uniform sampling: post-Fourier era of NMR data collection and processing. *Magn Reson Chem* 53(11):921–926. doi:10.1002/mrc.4284
25. Ying J, Delaglio F, Torchia DA, Bax A (2016) Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR*. doi:10.1007/s10858-016-0072-7
26. Hoch JC, Stern AS (1996) *NMR data processing*. Wiley-Liss, New York
27. Ulrich E, Akutsu H, Doreleijers J, Harano Y, Ioannidis Y, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte C, Tolmie D, Kent Wenger R, Yao H, Markley J (2008) BioMagResBank. *Nucleic Acids Res* 36 (Database issue):D402–D408. doi:10.1093/nar/gkm957
28. Friedrichs MS, Mueller L, Wittekind M (1994) An automated procedure for the assignment of protein ^1HN , ^{15}N , $^{13}\text{C}\alpha$, $^1\text{H}\alpha$, $^{13}\text{C}\beta$ and $^1\text{H}\beta$ resonances. *J Biomol NMR* 4(5):703–726
29. Lin G, Tegos T, Chen ZZ (2005) Heuristic search in constrained bipartite matching with applications to protein NMR backbone resonance assignment. *J Bioinforma Comput Biol* 3(6):1331–1350
30. Schmidt E, Guntert P (2012) A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc* 134(30):12817–12829. doi:10.1021/ja305091n

31. Jenetics: Java Genetic Algorithm Library, <http://jenetics.io>
32. Juneau J, Baker J, Wierzbicki F, Soto LM, Ng V (2010) *The definitive guide to Jython*. Springer, New York
33. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* 52(4):315–327
34. Becker S, Bobin J, Candès E (2011) NESTA: a fast and accurate first-order method for sparse recovery. *SIAM J Imaging Sci* 41:1–39
35. Wang L, Eghbalnia H, Bahrami A, Markley J (2005) Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications. *J Biomol NMR* 32(1):13–22. doi:[10.1007/s10858-005-1717-0](https://doi.org/10.1007/s10858-005-1717-0)

Protein Structure Elucidation from NMR Data with the Program Xplor-NIH

Guillermo A. Bermejo and Charles D. Schwieters

Abstract

Xplor-NIH is a popular software package for biomolecular structure determination from NMR (and other) experimental data. This chapter illustrates its use with the de novo structure determination of the B1 domain of streptococcal protein G (GB1), based on distances from nuclear Overhauser effects, torsion angles from scalar couplings, and bond-vector orientations from residual dipolar couplings. Including Xplor-NIH's latest developments, a complete structure calculation script is discussed in detail, and is intended to serve as a basis for other applications.

Key words NMR, Protein structure calculation, Xplor-NIH script

1 Introduction

In its simplest form, the biomolecular structure elucidation problem can be cast as follows: finding the three-dimensional all-atom model that optimally satisfies both the available experiments and general preconceived notions of what a good model should be (the so-called prior information). The latter involve covalent geometry (e.g., bond lengths and angles), nonbonded interactions (e.g., steric clashes), and conformation (e.g., protein side chain rotamers). The optimization is achieved by creating an energy expression, function of the molecular conformation, comprised of terms for both the experimental data and prior information, that is subsequently minimized using a conformational sampling technique. A large number of structural models (also simply referred to as “structures”) are independently calculated in this fashion. In the NMR field, it is standard practice to keep the k -lowest-energy structures (where k is typically 10 or 20) to indicate how tightly the system is defined by the data.

Xplor-NIH [1, 2] is a popular software package for biomolecular structure determination from experimental restraints, able to search conformational space with a variety of methods, including

molecular dynamics and gradient minimization. In active development since its introduction in 2002, Xplor-NIH has maintained and improved on the functionality of its predecessor XPLOR program [3], and added new functionality to take advantage of emerging experiments. Examples from NMR include improved treatment of residual dipolar coupling (RDC) data [2, 4] and paramagnetic relaxation enhancements [5], to name two. Increasingly, however, other experimental sources are being supported, such as small-angle X-ray and neutron scattering (reviewed in ref. [6]) and cryo-electron microscopy [7]. This experimental versatility, coupled with sophisticated (albeit easy-to-implement) modeling capabilities, such as the use of a conformational ensemble approach to account for the effects of molecular motion in the data [8], makes Xplor-NIH a powerful addition to the structural biologist's toolbox.

This chapter illustrates the use of Xplor-NIH with the *de novo* structure determination of the B1 domain of streptococcal protein G (GB1), based on NMR distances from nuclear Overhauser effects (NOEs) and inferred hydrogen bonds, torsion angles from scalar couplings, and bond-vector orientations from RDCs [9]. A structure calculation script is described in detail, and is intended to serve as a reference for other applications. Basic knowledge of the Python programming language is assumed.

2 Materials

2.1 Software

Xplor-NIH is freely available to academic and non-profit institutions (version 2.43 or higher is required to run the example script described below). It is officially supported on Unix/Linux and Mac-OS X operating systems, and can be run on Windows-based computers within a Linux virtual machine, such as VirtualBox (<https://www.virtualbox.org>), VMware (<https://www.vmware.com/products/workstation>), or QEMU (<http://qemu.org/>). Details on downloading, licensing, and installing Xplor-NIH can be found at <https://nmr.cit.nih.gov/xplor-nih>. Xplor-NIH provides a scripting interface based on the Python language (<http://python.org>).

2.2 Input Files

The Xplor-NIH structure calculation of a monomeric protein, such as the GB1 example considered in this chapter, requires the amino acid sequence and the available experimental data as inputs. The sequence (provided in a file named, e.g., `gb1.seq`) is expressed in the 3-character amino acid nomenclature: MET THR TYR LYS LEU ILE LEU ASN GLY LYS..., and is used to generate a PSF file (file extension `.psf`) that describes the specific covalent topology of the system (see below for details). The experimental data, in this case NOE-derived distances, torsion angles from scalar couplings, and

Table 1
GB1 NMR restraints

File name	Data description ^a	Restraint type
noe.tbl	NOEs and hydrogen bonds	Interatomic distance
dihedral.tbl	Scalar couplings	Torsion angle
bicelles_nh.tbl	RDCs for directly bound N–H ^N pairs in bicelle alignment medium	Bond-vector orientation
bicelles_hnc.tbl	RDCs for C _i –H _{i+1} ^N pairs in bicelle alignment medium	Bond-vector orientation
bicelles_nc.tbl	RDCs for C _i –N _{i+1} pairs in bicelle alignment medium	Bond-vector orientation
tmv107_nh.tbl	RDCs for directly bound N–H ^N pairs in phage alignment medium	Bond-vector orientation
tmv107_hnc.tbl	RDCs for C _i –H _{i+1} ^N pairs in phage alignment medium	Bond-vector orientation
tmv107_nc.tbl	RDCs for C _i –N _{i+1} pairs in phage alignment medium	Bond-vector orientation

^a*i* denotes the residue number

RDCs for different nuclear spin pairs in two molecular alignment media, are condensed into corresponding restraint files (arbitrarily named, although customarily ending with extension .tbl; see Table 1 for details). Entries (i.e., restraints) from each restraint file type are shown below as examples.

2.2.1 Distance Restraints

Distance restraint (from file noe.tbl):

```
assign (resid 2 and name HA) (resid 18 and name HN) 4.0 2.2 1.0
```

where the expressions in parenthesis (here and in all other restraint examples below) represent atom selections (in this case involving H^α of residue 2 and H^N of residue 18), and the sequence of numbers represents d , d_{minus} , and d_{plus} , respectively, which define the distance restraint bounds: $(d - d_{\text{minus}}, d + d_{\text{plus}})$. Atom selections can refer to more than one atom, for instance, in the cases of non-stereospecifically assigned NOEs or methyl groups [2]. In addition to directly NOE-derived restraints, such as the above example, there are distance restraints that enforce hydrogen bonds in secondary structure elements (e.g., as determined by characteristic NOE patterns). Each hydrogen bond is represented by two “hydrogen bond restraints” in order to loosely capture the proper hydrogen bond geometry. Here, an example hydrogen

bond between residue 26 and 30, in the middle of GB1's single α -helix (from file `noe.tbl`):

```
assign (resid 26 and name O) (resid 30 and name HN) 2.3 0.8 0.2
assign (resid 26 and name O) (resid 30 and name N ) 3.3 0.8 0.2
```

2.2.2 Torsion Angle Restraints

Torsion angle restraint (from file `dihedral.tbl`):

```
assign (resid 1 and name C) (resid 2 and name N)
      (resid 2 and name CA) (resid 2 and name C) 1.0 -105.0 40.0 2
```

where the atom selections define the torsion angle (in this case, ϕ of residue 2), the first number is a scale factor for the energy of this particular restraint, the second number is the target value (in degrees) to which the torsion is restrained, the third is a tolerance around the target torsion (in degrees), and the last number is the exponent that defines the shape of the energy bounds (in this case, quadratic). In general, the restraint scale factor and the exponent are uniformly given values of 1.0 and 2, respectively, in all torsion angle restraints. Although GB1's restraints are derived from scalar couplings, current common practice is to extract torsion angle restraints directly from chemical shifts (e.g., *see* ref. [10]).

2.2.3 RDC Bond-Vector Orientation Restraints

RDC restraint (directly bound N-H^N pair in bicelle alignment medium; from file `bicelles_nh.tbl`):

```
assign (resid 600 and name OO)
      (resid 600 and name Z )
      (resid 600 and name X )
      (resid 600 and name Y )
      (resid 2 and name N )
      (resid 2 and name HN) 2.2690 0.2000
```

where the first four selections (here involving a dummy residue 600) are ignored, are present solely for backwards compatibility with old XPLOR energy terms, and can be empty (although the parentheses must be present). The last two selections define the bond vector (in this case, the N-H^N of residue 2), and the two numbers are the experimental RDC value and its error (in Hz), respectively. RDCs associated with different alignment media involve independent alignment tensors, and must therefore be input in separate restraint files. Additionally, RDCs that stem from different nuclear spin pairs are also input in separate files, as it allows normalization relative to a reference spin pair (if RDCs have not been pre-normalized) and/or the assignment of different weights during structure calculations (see below).

3 Methods

3.1 Energy Function

Through changes in atomic positions Xplor-NIH minimizes the potential energy function

$$E_{\text{total}} = E_{\text{expt}} + E_{\text{prior}}. \quad (1)$$

E_{expt} is associated with the available experimental data; in the present example with protein GB1,

$$E_{\text{expt}} = w_{\text{dist}}E_{\text{dist}} + w_{\text{torsion}}E_{\text{torsion}} + w_{\text{RDC}}E_{\text{RDC}}, \quad (2)$$

where subscripts “dist,” “torsion,” and “RDC” refer to the distance, torsion angle, and RDC restraints, respectively, and w_j is the weight or energy scale for term j (as in all equations herein). Since the experimental data is insufficient to reliably define all atom positions, it is supplemented with information known a priori, via the E_{prior} energy term,

$$E_{\text{prior}} = w_{\text{bond}}E_{\text{bond}} + w_{\text{angle}}E_{\text{angle}} + w_{\text{impr}}E_{\text{impr}} + w_{\text{repel}}E_{\text{repel}} + \sum_{\text{db}} w_{\text{db}}E_{\text{db}}, \quad (3)$$

which relies on values of bond lengths, bond angles, improper dihedral angles that specify chirality and planarity of functional groups, and van der Waals radii (subscripts “bond,” “angle,” “impr,” and “repel,” respectively). Although not completely intractable (e.g., *see* ref. [11]), high computational expense and poor convergence during conformational sampling challenge a detailed description of electrostatic and van der Waals interactions. As a result, the latter are represented by a repulsive-only energy term (E_{repel} in Eq. 3) that simply discourages atomic overlap [12].

The remaining energy terms in E_{prior} (Eq. 3) are knowledge-based in nature (also known as empirical or statistical), derived from protein structure databases. Specifically,

$$\sum_{\text{db}} w_{\text{db}}E_{\text{db}} = w_{\text{torsionDB}}E_{\text{torsionDB}} + w_{\text{hbdb}}E_{\text{hbdb}}, \quad (4)$$

where E_{hbdb} represents a hydrogen bond potential used to improve backbone–backbone hydrogen bond geometry [13], and $E_{\text{torsionDB}}$ represents the torsionDB potential [14] that acts on torsion angles to improve backbone and side chain conformation.

The energy scales for experimental terms (Eq. 2) may require adjusting to the specific system at hand. For example, w_{dist} may be increased in cases of sparse NOE data. On the other hand, the relative scaling of terms within E_{prior} (Eq. 3) is carefully balanced and should not be modified.

3.2 Conformational Sampling

Minimization of the total potential energy (Eq. 1), starting from an arbitrary, extended conformation, results in a compactly folded

structure. The energy minimization is primarily achieved by molecular dynamics with simulated annealing, in a protocol that comprises the following stages:

1. *High Temperature Stage*. High temperature molecular dynamics in torsion angle space with small energy scales, and C $^{\alpha}$ -only repulsions.
2. *Simulated Annealing Stage*. Annealing in torsion angle space with energy scales ramped up, and all-atom repulsions.
3. *Torsion Angle Minimization*. Gradient minimization in torsion angle space.
4. *Cartesian Minimization*. Gradient minimization in Cartesian space.

The high temperature stage, with a relatively smooth energy surface and atoms able to go through each other, is intended for efficient sampling of conformational space. In the simulated annealing stage, the temperature is gradually decreased, while the energy surface becomes more restrictive, allowing the system to settle into an energy minimum. The final minimizations result in small conformational adjustments. A successful run yields a structure that fits the experimental restraints, has few or no atomic clashes, and does not deviate significantly from ideal covalent geometry.

3.3 Structure Calculation

Structure calculation protocols are implemented as Python scripts that take advantage of modules specifically written for Xplor-NIH. Below, the script used to calculate and analyze a set of GB1 structures is discussed. First, the outline of the script is presented, followed by a detailed description of the script itself.

3.3.1 Script Outline

The general script outline of an Xplor-NIH structure calculation protocol is as follows:

1. *System Definition*. Define the molecular covalent topology and associated covalent and van der Waals parameters.
2. *Extended Structure*. Generate a provisional extended molecular conformation.
3. *Important Collections*. Create empty collections to hold:
 - (a) All energy terms.
 - (b) Objects that handle changes (e.g., energy scale ramping) between and during the different stages of the protocol (as defined in Subheading 3.2 above).
4. *Energy Terms*. Set up each energy term and populate the above collections.

5. *Degrees of Freedom*. Set up objects that define degrees of freedom (e.g., torsion angle) allowed during the different stages of the protocol (as defined in Subheading 3.2).
6. *Structure Calculation*. Define the function that calculates a single structure.
7. *Structure Gathering*. Define the object that runs the above function N times and gathers statistics on the resulting structures.

While Python offers considerable flexibility in the implementation of the structure calculation protocol, this specific outline is dictated by adaptability of the script to other scenarios. For example, most actions associated with the use of a particular energy term are condensed into a snippet of adjacent lines of code, which facilitates the removal of the term when not applicable (e.g., when the associated experimental data is not available), and, similarly, the addition of extra energy terms. Thus, the same script can be easily reused for different purposes.

3.3.2 The Script

Below, the complete Python script, named `fold.py`, is presented in sections that correspond to those of the outline above. The script lines appear in Listings interspersed with the explanatory main text. A complete, uninterrupted version of this script can be found at the end of this chapter in a Supplemental Listing. The script can be executed from the command line by typing:

```
xplor -py fold.py > fold.out
```

where `fold.out` is the log file. All modules mentioned below, explicitly (in the main text) and/or implicitly (in the Listings), are further documented on the web at <http://nmr.cit.nih.gov/xplor-nih/doc/current/python/ref/>. In addition, a commented version of `fold.py`, along with all the files necessary to execute it, can be downloaded from <https://nmr.cit.nih.gov/xplor-nih/methodsMolBiol/>.

System Definition

The topology file is a library of amino acid definitions that specifies the atoms and their covalent connectivity, and provides “patches” for chemical modifications (e.g., peptide bond, disulfide bridge, N- and C-termini). The topology of a specific molecule is built from its sequence and stored in the PSF file. Although the PSF can be generated within the structure calculation script itself, it is recommended to create the PSF *ex situ* when the protein is not relatively standard (e.g., there are post-translational modifications or non-natural amino acids). The latter scenario is exemplified with protein GB1 (despite being a standard system) in **Note 1**, which

describes the generation of its PSF, `gbl.psf`. The `protocol` module provides a function to read in the PSF:

Listing 1

```
import protocol
protocol.initStruct("gbl.psf")
```

The topology file (and, consequently, the PSF) defines the atoms and covalent connectivity of the system, but does not contain specific values of bond lengths, bond angles, improper dihedral angles, and van der Waals radii (respectively required for energy terms E_{bond} , E_{angle} , E_{impr} , and E_{repel} ; *see* Eq. 3 above). Such information is provided by the associated parameter file, which is loaded as follows:

Listing 2

```
protocol.initParams("protein")
```

where the “protein” keyword specifies the use of the default parameter file for proteins. It is noteworthy that, as with any aspect of Xplor-NIH, topology and parameter files undergo updates; since old file versions are kept for backwards compatibility, reliance on the keyword (instead of the specific file name) ensures that the latest version is being used. (Alternatively, the file name `protein.par` is a shortcut to the latest parameter file.) The parameter file loaded must be consistent with the topology file used to create the PSF. In most instances these files come in pairs.

Extended Structure

Although the protein’s covalent topology and associated parameters are established at this point, atomic positions are still undefined (or “unknown” in Xplor-NIH terminology). While the structure calculation itself is still many lines away in the script, some energy terms require valid coordinates for their setup. Thus, an arbitrary extended conformation is generated with the function `genExtendedStructure`, from the `protocol` module. An arbitrary seed number initializes Xplor-NIH’s random number generator, which is used in the computation of this conformation and in subsequent molecular dynamics stages (*see* Note 2). These two operations are performed as follows:

Listing 3

```
protocol.initRandomSeed(3421)

protocol.genExtendedStructure()
```

Important Collections

A `PotList` instance (from the `potList` module) is an ordered collection (i.e., a sequence) specialized to hold objects that represent energy terms. It behaves itself as an energy term, so that, for example, a `PotList` that contains several RDC terms (e.g., associated with different restraint files) evaluates to the sum of its constituent terms. Although `PotList` instances are used throughout the script, a particularly central one is named `etotal`, created to group all the energy terms that will be used in the structure calculation (i.e., it corresponds to E_{total} in Eq. 1):

Listing 4

```
from potList import PotList
etotal = PotList()
```

Other important collections are the regular Python lists `highTempParams` and `rampedParams`:

Listing 5

```
from simulationTools import MultRamp, StaticRamp, InitialParams

highTempParams = [ ]
rampedParams = [ ]
```

They are used to handle (1) changes in both energy scales and atomic repulsion setups (e.g., C^α -only versus all-atom interactions) between the high temperature and simulated annealing stages, and (2) energy scale ramping within the simulated annealing stage itself. As its name suggests, objects added to `highTempParams` deal solely with the high temperature stage. Only instances of `MultRamp` or `StaticRamp` objects (imported here from the `simulationTools` module) will be added to `highTempParams` and `rampedParams`. The roles of the latter are exemplified below with the creation and configuration of the different energy terms, and should become completely clear when the structure calculation function is explained.

Energy Terms

The RDC energy term [2] requires the definition of the alignment tensor for each of the available media. The `tensors` dictionary, defined in Listing 6 below, maps a user-selected medium name (the key) to the corresponding tensor object (the value). In the generation of each tensor, an estimate of its axial component D_n (Da in the

RDC Restraints

Listing) and rhombicity (Rh in the Listing) (e.g., as obtained from the powder pattern method [15]; see Note 3) are provided:

Listing 6

```

from varTensorTools import create_VarTensor, calcTensorOrientation
tensors = {}
#           medium      Da      Rh
for (medium, Da, Rh) in [ ("tmv107", -6.5, 0.62),
                          ("bicelle", -9.9, 0.23) ]:
    tensor = create_VarTensor(medium)
    tensor.setDa(Da)
    tensor.setRh(Rh)
    tensors[medium] = tensor

```

Each molecular alignment tensor is represented by seven pseudoatoms, which describe the orientation, D_a , and the rhombicity. These pseudoatoms do not contact the protein (i.e., they are “van der Waals-invisible”), are generated automatically, and, thus, can be almost ignored. They are, however, present at the end of each output coordinate file, and can be helpful to understand and communicate the nature of the alignment, for instance, by visualizing the tensor orientation relative to the structure.

Next, a `StaticRamp` instance is added to `highTempParams`, instructing the function `calcTensorOrientation` (from the `varTensorTools` module) to optimize the orientation of each alignment tensor prior to the high temperature stage:

Listing 7

```

highTempParams.append(StaticRamp(" "
for medium in tensors.values():
    calcTensorOrientation(medium)
" "))

```

The “static” in `StaticRamp`’s name refers to the fact that the snippet of code used in its instantiation (i.e., the `for` loop) will be run as is (as opposed to a `MultiRamp` instance). This will become obvious below.

Next, the list `rdcData` is created, packed with the information associated with each input RDC restraint file (see Table 1). Each item within this list is a tuple that contains the following information (in this order): the medium’s name, an arbitrary user-selected experiment name (e.g., “NH” for N–H^N RDCs), the path name of the restraint file, and a relative scale factor.

Listing 8

```
#          medium    expt.  restraint file    scale
rdcData = [("tmv107", "NH", "tmv107_nh.tbl", 1),
           ("tmv107", "NCO", "tmv107_nc.tbl", 0.05),
           ("tmv107", "HNC", "tmv107_hnc.tbl", 0.108),
           ("bicelle", "NH", "bicelles_nh.tbl", 1),
           ("bicelle", "NCO", "bicelles_nc.tbl", 0.05),
           ("bicelle", "HNC", "bicelles_hnc.tbl", 0.108)]
```

The medium's name is used to identify the corresponding alignment tensor (i.e., it must be one of those in Listing 6). In this example, the RDCs in the restraint files have been normalized relative to the N–H^N spin pair, which increases the apparent magnitudes of the less sensitive (therefore, more error prone) non-N–H^N RDCs. So that the latter do not dominate the calculation, their respective contribution to the RDC energy term will be scaled down by the square of the inverse of the normalization factor. These relative scale constants are specified in the right-most column in Listing 8. (To use unnormalized RDCs in an Xplor-NIH calculation *see* Note 4.)

The RDC energy term itself is created next, by looping through each item in `rdcData`, and creating an individual energy term associated with the data therein:

Listing 9

```
from rdcPotTools import create_RDCPot, scale_toNH
rdcs = PotList("rdc")
for (medium, exp, table, scale) in rdcData:
    name = "%s_%s" % (exp, medium)
    rdc = create_RDCPot(name, table, tensors[medium])
    rdc.setScale(scale)
    # scale_toNH(rdc) # uncomment if unnormalized restraints
    rdcs.append(rdc)
etotal.append(rdcs)
rampedParams.append(MultRamp(0.01, 1.0, "rdcs.setScale(VALUE)"))
```

Specifically, for each loop iteration, a name is generated by combining the experiment and medium names (useful for reporting statistics later on), in turn, used in the creation of an RDC energy object. The iteration is concluded by setting the energy term's relative scale and adding the term to `rdcs` (a `PotList`). At loop exit, `rdcs` involves the entire RDC dataset and, thus, represents the complete RDC energy term (i.e., E_{RDC} in Eq. 2), which is added to `etotal`. Appended to `rampedParams` is a `MultRamp` instance, which defines

a range of values (from 0.01 to 1.0) for the overall RDC energy scale (w_{RDC} in Eq. 2). Throughout the high temperature stage, the energy scale will be given its “initial” value of 0.01 (i.e., the range’s minimum), whereas during the simulated annealing stage, it will be geometrically increased from this minimum to the maximum value of 1.0.

Distance Restraints

The NOE energy term is used to implement distance restraints (E_{dist} in Eq. 2). As opposed to the RDC term above, no loop is required to set up the NOE term because all GBI distance restraints are encapsulated in a single file:

Listing 10

```
import noePotTools
noe = noePotTools.create_NOEPot(name="noe", file="noe.tbl")
etotal.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale(VALUE)"))
```

With this caveat, the setup is very similar to that of the RDC term. The energy object is created by giving it a user-selected name and the path name of the restraint file. After being added to `etotal`, the term is further configured within `rampedParams` for energy scale handling, as explained above with the RDC term.

Torsion Angle Restraints

Xplor-NIH inherits facilities from the older XPLOR program on which it is based. With continued development of Xplor-NIH, older XPLOR facilities have been deprecated in favor of rewritten versions, such as the RDC and NOE energy terms discussed above (see Note 5). The legacy XPLOR energy terms, however, can still be accessed from Xplor-NIH Python scripts (via the `xplorPot` module), which is useful for terms that have no native counterpart in the Python interface. This is the case of the torsion angle restraint term (E_{torsion} in Eq. 2) (and other terms below):

Listing 11

```
from xplorPot import XplorPot
dihedralTable = "dihedral.tbl"
protocol.initDihedrals(dihedralTable)
etotal.append(XplorPot("CDIH"))
highTempParams.append(StaticRamp("etotal['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("etotal['CDIH'].setScale(200)"))
```

To facilitate reuse of the restraint file’s path name (also required below), it is assigned to a variable (`dihedralTable`). The call to the `protocol.initDihedrals` function sets up the term, which is

added to `etotal` as indicated. As opposed to the RDC and NOE terms above, where the energy scale is set to be uniform throughout the high temperature stage and ramped during the simulated annealing stage, here, it is configured to hold (different) uniform values in both stages. This is achieved by invoking the `highTempParams` list, to which a `StaticRamp` object is appended, specifying the constant value (of 10) for the high temperature stage. Similarly, the `StaticRamp` instance added to `rampedParams` specifies the energy scale value (of 200) for the entire annealing stage.

Knowledge-Based Energy Terms

With all the energy terms associated with experimental data, E_{expt} (Eq. 2), already configured, the first term in E_{prior} (Eq. 3) that we encounter is the knowledge-based hydrogen bond term E_{hbdb} (from Eq. 4) [13]:

Listing 12

```
protocol.initHBDB()
etotal.append(XplorPot("HBDB"))
```

Since the overall energy scale of this term remains uniform throughout the entire protocol (i.e., w_{hbdb} in Eq. 4 is always one), there is no need for entries in `highTempParams` and/or `rampedParams`. The term is added to `etotal` as usual.

Next, the knowledge-based torsion angle energy term, `torsionDB` ($E_{\text{torsionDB}}$ in Eq. 4) [14], is configured:

Listing 13

```
import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name="torsiondb",
                                                    system="protein")
etotal.append(torsiondb)
rampedParams.append(MultRamp(0.002, 2, "torsiondb.setScale(VALUE)"))
```

The similarities with the NOE energy term setup (Listing 10 above) are obvious. It is noteworthy that, since there is a version of this term for RNA [16] and DNA (unpublished), the `system` argument is used to choose the correct application upon the term's creation (i.e., "protein" for GB1).

Atomic Repulsions

The generation of the atomic repulsion energy term and its addition to `etotal` is performed as follows:

Listing 14

```

from repelPotTools import create_RepelPot,  initRepel
repel  = create_RepelPot("repel")
etotal.append(repel)

```

The behavior of this term throughout the structure calculation protocol is among the most complicated, as it not only involves changes in its energy scale (exemplified with previously discussed terms) but also in the atomic interactions that are active. During the high temperature stage, only repulsions between “inflated” C^α atoms are wanted in order to promote efficient conformational sampling. Since this configuration will not change throughout that stage, it is specified with a `StaticRamp` instance, correspondingly added to `highTempParams`:

Listing 15

```

highTempParams.append(StaticRamp( """initRepel(repel,
                                           use14=True,
                                           scale=0.004,
                                           repel=1.2,
                                           moveTol=45,
                                           interacting
                                           Atoms='name CA'
                                           ) """))

```

Here, noteworthy arguments to the `initRepel` function are `use14`, `repel`, and `interactingAtoms`, respectively specifying the use of 1–4 interactions (i.e., between atoms separated by three bonds), van der Waals radii (stored in the parameter file) enlarged by a factor of 1.2, and C^α repulsions only. Note that throughout the text, it is implied that interactions of (1) order lower than 1–4 are never active, and (2) order higher than 1–4 are always active.

During the simulated annealing stage, it is required for all atoms, exhibiting more realistic radii, to repel each other (excluding 1–4 interactions; see below for an explanation). Since this configuration remains unchanged during that stage, it is specified with a `StaticRamp` instance, added to `rampedParams`:

Listing 16

```

rampedParams.append(StaticRamp("initRepel(repel,use14=False)"))

```

The energy scale factor is familiarly configured (e.g., as in the RDC term):

Listing 17

```
rampedParams.append(MultRamp(0.004, 4, "repel.setScale(VALUE)"))
```

which specifies the geometrical increase during the simulated annealing stage from 0.004 to 4.

The torsionDB energy term introduced above (Listing 13) affects all the heavy-atom-based torsion angles of a protein (ϕ , ψ , χ_1 , \dots , χ_4) [14], hence the exclusion of 1–4 repulsions within the simulated annealing stage (implied in Listing 16 above with the `use14` argument). However, torsion angles of terminal, protonated groups, such as that involved in methyl rotation, are not considered by torsionDB. Thus, to prevent eclipsed conformations in such groups, the corresponding 1–4 interactions are introduced with an extra, repulsive energy term:

Listing 18

```
import torsionDBPotTools
repel14 = torsionDBPotTools.create_Terminal14Pot("repel14")
etotal.append(repel14)
highTempParams.append(StaticRamp("repel14.setScale(0)"))
rampedParams.append(MultRamp(0.004, 4, "repel14.setScale(VALUE)"))
```

The `StaticRamp` instance within `highTempParams` nullifies the term during the high temperature stage so that only the wanted C^α repulsions are active.

The two repulsive terms just discussed embody E_{repel} in Eq. 3.

*Bond Length, Bond Angle,
and Improper Dihedral
Energy Terms*

The remaining energy terms needed to complete E_{prior} (Eq. 3) and E_{total} (Eq. 1) are those that involve bond lengths, bond angles, and improper dihedral angles (to define group chirality and planarity), which are, respectively, set up below without introducing any new concepts:

Listing 19

```
etotal.append(XplorPot("BOND"))

etotal.append(XplorPot("ANGL"))
rampedParams.append(MultRamp(0.4, 1.0, "etotal['ANGL'].setScale(VALUE)"))

etotal.append(XplorPot("IMPR"))
rampedParams.append(MultRamp(0.1, 1.0, "etotal['IMPR'].setScale(VALUE)"))
```

Degrees of Freedom

Having already defined and set up all the energy terms, we focus on the configuration of the degrees of freedom used in the conformational search. This requires the introduction of `IVM` objects (from the `ivm` module, where the initials stand for Internal Variable Module [17]), which are central for performing molecular dynamics and gradient minimization. Briefly anticipating its functions, an `IVM` object works as follows: it is given particular degrees of freedom (e.g., torsion angle), associated with energy terms, set up according to a conformational sampling technique (e.g., for molecular dynamics: define the temperature, duration, etc.), and, finally, it is “run,” which performs the desired sampling operation. At this point in the script, however, the `IVM` objects are just created and assigned degrees of freedom, pending further configuration later on.

The protein evolves in torsion angle space during the entire protocol (*see* **Note 6**), except for the final Cartesian minimization stage. Thus, two `IVM` instances are created, starting with that configured for torsion angle space:

Listing 20

```
from ivm import IVM
dyn = IVM()

for tensor in tensors.values():
    tensor.setFreedom("fixDa, fixRh") # fix tensor Rh and Da (vary orientation)

protocol.torsionTopology(dyn)
```

where `dyn` is the `IVM` object instance. An `IVM` object controls the degrees of freedom of all the atoms in the system. As previously described, each RDC molecular alignment tensor is represented by a set of (pseudo)atoms, whose degrees of freedom—and, consequently, the variables they represent, namely the orientation, D_a , and rhombicity—are therefore under the purview of the `IVM` object. Since `dyn` will be used from the start of the structure calculation, when the protein conformation is far from optimal, the D_a and rhombicity of each tensor are fixed to their initial values (input in Listing 6), as allowing them to vary in an attempted optimization would likely distort them. This is achieved inside the above loop. Finally, the function `protocol.torsionTopology` configures the protein’s torsional setup.

The IVM instance reserved for the final Cartesian minimization stage, `minc`, is introduced next:

Listing 21

```
minc = IVM()

for tensor in tensors.values():
    tensor.setFreedom("varyDa, varyRh") # allow all tensor parameters to float

protocol.cartesianTopology(minc)
```

Since the overall protein shape is not expected to change significantly during that stage, the D_a and rhombicity of each alignment tensor are allowed to vary in order to optimize them.

Finally, all atoms are given a uniform mass and friction coefficient:

Listing 22

```
protocol.massSetup()
```

Structure Calculation

The calculation of a single structure is specified by the function `calcOneStructure` defined below. It relies on two useful variables, which correspond to the initial and final temperatures of the system:

Listing 23

```
temp_ini = 3500.0
temp_fin = 25.0
```

The initial temperature of 3500 K (`temp_ini` variable) is that of the high temperature stage, and the starting temperature of the simulated annealing stage. The final temperature of 25 K (`temp_fin` variable) is that at the end of the simulated annealing stage.

At this juncture, the protein atoms assume an arbitrary, extended conformation with satisfied covalent geometry, as produced by the function `protocol.genExtendedStructure` early on in the script (see Listing 3). The initial actions of `calcOneStructure` further modify the starting atomic coordinates in preparation for the high temperature stage. First, all torsion angles defined in the `dyn` IVM instance are randomized:

Listing 24

```
def calcOneStructure(loopInfo):
    """Calculate a single structure."""

    from monteCarlo import randomizeTorsions
    randomizeTorsions(dyn)
```

Second, each torsion angle with an associated torsion angle restraint is set to the restraint's target value:

Listing 25

```
import torsionTools
torsionTools.setTorsionsFromTable(dihedralTable)
```

[Here is where the `dihedralTable` variable, assigned to the restraint file path name in Listing 11 above, becomes useful.] Finally, it is ensured that the covalent geometry is satisfied:

Listing 26

```
protocol.fixupCovalentGeom(maxIters=100, useVDW=True)
```

The reason to start with the correct covalent geometry is that it cannot be improved in the calculations that take place in torsion angle space (i.e., all but the final Cartesian minimization) (*see Note 6*). It is also noteworthy that while the function `protocol.fixupCovalentGeom` silently performs molecular dynamics and minimization using the covalent energy terms and atomic repulsions, the latter (triggered by the `useVDW` argument) are only enabled sporadically, not necessarily enough to remove atomic clashes. Indeed, the purpose of this function (as well as `protocol.genExtendedStructure`) is to correct the covalent geometry and not atomic contacts. This is certainly not an issue because the generated structure is intended to go through the high temperature molecular dynamics stage, where atomic repulsions are also limited.

Before the start of the high temperature stage, the appropriate energy scale values and atomic repulsive interactions have to be established. This is achieved in the following two lines, which clearly reveal the inner workings of lists `highTempParams` and `rampedParams`:

Listing 27

```
InitialParams(rampedParams)
InitialParams(highTempParams)
```

The `InitialParams` function (discussed for the first time here, but originally imported from the `simulationTools` module in Listing 5 above) invokes the code snippet specified in each `MultRamp` or `StaticRamp` instance provided to it in an input list (here, `highTempParams` or `rampedParams`); in the case of a `MultRamp` instance, it does so with the first value in the associated range. The above Listing performs `InitialParams(highTempParams)` *after* `InitialParams(rampedParams)`. The effect is to override settings of a common parameter in `rampedParams` by those in `highTempParams`. For example, `InitialParams(rampedParams)` configures repulsive interactions between all atom types, which is immediately reconfigured to the C^α-only mode by `InitialParams(highTempParams)`, as desired for the high temperature stage. By contrast, `InitialParams(rampedParams)` sets the RDC energy scale to the initial value of the corresponding `MultRamp` instance, a setting that persists because it is not involved in `highTempParams`.

The stage is now set to configure the `IVM` object for torsion angle molecular dynamics. The object `dyn` has already been set up for torsion angle degrees of freedom (*see* Listing 20 above); it remains for the molecular dynamics details to be specified:

Listing 28

```
protocol.initDynamics(dyn,
    potList=etotal,
    bathTemp=temp_ini,
    initVelocities=True,
    finalTime=100,
    numSteps=1000,
    printInterval=100)

dyn.setETolerance(temp_ini/100) # used to set step size (default: temp/1000)
```

Here, with help from the function `protocol.initDynamics`, the input `dyn` is associated with the energy terms in `etotal` (via the `potList` argument), the initial temperature (via the `bathTemp` argument), and a dynamics duration (whichever is reached first: 100 ps, as specified by the `finalTime` argument, or 1000 molecular dynamics steps, as specified by the `numSteps` argument). In addition, it is directed that the initial atomic velocities be randomized (via the `initVelocities` argument) to values consistent

with the temperature value. The `printInterval` argument specifies that every 100 steps of molecular dynamics a report is printed in the output log file, which details the energy values of each term in `eTotal`, the kinetic energy, size of the molecular dynamics step and the current time.

An IVM object's `eTolerance` value (set on the above Listing's last line) controls the level of energy conservation to follow when determining the time step size [17]. During high temperature dynamics this value is increased from its default of $T/1000$ (where T is the temperature) so that conformational space is more widely sampled.

Finally, the IVM object is run, thus performing the molecular dynamics task for which it has been configured:

Listing 29

```
dyn.run()
```

Since the subsequent simulated annealing stage also relies on torsion angle molecular dynamics, the `dyn` IVM object can be repurposed. Here, however, the temperature will be lowered in discrete steps, and `dyn` run for each one of them. With this in mind, `dyn`'s configuration follows:

Listing 30

```
protocol.initDynamics(dyn,
                      finalTime=0.2,
                      numSteps=100,
                      printInterval=100)
```

The configuration is similar to that previously performed (*see* Listing 28 above), except that (1) atomic velocities are taken from the end of the previous molecular dynamics stage (the default behavior of the `initVelocities` argument, implied by its absence), (2) the energy terms are inherited from the previous configuration (`eTotal`) and, therefore, the `potList` argument is left unspecified, and (3) temperature is not specified, in anticipation of its variability within the annealing schedule. The latter is set up and run via a newly introduced object:

Listing 31

```
from simulationTools import AnnealIVM
AnnealIVM(initTemp=temp_ini,
          finalTemp=temp_fin,
          tempStep=12.5,
          ivm=dyn,
          rampedParams=rampedParams).run()
```

where the `dyn` IVM instance is provided (via the `ivm` argument), and the initial and final temperatures are specified (via the `initTemp` and `finalTemp` arguments, respectively), along with the temperature step (via the `tempStep` argument). That is, `dyn` is run with its latest setup (in Listing 30 above) for each of the $(3500 - 25)/12.5$ values of temperature. The `rampedParams` list is provided (via the `rampedParams` argument). For a `MultiRamp` instance in `rampedParams`, the corresponding code snippet is invoked at the beginning of each step (i.e., before `dyn` is run), with the involved value geometrically increased in a stepwise fashion using the associated range; this is the case of energy scales such as, for example, that of the RDC term, which geometrically increases from 0.01 to 1.0 throughout the annealing schedule. For a `StaticRamp` instance in `rampedParams`, the corresponding code snippet is simply invoked as it is, at the beginning of each step; this is the case of the torsion angle restraints' energy scale, which remains at a constant value of 200, as well as the configuration of atomic repulsions (except for their energy scaling).

Gradient minimization in torsion angle space follows, not surprisingly relying on the `dyn` IVM object, which is configured for the Powell algorithm and run:

Listing 32

```
protocol.initMinimize(dyn,
                      printInterval=50)
dyn.run()
```

Gradient minimization in Cartesian space is the final action of the `calcOneStructure` function. It relies on the IVM instance `minc`, already configured with Cartesian degrees of freedom (in Listing 21 above) for this purpose. The same function `protocol.initMinimize` above can be used to set up `minc` for Powell minimization, which is subsequently run:

Listing 33

```
protocol.initMinimize(minc,
                      potList=etotal,
                      dEPred=10)
minc.run()
```

Unlike `dyn`, `minc` was not assigned the applicable energy terms (`etotal`) prior to minimization, hence the need to do it here (via the `potList` argument). The `dEPred` argument provides the Powell conjugate gradient algorithm [18] an estimate of the expected drop in the energy. It has been found that specifying a

value of 10 for Cartesian minimization results in better optimization (the value would otherwise be 10^{-3}).

Structure Gathering

At this point, the end of indentation in the script lines of the following Listing signifies that the `calcOneStructure` function definition is finished. It remains to run `calcOneStructure` N times, and generate statistics on the resulting structures, a task conveniently handled by a `StructureLoop` object (from the `simulationTools` module), which is setup and run as follows:

Listing 34

```

from simulationTools import StructureLoop
StructureLoop(numStructures=100,
              structLoopAction=calcOneStructure,
              doWriteStructures=True,
              # Arguments for generating structure statistics:
              genViolationStats=True,
              averageSortPots=[etotal["BOND"], # terms for structure sorting.
                              etotal["ANGL"],
                              etotal["IMPR"], noe, rdcs, etotal["CDIH"]],
              averageTopFraction=0.1, # top fraction of structs. to report on.
              averagePotList=etotal, # terms analyzed.
              averageFitSel= "not (name H * or PSEUDO)", # selection to fit...
              ).run() # to average structure...
                   # and report precision.

```

In this instantiation two kinds of arguments are noted, related to either the structure calculation or the subsequent structural analysis. `numStructures`, `structLoopAction`, and `doWriteStructures` belong to the former type, and, respectively specify the total number of structures (N), the function used in the calculation of each of them (i.e., `calcOneStructure`), and whether the corresponding coordinate files (in PDB format) will be written. Each coordinate file is accompanied by an analysis file with the same name, except that the `.viols` extension is appended. This file contains information for each term in argument `averagePotList` (typically, all the terms used in the structure calculation, i.e., `etotal`), including detailed information about violations in each term, atomic clashes, etc. The more global structural analysis is triggered by the `genViolationStats` argument, and is based on the top fraction of structures specified by `averageTopFraction`, after all structures have been sorted from low to high energy using the terms in `averageSortPots`. The results of the analysis are stored in a file with the `.stats` extension, which reports on all the energy terms specified in `averagePotList`. The `.stats` file contains averages of fit metrics for each energy term, and a list of the most violated restraints for each energy term.

`averagePotList` is also used to compute a regularized average structure to which the analyzed structures are fit (using the selection string given in the argument `averageFitSel`) to report the coordinate precision. Although not exercised here, the average structure can be written out by specifying the extra argument `averageFilename` to `StructureLoop`. `StructureLoop` transparently handles simultaneous parallel calculation of multiple structures either on a local computer with multiple cores or on a cluster of computers. It is noteworthy that, after the initial folding, as done here with GB1, structures are usually refined by running a similar script (*see* **Note 7** for details).

4 Notes

1. The PSF file of a standard protein, such as GB1, can be generated directly from the command line, by using one of the helper programs provided with Xplor-NIH (within the `bin` directory of Xplor-NIH's distribution package). Assuming GB1's amino acid sequence is stored in the file `gb1.seq` (*see* Subheading 2.2 for more details), the PSF, `gb1.psf`, is generated simply by typing:

```
seq2psf gb1.seq
```

The command automatically recognizes the system type (in this case, protein), and uses the corresponding default topology file (i.e., the latest version). Alternatively, when a PDB coordinate file is available (e.g., representing a conformation different from that under study), it can be similarly used to generate the PSF with the helper program `pdb2psf`. Additional examples of PSF generation, including less standard cases, can be found in Xplor-NIH's distribution package, within the directory `eginput/PSF_generation`.

2. Use of the same seed number on the same computer yields the same results.
3. Initial values for the alignment tensor's D_a and rhombicity can be obtained using a maximum likelihood approach [19] which is implemented in the Xplor-NIH helper program `calcDaRh`. For example, based on the RDCs measured in bicelle media for GB1, this program can be run from the command line by typing:

```
calcDaRh -normtype none bicelles_nh.tbl bicelles_hnc.tbl bicelles_nc.tbl
```

The `-help` option of the `calcDaRh` command can be invoked for more details.

4. It is frequently simpler to use unnormalized RDCs. In such case, the relative energy scales in Listing 8 should be set to one, and the `scale_toNH` line in Listing 9 should be

uncommented. Additionally, the Xplor-NIH convention is for the gyromagnetic ratio of ^{15}N to be effectively treated as positive, when it is in fact negative. If the RDC restraints have not been prepared using the Xplor-NIH convention, the function `correctGyromagneticSigns` (from the `rdcPotTools` module) should be called before `create_RDCPot` in Listing 9.

5. Care has been taken to ensure that restraint formats supported by the old XPLOR program [3] are compatible in the newer Xplor-NIH [1, 2].
6. It is tempting (and erroneous) to think that when an `IVM` object (i.e., an instance of `ivm.IVM`) is configured with torsion angle degrees of freedom, as is the case with `dyn` in the main text (*see* Listing 20 above), there is no need to include covalent energy terms (i.e., for bond lengths and angles, and improper dihedral angles) in molecular dynamics or minimization. Indeed, while a great part of the covalent geometry remains unchanged in torsion angle space, making such energy terms unnecessary, some functional groups, however, require them. This is the case for conformationally flexible rings, such as that of proline residues, which require one endocyclic bond to behave as a spring (as restrained by the bond energy term) instead of being rigid. (Note that bond angles and improper dihedral angles involving the atoms in the mentioned bond also need to be restrained by the corresponding terms.) In addition, the default torsion angle configuration of an `IVM` object (applied to `dyn` via the `protocol.torsionTopology` function in Listing 20) allows flexibility in the peptide ω angle, increasing the reliance on covalent energy terms (this behavior is controlled by the optional `fixedOmega` argument of the above function).
7. A script for refinement of a reasonable starting structure is quite similar to `fold.py` presented above. The per-Listing differences are enumerated here:
 - (a) The call to `protocol.genExtendedStructure` in Listing 3 should be replaced by `protocol.initCoords(file.pdb)`, where `file.pdb` is an input structure file to be refined.
 - (b) The call to `calcTensorOrientation` in Listing 7 should be replaced with a call to function `calcTensor`, after importing it from the `varTensorTools` module. This function computes the D_a , rhombicity, and orientation of an input alignment tensor object via singular value decomposition [20]. This full calculation (as opposed to just the orientation with `calcTensorOrientation`) is possible in the refinement case because a reasonable

- starting structure is available at the beginning of the calculation.
- (c) The initial temperature (`temp_ini` in Listing 23) should be lowered to 3000 K.
 - (d) The calls to `randomizeTorsions` in Listing 24 and to `setTorsions` in Listing 25 should be removed.
 - (e) Finally, if one desires to refine more than one structure, a glob wildcard expression can be specified by adding the `pdbFilesIn` argument of `StructureLoop` in Listing 34 (e.g., `pdbFilesIn=fold*.best` would refine all the structures whose file names match the pattern `fold*.best`).

A refinement script is included in the file package associated with this chapter (downloadable from the web at <https://nmr.cit.nih.gov/xplor-nih/methodsMolBiol/>).

Supplemental Listing: File `fold.py` (a commented version can be downloaded from the web at <https://nmr.cit.nih.gov/xplor-nih/methodsMolBiol/>).

```
import protocol
protocol.initStruct("gbl.psf")
protocol.initParams("protein")
protocol.initRandomSeed(3421)
protocol.genExtendedStructure()

from potList import PotList
etotal = PotList()

from simulationTools import MultRamp, StaticRamp, InitialParams

highTempParams = [ ]
rampedParams = [ ]

from varTensorTools import create_VarTensor, calcTensorOrientation
tensors = {}
#           medium      Da      Rh
for (medium, Da, Rh) in [("tmv107", -6.5, 0.62),
                        ("bicelle", -9.9, 0.23)]:
    tensor = create_VarTensor(medium)
    tensor.setDa(Da)
    tensor.setRh(Rh)
    tensors[medium] = tensor
```

(continued)

```

highTempParams.append(StaticRamp("""
for medium in tensors.values():
    calcTensorOrientation(medium)
"""))

#           medium      expt.  restraint file      scale
rdcData = [("tmv107",   "NH",   "tmv107_nh.tbl",    1),
           ("tmv107",   "NCO", "tmv107_nc.tbl",   0.05),
           ("tmv107",   "HNC", "tmv107_hnc.tbl",  0.108),
           ("bicelle",  "NH",   "bicelles_nh.tbl", 1),
           ("bicelle",  "NCO", "bicelles_nc.tbl", 0.05),
           ("bicelle",  "HNC", "bicelles_hnc.tbl", 0.108)]

from rdcPotTools import create_RDCPot, scale_toNH
rdcs = PotList("rdc")
for (medium, exp, table, scale) in rdcData:
    name = "%s_%s" % (exp, medium)
    rdc = create_RDCPot(name, table, tensors[medium])
    rdc.setScale(scale)
    # scale_toNH(rdc) # uncomment if unnormalized restraints
    rdcs.append(rdc)
etotal.append(rdcs)
rampedParams.append(MultRamp(0.01, 1.0, "rdcs.setScale(VALUE)"))

import noePotTools
noe = noePotTools.create_NOEPot(name="noe", file="noe.tbl")
etotal.append(noe)
rampedParams.append(MultRamp(2, 30, "noe.setScale(VALUE)"))

from xplorPot import XplorPot
dihedralTable = "dihedral.tbl"
protocol.initDihedrals(dihedralTable)
etotal.append(XplorPot("CDIH"))
highTempParams.append(StaticRamp("etotal['CDIH'].setScale(10)"))
rampedParams.append(StaticRamp("etotal['CDIH'].setScale(200)"))

protocol.initHBDB()
etotal.append(XplorPot("HBDB"))

import torsionDBPotTools
torsiondb = torsionDBPotTools.create_TorsionDBPot(name="torsiondb",
                                                    system="protein")
etotal.append(torsiondb)
rampedParams.append(MultRamp(0.002, 2, "torsiondb.setScale(VALUE)"))

```

(continued)

```
from repelPotTools import create_RepelPot, initRepel
repel = create_RepelPot("repel")
etotal.append(repel)

highTempParams.append(StaticRamp("""initRepel(repel,
                                             use14=True,
                                             scale=0.004,
                                             repel=1.2,
                                             moveTol=45,
                                             interactingAtoms='name CA'
                                             ) """))

rampedParams.append(StaticRamp("initRepel(repel, use14=False)"))
rampedParams.append(MultRamp(0.004, 4, "repel.setScale(VALUE)"))

import torsionDBPotTools
repel14 = torsionDBPotTools.create_Terminal14Pot("repel14")
etotal.append(repel14)
highTempParams.append(StaticRamp("repel14.setScale(0)"))
rampedParams.append(MultRamp(0.004, 4, "repel14.setScale(VALUE)"))

etotal.append(XplorPot("BOND"))

etotal.append(XplorPot("ANGL"))
rampedParams.append(MultRamp(0.4, 1.0, "etotal['ANGL'].setScale(VALUE)"))

etotal.append(XplorPot("IMPR"))
rampedParams.append(MultRamp(0.1, 1.0, "etotal['IMPR'].setScale(VALUE)"))

from ivm import IVM
dyn = IVM()

for tensor in tensors.values():
    tensor.setFreedom("fixDa, fixRh") # fix tensorRh and Da (vary orientation)

protocol.torsionTopology(dyn)

minc = IVM()

for tensor in tensors.values():
    tensor.setFreedom("varyDa, varyRh") # allow all tensor parameters to float

protocol.cartesianTopology(minc)

protocol.massSetup()
```

(continued)

```

temp_ini=3500.0
temp_fin=25.0

def calcOneStructure ( loopInfo ):
    """Calculate a single structure. """

    from monteCarlo import randomizeTorsions
    randomizeTorsions (dyn)

    import torsionTools
    torsionTools.setTorsionsFromTable (dihedralTable)

    protocol.fixupCovalentGeom(maxIters=100, useVDW=True)

    InitialParams (rampedParams)
    InitialParams (highTempParams)

    protocol.initDynamics (dyn,
                           potList=etotal,
                           bathTemp=temp_ini,
                           initVelocities=True,
                           finalTime=100,
                           numSteps=1000,
                           printInterval=100)

    dyn.setETolerance (temp_ini / 100) # used to set step size (default: temp/1000)

    dyn.run ()

    protocol.initDynamics (dyn,
                           finalTime=0.2,
                           numSteps=100,
                           printInterval=100)

    from simulationTools import AnnealIVM
    AnnealIVM (initTemp=temp_ini,
               finalTemp=temp_fin,
               tempStep=12.5,
               ivm=dyn,
               rampedParams=rampedParams).run ()

    protocol.initMinimize (dyn,
                           printInterval=50)

    dyn.run ()

```

(continued)

```

protocol.initMinimize(minc,
                      potList=etotal,
                      dEPred=10)

minc.run()

from simulationTools import StructureLoop
StructureLoop(numStructures=100,
              structLoopAction=calcOneStructure,
              doWriteStructures=True,
              # Arguments for generating structure statistics:
              genViolationStats=True,
              averageSortPots=[etotal["BOND"], # terms for structure sorting.
                               etotal["ANGL"],
                               etotal["IMPR"], noe, rdc, etotal["CDIH"]],
              averageTopFraction=0.1, # top fraction of structs. to report on.
              averagePotList=etotal, # terms analyzed.
              averageFitSel="not (name H* or PSEUDO)", # selection to fit...
              ).run() # to average structure...
                    # and report precision.

```

Acknowledgements

This work was supported by the NIH Intramural Research Programs of CIT (to G.A.B. and C.D.S.), NIDDK (through G. Marius Clore), NCI (through R. Andrew Byrd), and NHLBI (through Nico Tjandra).

References

- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:66–74
- Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* 48:47–62
- Brünger AT (1993) XPLOR Manual Version 3.1. Yale University Press, New Haven. <http://xplor.csb.yale.edu/xplor-info/>
- Huang J-R, Grzesiek S (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. *J Am Chem Soc* 132:694–705
- Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against ^1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126:5879–5896
- Schwieters CD, Clore GM (2014) Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc* 80:1–11
- Gong Z, Schwieters CD, Tang C (2015) Conjoined use of EM and NMR in RNA structure refinement. *Plos One* 10:e0120445
- Clore GM, Schwieters CD (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation. *J Am Chem Soc* 126:2923–2938
- Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J Am Chem Soc* 121:2337–2338

10. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56:227–241
11. Tian Y, Schwieters CD, Opella SJ, Marassi FM (2014) A practical implicit solvent potential for NMR structure calculation. *J Magn Res* 243:54–64
12. Nilges M, Clore GM, Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229:317–324
13. Grishaev A, Bax A (2004) An empirical backbone-backbone potential in proteins and its application to NMR structure refinement and validation. *J Am Chem Soc* 126:7281–7292
14. Bermejo GA, Clore GM, Schwieters CD (2012) Smooth statistical torsion angle potential derived from a large conformational database via adaptive kernel density estimation improves the quality of NMR protein structures. *Protein Sci* 21:1824–1836
15. Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* 133:216–221
16. Bermejo GA, Clore GM, Schwieters CD (2016) Improving NMR structures of RNA. *Structure* 24:806–815
17. Schwieters CD, Clore GM (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 152:288–302
18. Powell MJD (1977) Restart procedures for the conjugate gradient method. *Math Program* 12:241–254
19. Warren JJ, Moore PB (2001) A maximum likelihood method for determining D_a^{PQ} and R for sets of dipolar coupling data. *J Magn Reson* 149:271–275
20. Losonczi JA, Andrec M, Fischer MW, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 138:334–342

Practical Nonuniform Sampling and Non-Fourier Spectral Reconstruction for Multidimensional NMR

Mark W. Maciejewski, Adam D. Schuyler, and Jeffrey C. Hoch

Abstract

A general approach to accelerating multidimensional NMR experiments via nonuniform sampling and maximum entropy spectral reconstruction was first demonstrated by Laue and colleagues in 1987. Following decades of continual improvements involving dozens of software packages for non-Fourier spectral analysis and many different schemes for nonuniform sampling, we still lack a clear consensus on best practices for sampling or spectral reconstruction, and programs for processing nonuniformly sampled data are not particularly user-friendly. Nevertheless, it is possible to discern conservative and general guidelines for nonuniform sampling and spectral reconstruction. Here, we describe a robust semi-automated workflow that employs these guidelines for simplifying the selection of a sampling schedule and the processing of the resulting nonuniformly sampled multidimensional NMR data. Our approach is based on **NMRbox**, a shared platform for NMR software that facilitates workflow development and execution, and enables rapid comparison of alternate approaches.

Key words Nonuniform sampling, NMR, Multidimensional, Maximum entropy reconstruction (MaxEnt), Spectral reconstruction, Nus-tool

1 Introduction

The difficulty of obtaining high-resolution spectral estimates from short data vectors using the discrete Fourier Transform (DFT) was well understood from the inception of pulsed FT-NMR [1], but did not present a serious obstacle until the advent of multi-dimensional NMR experiments. These experiments employ a paradigm introduced by Jeener [2], in which one or more indirect time dimensions are sampled parametrically, by repeating the experiment while incrementing the time delays corresponding to the indirect time dimensions. Consequently, the total experiment time is directly proportional to the number of sampled evolution times along the indirect dimensions; long data vectors needed for high resolution require lengthy experiments. With conventional sampling at uniform intervals required by the DFT, sampling sufficient for

high resolution in the indirect dimensions can become prohibitive for three and higher dimensional experiments. With sampling constrained by the Nyquist sampling theorem, the sampling problem for multidimensional experiments becomes more acute with increasing magnetic field.

A general solution to this sampling problem was introduced by Laue and colleagues [3, 4]: instead of sampling uniformly in the indirect dimensions, a subset of evolution times from the uniform (Nyquist) grid is sampled, and a non-Fourier method of spectrum analysis that is capable of computing high-resolution spectra despite the “missing” data. The method employed by Laue and colleagues was maximum entropy (MaxEnt) reconstruction [5], a powerful and robust regularization method that makes no assumption about the nature of the signals. A closely related regularization method that employs the l_1 -norm as the regularization functional instead of the entropy is used in an approach known as Compressed Sensing (CS) [6]. Both the methods utilize data sampled nonuniformly from the uniform Nyquist grid. They both iteratively determine the spectrum by regularizing a trial spectrum (minimizing the l_1 -norm or maximizing the entropy), and inverting the spectrum via inverse DFT to generate a “mock” data set that is constrained (either approximately or exactly, in different approaches) to agree with the measured data. There are many other methods capable of computing spectra from nonuniformly sampled data [7], including a large class of methods that model the time domain data as consisting of a sum of exponentially decaying sinusoids. The algorithms employed by different methods have been described elsewhere, but a thorough critical comparison of the results from different approaches has yet to be reported.

The results, regardless of the method used to compute the spectrum, depend on the number and distribution of indirect evolution times that are sampled [8–10]. The field of compressed sensing (CS) provides theorems [11] specifying the minimal number of samples needed to faithfully reconstruct the spectrum via l_1 -norm minimization that depends on the sparsity of the spectrum to be recovered, which is the fraction of elements in the spectrum that are nonzero. While these theorems indicate the presence of minimal sampling thresholds for accurate recovery of the spectrum from nonuniformly sampled data, their quantitative meaning for NMR spectroscopy remains unclear, because NMR resonances have finite linewidths that span multiple elements of the spectrum; in fact, the Lorentzian lineshape is not band-limited. Supporting evidence that the CS theorems do not directly apply to Lorentzian lines is that steps taken to make the spectrum more sparse improve the performance of l_1 -norm minimization for spectral recovery [12], mitigating some artifacts [13]. Since we cannot rely on CS theorems to estimate the amount of sampling needed for accurate recovery of

NMR spectra, we rely instead upon empirical experience to provide conservative guidelines.

Two main principles governing the design of robust and efficient sampling schemes are that the sampling should be biased toward the part of the signal with the highest data content and that it should be as incoherent as possible. The latter helps to ensure that sampling artifacts are as small as possible, since no non-Fourier method is capable of perfect suppression of sampling artifacts in the presence of noise. The former is needed to ensure high sensitivity. Quantitative measures of both properties exist that can be computed for sample schedules a priori. We define the sampling function as a real-valued vector that is isomorphic with the sampling grid, having the value one at elements corresponding to indirect evolution times that are sampled, and the value zero for evolution times not sampled. For an exponentially decaying signal with two indirect dimensions, the relative sensitivity of a scheme with sampling function \mathbf{K} spanning a two-dimensional grid with size n_1 by n_2 is given by

$$r(\mathbf{K}) = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K_{ij} p_{ij}}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} p_{ij}} \tag{1}$$

where the elements of \mathbf{p} are given by

$$p_{ij} = \exp\left\{-\left(\frac{R_2(1)}{SW_1}\right) - \left(\frac{R_2(2)}{SW_2}\right)\right\} \tag{2}$$

and $R_2(1)$ and $R_2(2)$ are the signal envelope decay rates, and SW_1 and SW_2 are the spectral widths in the two dimensions. Eqs. (1) and (2) generalize to other signal envelopes (i.e., non-exponential decay) and arbitrary dimensionality.

A measure of the incoherence of a sampling schedule is the peak-to-side lobe ratio (PSR) for the point-spread function (PSF) corresponding to the sampling scheme [14, 15]. The PSF is the DFT of the sampling function. The PSF has a strong central (zero frequency) component, with nonzero frequency values corresponding to sampling artifacts (for uniform sampling the PSF has a single nonzero value, the zero frequency component). The PSR is the ratio of the value of the zero-frequency component to the value of the largest nonzero-frequency component. Coherent sampling schemes have small PSR, and large sampling artifacts; incoherent sampling schemes have large PSR and small sampling artifacts.

**1.1 General
Guidelines for
Designing Robust
Sampling Schedules**

Here, we consider parameters of the experiment and sample that influence the design of robust sampling schedules. A more detailed discussion was recently provided by Rovnyak and colleagues [16].

1. *Dimensionality.* Nonuniform sampling of a single indirect dimension for two-dimensional experiments presents special challenges, because sampling artifacts are coherent across the orthogonal, uniformly sampled direct dimension [17, 18]. Nonetheless, conservative application of NUS, with sampling coverage of 0.3 or greater of the Nyquist grid, works well. If the dynamic range of the experiment is high, higher sampling coverage is appropriate. NUS from an oversampled grid has been shown to help suppress sampling artifacts by shifting them out of the spectral window. For three- and higher dimensional experiments with two or more indirect dimensions, sampling coverage of $(0.3)^k$, where k is the number of indirect dimensions, works well even for experiments with high dynamic range. When dynamic range is low and sensitivity is high, more aggressive undersampling with coverage of $(0.1)^k$ has been reported to be successful.
2. *Signal envelope.* There are three main classes of signal envelopes encountered in NMR, exponential decay, sine-modulated exponential decay, and stationary (constant-time experiments). Knowledge of the envelope type and decay rates is especially important for designing sampling schedules that attain high sensitivity. For the majority of experiments that yield exponentially decaying sinusoids, an exponentially biased sampling distribution is appropriate. For J-modulated experiments, a cosine-modulated exponential decay is appropriate, with the cosine frequency matched to the signal anti-node. For constant-time experiments, an unweighted random (or other suitable sampling distributions; *see* below) is appropriate. For exponentially biased sampling, it has been shown that an “over-matched” bias (i.e., a bias at a decay rate faster than the rate of signal envelop decay) is beneficial [19] and that sensitivity per unit measuring time is improved by a factor of two when an over-matched exponential sample weighting of $2 \times T_2$ is used [20].
3. *Maximum increment.* For Fourier methods of spectral analysis, resolution is largely determined by the maximum evolution time t_{\max} . Non-Fourier methods are in principle capable of super-resolution, that is, able to resolve frequencies closer than $1/t_{\max}$. Nevertheless, the attainable resolution still depends on the maximum evolution time sampled, and we find a maximum evolution time of $3 \times T_2$ (close to the $\pi \times T_2$ maximum evolution time needed for the DFT to achieve a digital resolution sufficient to resolve peaks separated

by their natural linewidths [1, 21]) helps ensure that resolution is determined by the magnet and sample characteristics, and is not limited by sampling. Of course, the advantage when using NUS is that not all samples corresponding to the maximum evolution times (the “far edges” of the sampling grid) need to be sampled.

4. *Sampling distribution.* The Compressed Sensing theorems are based on fully incoherent, random sampling. However, random sampling leads to poor sensitivity when used to detect exponentially decaying sinusoids. When the sampling distribution is biased to sample more of the signal power, it has been shown [16, 22] that biased (e.g., exponential) random sampling leads to large gaps in the sampling coverage that can cause unwanted sampling artifacts. Approaches that utilize nonrandom sampling distributions to distribute gaps more evenly, eliminating large gaps, have been proposed, including Poisson Gap [23], quantile (D. Rovnyak, personal communication; manuscript in preparation; software available at http://www.facstaff.bucknell.edu/drovnyak/dsrlab_downloads.html), and quasi-random [24] distributions. There is not yet a consensus on the best approach.

1.2 A Tool for Generating and Analyzing Sample Schedules

In addition to the variety of non-Fourier reconstruction techniques available to the spectroscopist, there are numerous schemes for generating sample schedules. These schemes largely follow the same set of heuristic guidelines noted above, but their differences in implementation and variety of configuration parameters are daunting. In addition, a systematic quantitative analysis of many popular sampling schemes has not been performed, largely because there has been no common platform for such an analysis and there is no consensus on the appropriate quantitative metrics. In order to address both these needs, we introduce `nus-tool`, a sample schedule utility distributed through NMRbox. The goal of `nus-tool` is to serve as a flexible interface that is able to incorporate various existing sampling schemes and analysis metrics as modular additions. The `nus-tool` will facilitate the utilization of NUS and provide a platform for systematic analysis of NUS schedules.

The `nus-tool` utility is a GUI interface that guides a user through sample schedule construction and provides facilities for writing the schedule to a file, reading a schedule from a file, performing basic a priori analyses, and plotting. The components of `nus-tool` are described below, and a screenshot of the utility is shown in Fig. 1.

1. **Generate.** Sample schedules may be generated using random sampling, exponentially biased sampling, and a set of schemes based on quantiles (D. Rovnyak, personal communication) that prevent large gaps. When sample schedules are generated,

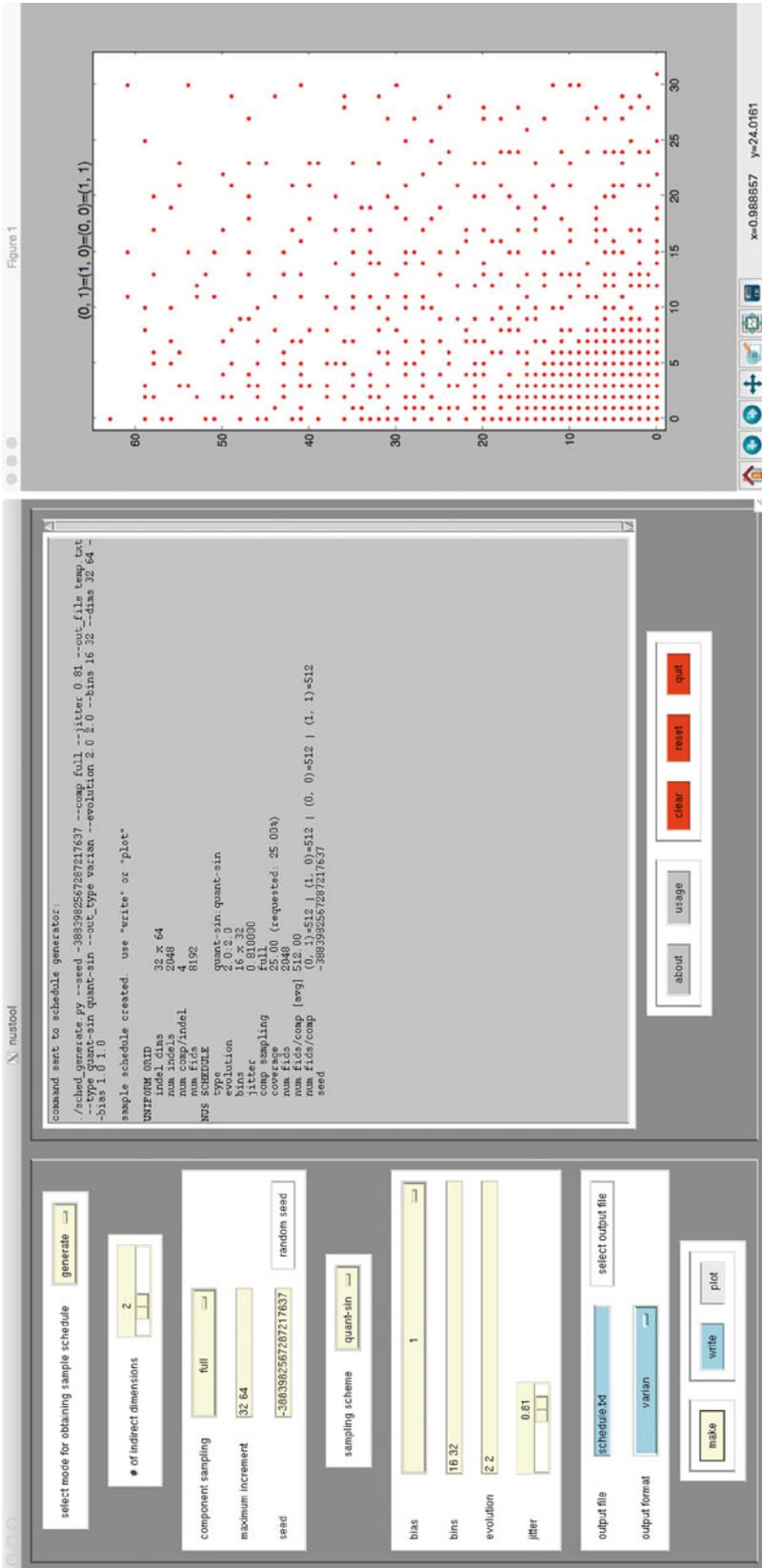


Fig. 1 The `nus-tool` GUI (left) and sample schedule (right). The left panel of the `nus-tool` GUI collects parameter inputs from the user and the right panel displays a status box for text output. The example schedule is full-component (i.e., the same sampling schedule is used across all four hypercomplex components for this 2D case). The `nus-tool` fully supports partial-component sampling (i.e., independent sampling of each hypercomplex component) and will be useful in further characterizing the benefits which it affords

the command executed behind the GUI is reported in the status window, so that it may be captured and run from a script.

2. **Plot.** Sample schedules may be shown in a scatter plot. Plots are currently limited to 2D, but tools for selecting 2D planes through higher dimensional schedules will be implemented. In addition, projections that sum over dimensions will be included; this will allow the sampling distribution along each axis to be visually inspected.
3. **Read/Write.** Sample schedule file formats supported are Varian, Bruker, and RNMRTK. There is additional ongoing work by the NMRbox team to develop a Nonuniform Exchange (NEX) file format. This universal format will aid with workflow analysis in [25]. The draft specification for NEX is currently included.
4. **Analyze.** Metrics for sensitivity (intrinsic sensitivity), resolution (mean/max evolution times), and artifact level (PSR) are reported.

Once generated, sampling schedules produced by `nus-tool` are used by the spectrometer to perform the experiment. Details of implementing NUS experiments vary among instruments from different manufacturers, and among different versions of operating software from the instrument vendors.

1.3 General Guidelines for Spectral Reconstruction

Maximum entropy (MaxEnt) reconstruction belongs to the class of regularization methods (which includes compressed sensing, CS) that make no assumptions about the nature of the signals, and thus are among the most robust non-Fourier methods. The two parameters governing MaxEnt reconstruction are *aim*, the level of agreement between the mock-data obtained by inverse DFT of the final spectrum and the empirically measured data, and *def*, a parameter related to the sensitivity of the spectrometer. There are two main regimes for values of *aim*; in the Bayesian regime *aim* is comparable to the noise in the data. In the “MINT” or “FM” regime [20, 26], the spectral reconstruction is nearly linear. The latter regime gives smoother reconstructions and few false positives. Although the MINT regime leads to noisier reconstructions, MINT is useful when quantification of peak intensities or preservation of spectral lineshapes is important.

2 Methods-General Spectrum Analysis Workflow

Here, we describe a general workflow for processing data from a 3D NUS NMR experiment with MaxEnt, as implemented in the Rowland NMR Toolkit (RNMRTK). Figure 2 illustrates various steps in the workflow, and compares MaxEnt reconstruction of the NUS data with DFT of the data with zeroes used in place of FIDs not sampled (nuDFT).

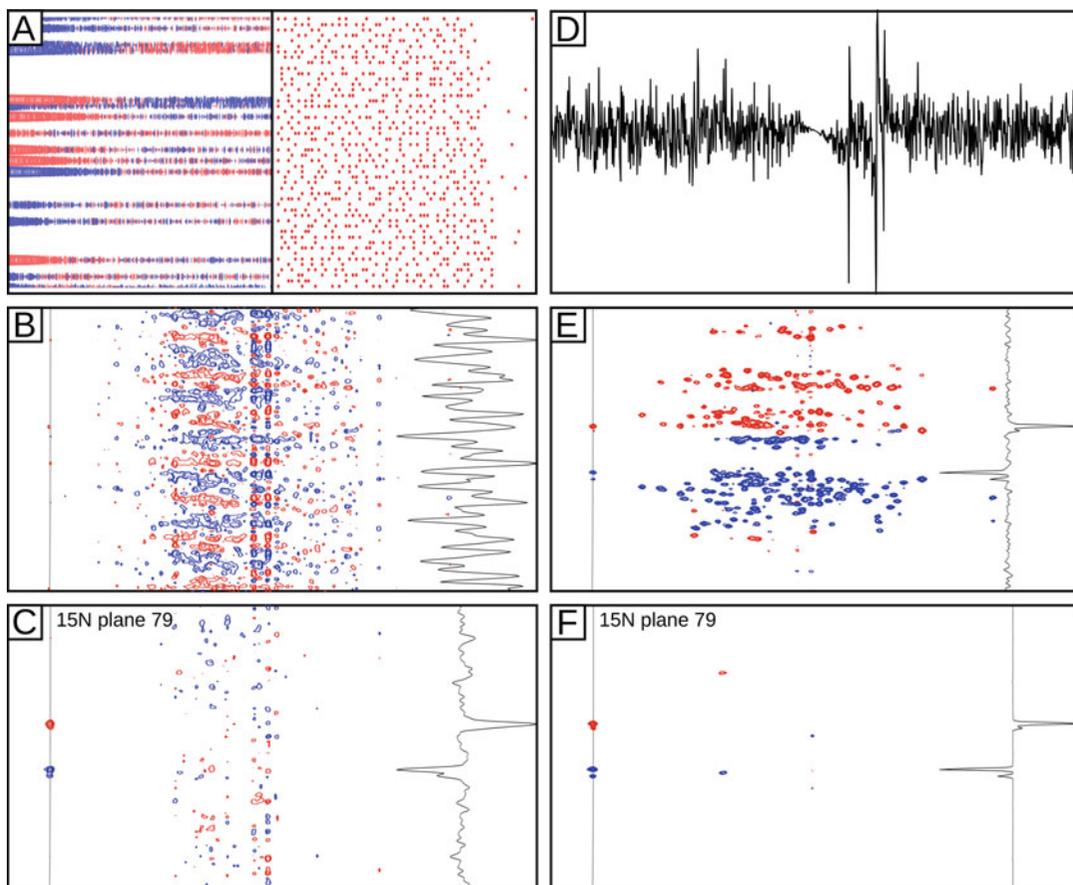


Fig. 2 Steps in the workflow for processing of 3D NUS data in NMRbox. *Panel A*—(left) The first t_1/t_3 plane of the fully expanded time domain data with zeros filled in for FIDs not in the sample schedule. (right) The sample schedule indicating t_2 (CACB, x axis) and t_1 (N, y axis) intervals is collected. The first column in the sample schedule, corresponding to the first t_2 plane, corresponds to the collected FIDs on the left. *Panel B*—A 2D HN/CACB summed projection of the 3D nuDFT spectrum with a 1D slice corresponding to the position of the solid vertical line. Here, the sampling noise summed across the projection overwhelms the signals. The same 1D slice is also shown in panels C, E, and F. *Panel C*—A single HN/CACB plane (plane 79) of the 3D nuDFT spectrum. While significant sampling artifacts are present, the spectrum is suitable for determining phase and other processing parameters. *Panel D*—The spectrum of a single FID with the longest combined evolution delay in the sample schedule, prior to extraction along the amide proton region, which is used by the program *noisecal* to estimate values for the MaxEnt parameters *def* and *aim*. *Panel E*—The 2D HN/CACB summed projection of the automated MaxEnt reconstruction. *Panel F*—A single HN/CACB plane of the 3D MaxEnt reconstruction. The significant reduction in sampling noise in the MaxEnt reconstruction as compared to the nuDFT is readily apparent in Panels E and F

2.1 Manual Approach

1. **Expand.** Expand the Agilent or Bruker NUS time domain data into a full multi-dimensional data set with zeros replacing any FIDs not in the sample schedule. This is most easily implemented with *nusExpand.tcl* which is part of the NMRPipe conversion utility and whose syntax is generally determined automatically by the “*bruker*” or “*varian*” commands in the

NMRPipe [27] distribution. Several significant advantages occur by expanding the data prior to data conversion including the ability to process the data with a nonuniform DFT (nuDFT). While the nuDFT will give the poorest result with no suppression of sampling artifacts, it is a quick method of determining important processing parameters that are needed such as data reversals, sign alterations, regions to extract, phase values, and to validate the data conversion. In addition, an intermediate file from the nuDFT will be utilized later in the workflow for determining reasonable values for the MaxEnt parameters *def* and *aim*.

2. **Convert to NMRPipe.** Convert time domain data from the vendor specific to NMRPipe file format. During the conversion meta-data regarding referencing can be modified, Bruker digital oversampled FIDs are corrected, and any data shuffling due to sensitivity enhancement in indirect dimensions is performed. The data conversion script is created automatically by the same “*bruker*” or “*varian*” commands as with **step 1**.
3. **Convert to RNMRTK.** Data is converted from NMRPipe to RNMRTK format with the program *spectrum-translator*. Note that the NMRPipe formatted data must be in a single file for *spectrum-translator* to function properly.
4. **Process the spectrum with a nonuniform DFT (nuDFT).** Data are processed along all dimensions with DFT with traditional processing steps such as solvent suppression, data reversals, apodization, zero filling, and phasing. However, a key difference is that an intermediate file is saved after the nuDFT along the acquisition dimension prior to extracting a region of interest and deleting the imaginary data. The intermediate saved data will be used in the next step to automatically determine reasonable values for the MaxEnt parameters; *def*.
5. **Determine reasonable values for *def*, and *aim*.** The 1D spectrum with the longest combined t_1/t_2 evolution time from the intermediate saved data is analyzed by the program *noisecalc* which determines reasonable values for *def* and *aim* based on the RMS noise of the spectrum after purging the highest amplitude signals and skipping the central solvent region [28, 29].
6. **Perform preliminary maximum entropy reconstructions to determine λ .** When processing t_1/t_2 planes of 3D data sets with MaxEnt in constant-*aim* mode each plane stacked along the t_3 dimension is processed independently and produces its own value of λ , the Lagrange multiplier that determines the weight applied to the constraint term relative to the entropy [30]. The variation of λ values among the t_1/t_2 planes introduces a variation of intensity values, thus causing distortions in peak shapes. To resolve this, the λ values are averaged to

produce a single representative λ value, which is then used to reprocess the t1/t2 planes in constant- λ [30] mode.

7. **Process the spectrum with maximum entropy reconstruction.** A MaxEnt reconstruction along all t1/t2 planes is performed in constant- λ mode using the *def* value from earlier and the averaged λ value from the previous step.
8. **Examine the spectrum and adjust the def multiplication factor if desired.** It is often desirable to set *def* slightly less than the value determined by *noisecalc*. By setting *def* lower the converged λ value will also decrease and the number of loops to converge will increase. This can lead to a spectrum with a better cosmetic appearance, but if pushed too far will lead to noise spikes. Good results are generally obtained by multiplying *def* by a value from 1.0 to 0.1.

2.2 Automated Approach

The steps outlined in the above workflow can be tedious and rely on having strong technical skills with the Rowland NMR Toolkit program suite. To alleviate this burden we have created a program called *auto-maxent* which handles all of the steps outlined in the workflow with the exception of the generation of the NMRPipe file conversion script. To implement the workflow these steps are performed:

1. The NMRPipe “*bruker*” or “*varian*” command is run from the directory with the raw time domain data.
 - (a) The “Read Parameters” button is selected.
 - (b) For Bruker data the checkbox “During Conversion (Normal FID)” is checked.
 - (c) Meta data for referencing and axis labels can be adjusted if desired.
 - (d) The “Save Script” button is selected to save the script as “[fid.com](#)” and the dialog box closed.
2. The data is converted by running “*auto-maxent convert.*” This will run the NMRPipe conversion script, which will expand the data and then convert to NMRPipe format. Data will then be converted to RNMRTK format with *spectrum-translator*.
3. A configuration file named *process.cfg* is created by running “*auto-maxent setup.*”
4. The data are processed with nuDFT by running “*auto-maxent dft.*” The results are viewed in *NMRDraw* or *contour* and the values in the *process.cfg* configuration file are updated iteratively until a correct looking spectrum is attained, although with significant sampling noise present.
5. Once the nuDFT produces a spectrum with the correct processing parameters and a region of interest has been selected, the

data are processed with MaxEnt reconstruction with “*auto-maxent maxent*.” The program will perform **steps 5–7** in the workflow above; FIDs are analyzed for reasonable values of *def* and *aim*, preliminary MaxEnt calculations are performed on t1/t2 planes to determine λ , and then the whole spectrum is processed with MaxEnt in constant- λ mode.

6. The last step is to check the maximum number of loops for convergence and adjust the *def multiplier* if desired. Typically, the *def multiplier* is set between 1.0 and 0.1 and careful examination of weaker signals should be compared with the different *def multipliers*. Note that when the *def multiplier* is set low the spectrum may look cosmetically better, but 1D strips should also be examined to be sure the noise has a normal distribution and does not become spiky which would indicate *def* was set too low.

All software described here, as well as sample data sets, sampling schedules, and a detailed step-by-step tutorial are available on the NMRbox platform. Access to NMRbox is free for academic and non-profit users by visiting <https://nmrbox.org>. NMRbox is provided by the National Center for Biomolecular NMR Data Processing and Analysis, an NIH/NIGMS Biomedical Technology Research Resource.

Acknowledgments

We thank Alan S. Stern, Gerard Weatherby, Frank Delaglio, David Rovnyak, and Levi Craft for useful discussions and technical support. Support for NMRbox from the US National Institutes for Health (via grant P41GM111135) is gratefully acknowledged. Support from NIH (via grant R21GM104517) for research on MaxEnt reconstruction is also gratefully acknowledged.

References

1. Hoch JC, Stern AS (1996) NMR data processing. Wiley-Liss, New York
2. Jeener J (1971) Oral presentation. Ampere International Summer School, Basko Polje, Yugoslavia
3. Barna JCJ, Laue ED (1987) Conventional and exponential sampling for 2D NMR experiments with application to a 2D NMR spectrum of a protein. *J Magn Reson* 75:384–389
4. Barna JCJ, Laue ED, Mayger MR, Skilling J, Worrall SJP (1987) Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *J Magn Reson* 73:69–77
5. Skilling J, Bryan RK (1984) Maximum entropy image reconstruction: general algorithm. *Mon Not R Astron Soc* 211:111–124
6. Donoho DL (2006) Compressed sensing. *IEEE Trans Inform Theory* 52:1289–1306
7. Mobli M, Hoch JC (2014) Nonuniform sampling and non-Fourier signal processing methods in multidimensional NMR. *Prog Nucl Magn Reson Spectrosc* 83C:21–41. doi:10.1016/j.pnmrs.2014.09.002
8. Mobli M, Maciejewski MW, Schuyler AD, Stern AS, Hoch JC (2012) Sparse sampling methods in multidimensional NMR. *Phys Chem Chem Phys* 14:10835–10843. doi:10.1039/c2cp40174f

9. Hyberts SG, Arthanari H, Wagner G (2012) Applications of non-uniform sampling and processing. *Top Curr Chem* 316:125–148. doi:10.1007/128_2011_187
10. Kazimierczuk K, Orekhov V (2015) Non-uniform sampling: post-Fourier era of NMR data collection and processing. *Magn Reson Chem* 53:921–926. doi:10.1002/mrc.4284
11. Donoho DL, Tanner J (2009) Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Phil Trans R Soc* 367:4273–4293
12. Stern AS, Hoch JC (2015) A new approach to compressed sensing for NMR. *Magn Reson Chem* 53:908–912. doi:10.1002/mrc.4287
13. Stern AS, Donoho DL, Hoch JC (2007) NMR data processing using iterative thresholding and minimum l_1 -norm reconstruction. *J Magn Reson* 188:295–300
14. Lustig M, Donoho D, Pauly JM (2007) Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med* 58:1182–1195
15. Schuyler AD, Maciejewski MW, Stern AS, Hoch JC (2013) Formalism for hypercomplex multidimensional NMR employing partial-component subsampling. *J Magn Reson* 227:20–24. doi:10.1016/j.jmr.2012.11.019
16. Palmer MR, Wenrich BR, Stahlfeld P, Rovnyak D (2014) Performance tuning non-uniform sampling for sensitivity enhancement of signal-limited biological NMR. *J Biomol NMR* 58:303–314. doi:10.1007/s10858-014-9823-5
17. Hoch JC, Maciejewski MW, Filipovic B (2008) Randomization improves sparse sampling in multidimensional NMR. *J Magn Reson* 193:317–320
18. Monajemi H (2016) Phase transitions in deterministic compressed sensing, with application to magnetic resonance spectroscopy. Stanford University, Stanford, CA
19. Schuyler AD, Maciejewski MW, Arthanari H, Hoch JC (2011) Knowledge-based nonuniform sampling in multidimensional NMR. *J Biomol NMR* 50:247–262
20. Paramasivam S, Suiter CL, Hou GJ, Sun SJ, Palmer M, Hoch JC, Rovnyak D, Polenova T (2012) Enhanced sensitivity by nonuniform sampling enables multidimensional MAS NMR spectroscopy of protein assemblies. *J Phys Chem B* 116:7416–7427. doi:10.1021/Jp3032786
21. Rovnyak D, Sarccone M, Jiang Z (2011) Sensitivity enhancement for maximally resolved two-dimensional NMR by nonuniform sampling. *Magn Reson Chem* 49:483–491
22. Hyberts SG, Takeuchi K, Wagner G (2010) Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* 132:2145–2147. doi:10.1021/ja908004w
23. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling. *J Biomol NMR* 52:315–327. doi:10.1007/s10858-012-9611-z
24. Worley B, Powers R (2015) Deterministic multidimensional nonuniform gap sampling. *J Magn Reson* 261:19–26. doi:10.1016/j.jmr.2015.09.016
25. Fenwick M, Weatherby G, Vyas J, et al. CONNJUR Workflow Builder: A software integration environment for spectral reconstruction. *Journal of Biomolecular NMR*. 2015;62(3):313–326. doi:10.1007/s10858-015-9946-3
26. Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luithardt H, Fejzo J, Chorev M, Aktas H, Colson K, Falchuk KH, Halperin JA, Wagner G (2007) Ultrahigh-resolution $(1)H-(13)C$ HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J Am Chem Soc* 129:5108–5116
27. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
28. Mobli M, Maciejewski MW, Gryk MP, Hoch JC (2007) An automated tool for maximum entropy reconstruction of biomolecular NMR spectra. *Nat Methods* 4:3–4
29. Mobli M, Maciejewski MW, Gryk MP, Hoch JC (2007) Automatic maximum entropy spectral reconstruction in NMR. *J Biomol NMR* 39:133–139
30. Schmieder P, Stern AS, Wagner G, Hoch JC (1997) Quantification of maximum entropy spectrum reconstructions. *J Magn Reson* 125:332–339

Covariance NMR Processing and Analysis for Protein Assignment

Bradley J. Harden and Dominique P. Frueh

Abstract

During NMR resonance assignment it is often necessary to relate nuclei to one another indirectly, through their common correlations to other nuclei. Covariance NMR has emerged as a powerful technique to correlate such nuclei without relying on error-prone peak peaking. However, false-positive artifacts in covariance spectra have impeded a general application to proteins. We recently introduced pre- and postprocessing steps to reduce the prevalence of artifacts in covariance spectra, allowing for the calculation of a variety of 4D covariance maps obtained from diverse combinations of pairs of 3D spectra, and we have employed them to assign backbone and sidechain resonances in two large and challenging proteins. In this chapter, we present a detailed protocol describing how to (1) properly prepare existing 3D spectra for covariance, (2) understand and apply our processing script, and (3) navigate and interpret the resulting 4D spectra. We also provide solutions to a number of errors that may occur when using our script, and we offer practical advice when assigning difficult signals. We believe such 4D spectra, and covariance NMR in general, can play an integral role in the assignment of NMR signals.

Key words NMR, Covariance, Resonance assignment, Peak lists, Spectral derivative, 4D spectra

1 Introduction

1.1 Assignment Without Peak Lists

A central task in NMR assignment involves relating nuclei to each other either through direct correlations or through their mutual correlation to a common nucleus or set of nuclei. For example, protein amide H^N and N resonances are directly correlated to each other in 2D HN-HSQC spectra, forming (H^N, N) correlations. Similarly, methyl H^M and C^M resonances form (H^M, C^M) correlations in 2D HC-HSQC spectra. On the other hand, for large proteins, amide resonances are related to methyl resonances through their mutual correlation to C^α and C^β nuclei in 3D HNCA, HN(CA)CB, and HMCMBCA spectra [1, 2]. Conventional assignment techniques accomplish this by first abstracting the raw data into a peak list made of frequency coordinates and then comparing frequencies to identify shared resonances. In this

instance, HNCA and HN(CA)CB spectra give rise to (H^N , N, C^α) and (H^N , N, C^β) correlations respectively, while HMCMBCA spectra give rise to both (H^M , C^M , C^α) and (H^M , C^M , C^β) correlations for valine residues. Software algorithms can identify instances of common C^α and C^β frequencies among these coordinates and translate them into proposed assignments. While effective, the conventional approach relies crucially on the premise that the frequency coordinates in each abstracted peak list are a faithful representation of the spectrum's true, underlying correlations. However, if this assumption is not fulfilled, the approach will fail outright. In such a case, only revising the peak list can rescue the assignment, yet this task is often fraught with ambiguity and error. Over the past 12 years, covariance NMR has emerged as a complementary tool to conventional assignment techniques, because it does not rely on any form of abstraction. It accomplishes the same task as traditional approaches, namely relating unconnected nuclei through their mutual correlation to a common nucleus, but it does so by directly manipulating the raw data itself.

Covariance NMR was developed in 2004 by Brüschweiler and coworkers [3–7] as an alternative to conventional Fourier transform processing. Initially, spectra were covaried with themselves as a means to enhance resolution in indirectly detected dimensions. Soon after, Blinov et al. introduced unsymmetrical covariance NMR [8, 9], whereby novel correlation maps could be created by covarying two *different* spectra. Later, related techniques broadened to include hyperdimensional NMR [10, 11], assignment without peak lists [12] and cross-spectra [13]. Finally, Snyder et al. codified the theory of covariance NMR [14, 15] and expanded its applications to higher dimensional spectra [16, 17].

Although covariance NMR is a promising alternative to traditional methods, the technique has not been widely adopted by the NMR community. This is perhaps due to the high prevalence of false-positive signals in covariance spectra. In an effort to address this limitation, our lab recently introduced pre- and postprocessing steps to help reduce or eliminate such artifacts [18, 19]. We discuss our methods in Subheading 1.2. In Subheading 1.3, we discuss specific applications using 4D covariance spectra made possible with our procedures that we have successfully employed in both the backbone and sidechain assignment of two large and challenging proteins [18, 20].

1.2 Eliminating False-Positive Correlations

To better understand the source of false-positive correlations, we can recast covariance NMR as a series of inner products between 1D slices from two different spectra. Figure 1 illustrates that a spectrum of (I, K) correlations can be covaried with a spectrum of (L, K) correlations to form a covariance spectrum of (I, L) correlations. The covariance operation, in this case, is equivalent to the matrix product of the (I, K) and (L, K) spectra. If the two input

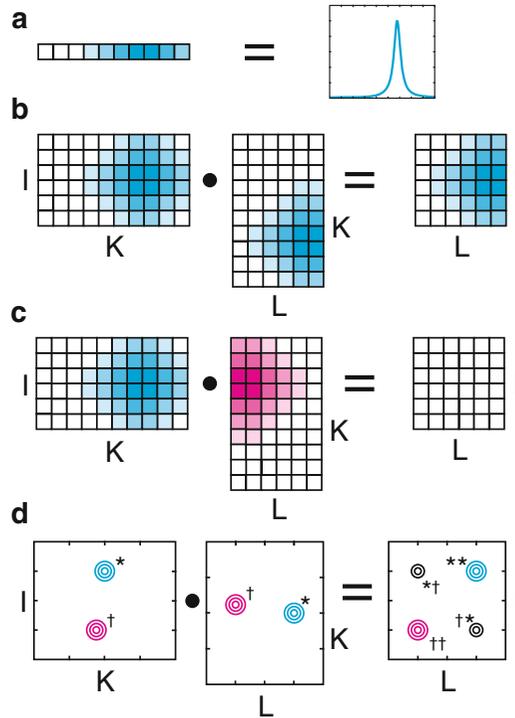


Fig. 1 Unsymmetrical covariance is equivalent to matrix multiplication. (a) The series of colored squares to the *left* represent the peak plotted on the *right* as a row vector. The intensity of the color in each square corresponds to the amplitude of the peak. (b) 2D spectra can be represented as matrices. An $I \times K$ matrix multiplied by a $K \times L$ matrix results in an $I \times L$ matrix, subsuming the K dimension. If two spectra share a signal along the K dimension, there will be a peak in their resulting matrix product. (c) If two spectra do not share a peak in the K dimension, their matrix product will be devoid of any peaks. (d) If two sets of peaks are nearly degenerate in the K dimension, the covariance spectrum will feature false-positive artifacts. Here, each pair of peaks in the input spectra, marked with \dagger or $*$, produces a peak in the covariance spectrum. The peaks marked $\dagger\dagger$ and $**$ are the desired correlations, while the peaks marked $\dagger*$ and $*\dagger$ are false-positive artifacts

spectra share a common signal along the subsumed (K) dimension, such as in panel (b), then the resulting covariance spectrum will feature a correlation at the corresponding row and column of the signals in the first and second spectra respectively. However, if the two spectra do not share a common signal in the subsumed dimension, as in panel (c), then the covariance spectrum will be empty. False-positive artifacts originate from covariance between 1D slices that share a signal purely by coincidence. Such an example is shown in panel (d). Here, each original spectrum contains two peaks, marked with $*$ and \dagger , that are nearly degenerate in the K dimension. As expected, the covariance spectrum features the desired

correlations at the ** and †† positions. However, as a result of near degeneracy, it also features undesirable peaks at the *† and †* positions. Such artifacts are an inevitable result of traditional covariance methods and are especially prevalent when dealing with crowded spectra, for example those of large and/or intrinsically disordered proteins. Panel (d) also underscores the importance of carefully processing the subsampled dimension. Any signals shared between the input spectra, even those originating from spurious signals or artifacts, will be carried over to the covariance spectrum (*see Note 1*).

Our lab has introduced a preprocessing step that helps to reduce the number and severity of such false-positive artifacts. Figure 2 illustrates that the inner product is separable into two operations, element-wise multiplication and summation. With standard covariance NMR processing (panels (a)–(f)), if two 1D slices feature even partially overlapping signals, then their element-wise product will be strictly positive, and the subsequent summation will inevitably produce an incorrect and undesirable peak in the covariance spectrum. However, if we first take the derivative of each slice (panels (g) and (h)), prior to element-wise multiplication, the product will be strictly positive only in the case of exact alignment between the signals (panels (i)–(k)). If the signals are not perfectly aligned (panels (l)–(n)), the element-wise product will feature both positive and negative signals, and after summation, the negative signals will act to reduce, change the sign, or eliminate false-positive correlations. Spurious negative correlations can then be eliminated by restricting the spectrum to only its positive elements.

In addition to the improvements offered by the derivative preprocessing step, we have also implemented postprocessing enhancements to the standard covariance protocol. Because we are often able to find more than one common nucleus through which we can relate unconnected nuclei, we can combine covariance spectra obtained through different nuclei to further decrease false-positive artifacts. Figure 3 demonstrates that two nuclei I and L can be related to each other through common correlations to two different nuclei K_1 and K_2 . A spectrum of (I, K_1) correlations can be covaried with a spectrum of (L, K_1) correlations to form a covariance spectrum of (I, L) correlations, and spectra with (I', K_2) and (L', K_2) correlations create a covariance spectrum of (I', L') correlations. Here, K_2 represents a completely different nucleus than K_1 , e.g., C^β rather than C^α , whereas the apostrophes indicate a second instance of the same nucleus, e.g., I and I' are both H^N . Although both the (I, L) and (I', L') spectra may each contain false-positive signals originating from truly degenerate K_1 or K_2 frequencies, the artifacts in each spectrum will likely be different. As a result, taking the element-wise product of the two covariance maps will reinforce the shared peaks while reducing erroneous signals. Furthermore, if we restrict ourselves to only the positive correlations from each

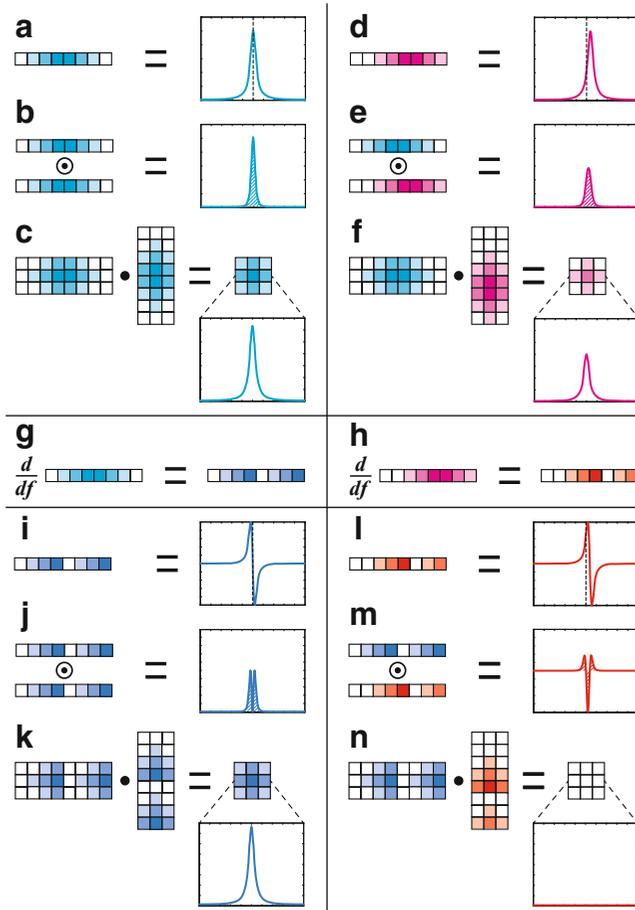


Fig. 2 Derivatives and covariance spectra. **(a)** The series of colored squares to the *left* represents the peak to the *right* as a row vector. The intensity of the color in each square corresponds to the amplitude of the peak. The peak is centered on the dotted line. **(b)** When two peaks align perfectly, the element-wise product of their row vectors is strongly positive. **(c)** The matrix product of two 2D spectra is a covariance spectrum. Each of its elements is a sum of an element-wise product, like that shown in **(b)**. In this case, a strong signal is produced in the covariance spectrum. **(d)** A different peak, slightly offset from the peak in **(a)**, is represented by a shifted row vector. **(e)** The shift between the peaks in **(a)** and **(d)** reduces their element-wise product. **(f)** The two signals from **(a)** and **(d)** still produce an undesirable peak in the covariance spectrum. **(g)** and **(h)** Taking the derivatives of the vectors in **(a)** and **(d)** yields the vectors in **(i)** and **(l)** respectively. The colored squares have been reordered for clarity, even though they no longer correctly represent the changes in intensity. **(i)** The derivative of the signal in **(a)** is still centered on the dotted line. **(j)** If two peaks align perfectly, then their derivatives will as well, and the element-wise product of the two vectors is strictly positive. **(k)** Upon summation, the element-wise product produces a strong peak in the covariance spectrum, as desired. **(l)** The derivative of the signal in **(d)** is offset from the dotted line. **(m)** The element-wise product of the vectors from **(i)** and **(l)** has both positive and negative values. **(n)** Upon summation, the positive and negative values from the element-wise product cancel each other, producing an empty covariance spectrum. The false-positive artifact has been eliminated

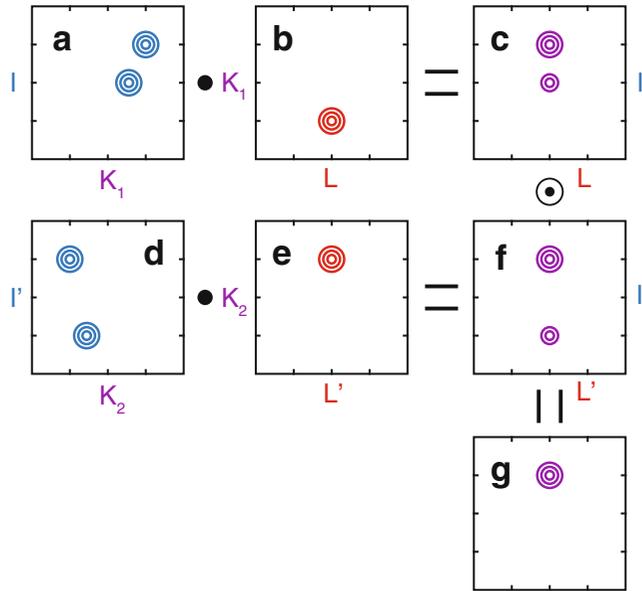


Fig. 3 The element-wise product of covariance spectra can eliminate artifacts. Spectrum (a) is covaried with spectrum (b) to produce spectrum (c). The dot symbol indicates matrix multiplication/covariance. The top-right signal in spectrum (a) gives rise to the top signal in spectrum (c). The second signal in spectrum (a) does not align perfectly with the signal in spectrum (b) in the subsumed dimension, and as a result, gives rise to the smaller, artifact signal in spectrum (c). Similarly, spectra (d) and (e) are covaried to produce spectrum (f). The small, artifact peak in spectrum (f) is again the result of covariance between imperfectly aligned signals in (d) and (e). (g) The element-wise product of (c) and (f) reinforces the correct peak while eliminating the artifacts. We represent the element-wise product operator with a *circled dot*

individual covariance map prior to element-wise multiplication (*see* the above), we eliminate any possibility of introducing erroneous signals through the multiplication of two negative peaks.

1.3 4D Covariance Spectra

Our applications so far [19, 20] have primarily focused on covarying two 3D spectra with correlations of the form (I, J, K) and (L, M, K) to create a 4D spectrum with (I, J, L, M) correlations (Fig. 4). A 4D covariance spectrum of this type amounts to the inner product of every possible combination of 1D slices along the K dimension of each spectrum. As a result, all of the previously discussed improvements remain applicable. The derivative can be used as a preprocessing step before covariance, and the element-wise product of multiple 4D spectra can be used after covariance to combine spectra and reduce or eliminate artifacts.

Typically, these 4D spectra map one set of “spin anchors” to another set of “spin anchors” through an intermediary set of nuclei. Here, we borrow the term spin anchor from the software

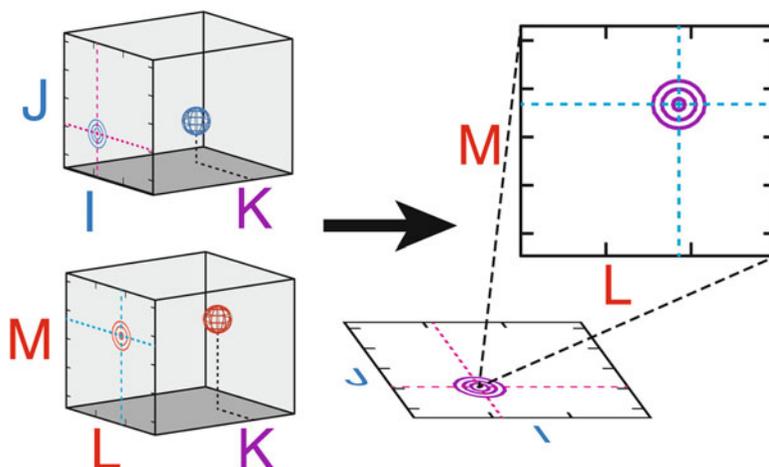


Fig. 4 4D covariance spectra. A 3D spectrum of (I, J, K) correlations can be covaried with a second 3D of (L, M, K) correlations to form a 4D covariance spectrum with (I, J, L, M) correlations. If the two 3D spectra share a signal at a common frequency in the K dimension, as shown, there will be a peak in the covariance spectrum at the corresponding frequencies in the I, J, L, and M dimensions. A 4D spectrum can be visualized as a plane of planes. As a result, each point in the I/J plane corresponds to an entire L/M plane

application CARRA [21] to refer to a correlation between a hydrogen nucleus and its directly attached, heavy atom nucleus. Much of the NMR assignment process revolves around spin anchors. They are the correlations observed in HN- and HC-HSQC spectra, and many multidimensional NMR experiments correlate a spin anchor to some other nucleus or nuclei. For example, HNCA correlates the H^N anchor with C^α and HMCMBCA correlates valine methyl $H^M C^M$ anchors with C^α and C^β . 4D spectra correlating spin anchors are particularly useful and simple to conceptualize. They are easy to navigate, and they lend themselves well to assignment without peak lists.

In the following protocol, we will outline the steps necessary to calculate, navigate, and interpret 4D covariance spectra. We will discuss how to prepare 3D spectra, including any necessary re-processing. We will explain how to run our covariance script and adjust the available options. Finally, we will illustrate how to traverse and interpret the spectrum.

2 Materials

Covariance processing with our script requires four principal components: a working installation of MATLAB or GNU Octave, the Covariance NMR Toolbox [22] by Snyder et al., the script itself and a set of 3D spectra. Links to download our script, the Covariance NMR Toolbox and GNU Octave are available on our website at <http://frueh.med.jhmi.edu/software-downloads/>.

Our script has been tested in both MATLAB 8.3+ (2014+) [23] and GNU Octave 4.0+ [24] on Linux, Mac, and Windows operating systems. However, due to the simplicity of the code, it is anticipated that our script will run successfully on most modern versions of either MATLAB or Octave. Users are directed to their institution for MATLAB licensing. Octave is available free of charge. The Covariance NMR Toolbox should be decompressed, placed in a permanent location on the disk, and added to the MATLAB or Octave path with the `addpath` command. The script itself should be copied and edited for each separate instance of its execution, much like NMRPipe processing scripts. The input 3D spectra should be in NMRPipe format, but they may be in either plane-by-plane format or in a single, monolithic file.

Our script can be used to create 4D spectra connecting any combination of HN or HC spin anchors. Figure 5 presents a number of possible combinations. Each pair of 3D spectra to be covaried should feature correlations to a common nucleus that will be used to correlate the two different spin anchors. Further, if the same spin anchors can be related to each other through different

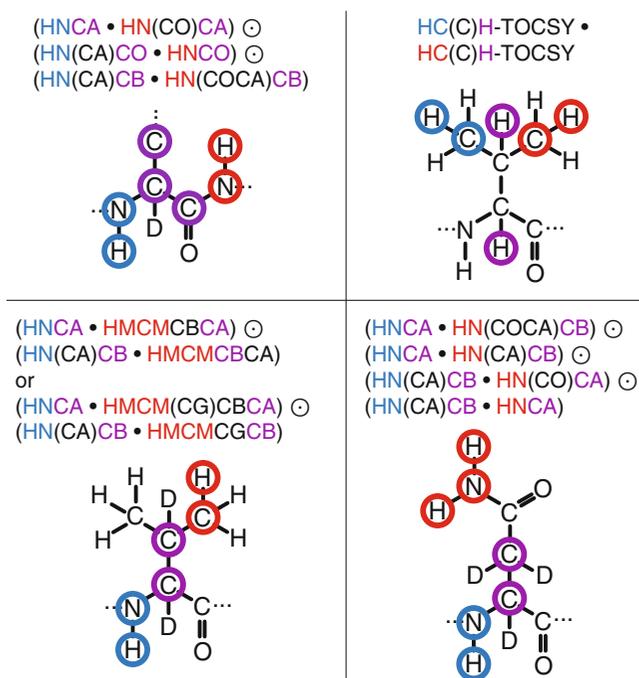


Fig. 5 Types of 4D covariance spectra that can be generated with the script described in this chapter. Each panel illustrates the amino acid nuclei involved in a particular 4D covariance spectrum. The nuclei forming the spin anchors that will appear in the 4D spectra are circled in *blue* or *red*, while the nuclei subject to covariance and connecting the two spin anchors are circled in *purple*. Above each amino acid is a diagram indicating the 3D spectra involved and their proper combination during covariance calculations. A *single dot* indicates that two spectra should be covaried, subsuming the purple dimensions. A *circled dot* indicates that two covariance spectra should be multiplied in an element-wise fashion

nuclei, the resulting set of 4D covariance spectra can be combined with element-wise multiplication to reduce artifacts.

Once the requisite software packages and spectra have been gathered, users may prepare their spectra for covariance.

3 Methods

3.1 Preparing Spectra

Before using our script to calculate a 4D covariance spectrum, users must first prepare their 3D spectra appropriately. To successfully use our script each spectrum must be transposed correctly, the digital resolution must match appropriately among all 3D spectra, and the spectra must have a self-consistent calibration.

Our script requires that the spin anchor nuclei be placed in the X and Y dimensions of the NMRPipe spectrum and that the shared nucleus be placed in the Z dimension. The Z dimension corresponds to the K dimension in Fig. 4 and is the dimension that will be subsumed during covariance. To satisfy this requirement, users must transpose their spectra appropriately in nmrPipe. This can be accomplished using any combination of the **nmrPipe** functions TP and ZTP as well as when reading (**xyz2pipe**) and writing (**pipe2xyz**) data with the options -x, -y, and -z. Table 1 describes the effect of each function or option when reading the data from the disk, transforming it in memory and writing back to the disk. Many combinations amount to the same result. For example, to transfer data from the Y dimension to the Z dimension, use a combination of **xyz2pipe -y** and **pipe2xyz -z**. The same could be accomplished with **xyz2pipe -z** and **pipe2xyz -x** or even **xyz2pipe -**

Table 1

Effect of each NMRPipe function/option on the transposition of the data. A, B, and C each represent a dimension name. The order of the three characters in each box indicates the order of the data in the dimensions X, Y, and Z of the spectrum respectively. The right-most column specifies the location of the data before and after using the function/option

Function/option	Before	After	Start -> End
xyz2pipe -x	ABC	ABC	Disk -> Memory
xyz2pipe -y	ABC	BAC	Disk -> Memory
xyz2pipe -z	ABC	CAB	Disk -> Memory
nmrPipe -fn TP	ABC	BAC	Memory -> Memory
nmrPipe -fn ZTP	ABC	CBA	Memory -> Memory
pipe2xyz -x	ABC	ABC	Memory -> Disk
pipe2xyz -y	ABC	BAC	Memory -> Disk
pipe2xyz -z	ABC	BCA	Memory -> Disk

x, **nmrPipe** -fn TP, **nmrPipe** -fn ZTP and **pipe2xyz** -x. Users can verify that the data are correctly transposed by using the command **showhdr**. The shared nucleus should be listed under the column “Z-Axis” for each spectrum.

If multiple pairs of 3D spectra will be covaried and later multiplied together, then the order of the spin anchor dimensions should be consistent across each pair of 3D spectra. For example, if H^N spin anchors in the HNCA and HN(CA)CB spectra will be connected to $H^M C^M$ spin anchors in HMCBCA spectra through both C^α and C^β , then H^N should be found in the same dimension of both the HNCA and HN(CA)CB spectra, e.g., along the X dimension. The same is true for the other three nuclei comprising the two spin anchors that will be correlated. Note that once the shared nucleus has been properly placed in the Z dimension, correctly placing the remaining nuclei should require *at most* an XY transpose with the **nmrPipe** function TP or **xyz2pipe** and **pipe2xyz** options -y.

The next step when preparing spectra is to match the digital resolution along any dimensions that undergo covariance or element-wise multiplication. This is essential to ensure that each discrete element of one spectrum is multiplied with its corresponding element in the other spectrum. As discussed in Subheading 1.2, element-wise multiplication is part of both the covariance operation itself and our postprocessing scheme. If a 4D spectrum is calculated from a pair of 3D spectra with correlations of the forms (I, J, K_1) and (L, M, K_1), as in Fig. 4, then the digital resolutions of the two K_1 dimensions must be matched before covariance. Furthermore, if there is a second pair of 3D spectra with correlations (I', J', K_2) and (L', M', K_2), then not only must the digital resolutions of the two K_2 dimensions match, but the resolution of I must match that of I', J must match J', L must match L', and M must match M'. Stated differently, the digital resolutions of each K dimension must match *within each pair* of covaried 3D spectra, and the digital resolutions of the I, J, L, and M dimensions must match respectively *across all pairs* of 3D spectra.

Figure 6 illustrates why and how the digital resolutions are matched. Specifically, we consider spectra with different spectral widths, carrier frequencies, and initial resolutions. Panel (a) represents a slice of 120 points along a ^{13}C dimension centered at 55 PPM with a spectral width of 30 PPM. Panel (b) represents a slice of 96 points centered at 58 PPM with a spectral width of 28 PPM. Both spectra contain a signal at 52 PPM. Beneath each spectrum is a series of dots representing the points at which each spectrum is sampled. Each space between dots represents eight points of the spectrum. Panel (c) demonstrates that although the two spectra represent the same underlying signal, they sample that signal at very different data coordinates. Panel (d) shows that if we compress the second spectrum to align each of its points with the first spectrum,

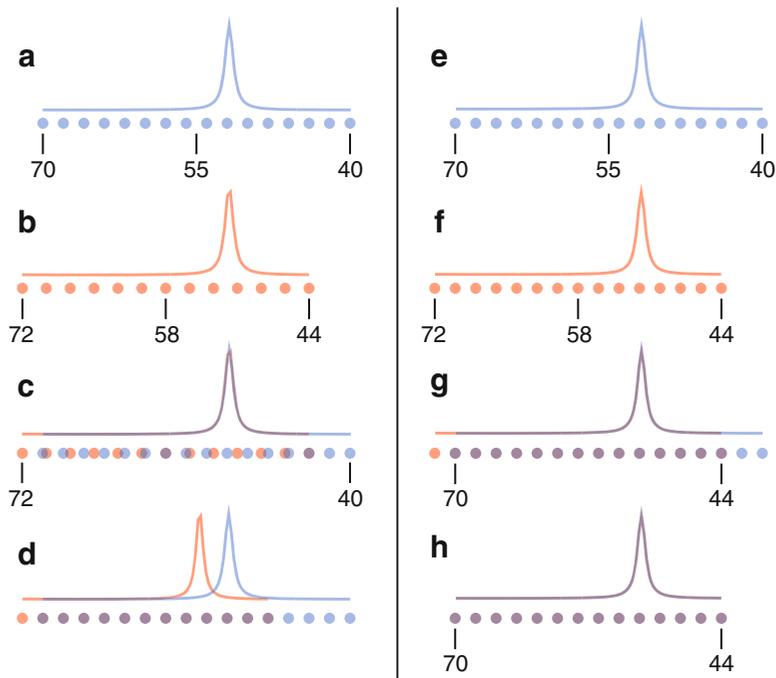


Fig. 6 Matching the digital resolution between spectra. The dots below the spectra correspond to the points at which the frequency axis is sampled. Each space between the dots represents eight sampled points. **(a)** A 1D spectrum of 120 points, with a spectral width of 30 PPM and centered at 55 PPM (digital resolution of 4.00 points per PPM). A single signal is present at 52 PPM. **(b)** A 1D spectrum of 96 points, with a spectral width of 28 PPM and centered at 58 PPM (digital resolution of 3.43 points per PPM). This spectrum also has a signal at 52 PPM. **(c)** When overlaying the two spectra from **(a)** and **(b)**, the signals properly align at 52 PPM. However, each spectrum is sampled at very different locations along the frequency axis. **(d)** If we compress the spectrum from **(b)** to align with the spacing of those in **(a)**, our signals no longer align. Multiplying the aligned points would not produce a correlation in the covariance spectrum. Consequently, element-wise multiplication is not well defined. **(e)** The same spectrum as in **(a)**. **(f)** The spectrum from **(b)** zero padded from 96 to 112 points before Fourier transformation. It now has a digital resolution of 4.00 as well. **(g)** When overlaying the signals from **(e)** and **(f)**, both the signals and the sampling locations align. **(h)** Each spectrum has been restricted to the region common to both spectra. Element-wise multiplication is well defined and will lead to a correlation in the covariance spectrum

the signals themselves no longer align. As the two spectra are currently processed, there is simply no way to correctly multiply them in an element-wise fashion. The solution is simply to zero pad the two spectra to reach the same digital resolution after Fourier transformation. In this case, the first spectrum contains 120 points over a spectral width of 30 PPM, giving a digital resolution of 4.00 points per PPM, whereas the second spectrum features 3.43 points per PPM. Zero padding the second spectrum from 96 points to 112 points will increase its digital resolution to 4.00 points per PPM, matching the first. It is important that the digital resolutions match exactly (*see Note 2*). The newly zero padded spectra in panels **(e)** and **(f)** are overlaid in panel **(g)**. Now both spectra sample the

underlying signal at the same locations. Finally, the limits of each spectrum are restricted to a common region, here 44 to 70 PPM as in panel (h), and element-wise multiplication is well defined. This last step is performed within our processing script.

Finally, users should be sure that all NMRPipe spectra are self-consistently calibrated in every dimension. Combining concepts from Figs. 3 and 6, we note that it is extremely important for each signal to align exactly with its counterpart in another spectrum when undergoing element-wise multiplication. Misalignment due to incorrect calibration, even by one point, can substantially reduce the amplitudes of covariance correlations. We advise users to find a set of clearly identifiable signals and use **nmrDraw** to verify that those signals occur at *exactly* the same chemical shift in every spectrum, usually by updating the values of CAR in **fid.com** scripts.

Once the spectra have been transposed and calibrated and the digital resolutions have been matched, the final consideration before running our script is the size of the resulting 4D spectrum. The size of the 4D spectrum can be calculated by multiplying the total number of points (the product of the dimension sizes I, J, L, and M) by the size of each point (4 bytes), and adding the size of the NMRPipe header (2048 bytes). Dividing this number by 2^{20} will give the size in megabytes (MB), whereas dividing it by 2^{30} will give the size in gigabytes (GB). Users should be aware that it may not be possible to manipulate and view large files properly in all software (*see Note 3*). If users determine that the size will be too large, they should either reduce the size of the input spectra in the I, J, L, and/or M dimensions or downsample the resulting 4D spectrum (*see below*). Reducing the size of the input spectra can be accomplished in three ways: extract only the regions of the spectrum relevant to analysis; reduce the amount of zero padding if any; or truncate the time-domain data at the cost of resolution. The option to downsample the 4D spectrum will be discussed further below.

3.2 Running the Script

Users should make a new copy of our script for each calculation. This has the beneficial effect of maintaining a record of processing parameters. Users will need to modify the eight parameters at the beginning of the file: *filenames*, *extract_IJLM*, *extract_K*, *filename_4D*, *labels_4D*, *lambda*, *with_mrs*, and *downsample*.

The variable *filenames* is a two-column cell array representing the input spectra. Each row specifies a pair of 3D spectra that will be covaried to produce a 4D spectrum. If multiple rows are present, the 4D spectrum from each row will be combined with all others using element-wise multiplication to form the final spectrum. Each element of the cell array is a string indicating the location of a 3D spectrum. The string may indicate a single file, if it points to a monolithic NMRPipe spectrum, or it may be a standard NMRPipe formatting string of the form “/path/to/test%03d.ft3.” Users

should note the order of their spectra in each row. The first column should always indicate the spectrum with dimensions I, J, and K, while the second column should always indicate the spectrum with dimensions L, M, and K. For example, a correlation map for sequential amide resonance assignment employing HNCA and HN(CO)CA, as well as HN(CA)CB and HN(COCA)CB, would require the lines:

```
filenames = {"/path/to/HNCA/ft3/test%03d.ft3", ...
"/path/to/HNCOCA/ft3/test%03d.ft3"; ...
"/path/to/HNCACB/ft3/test%03d.ft3", ...
"/path/to/HNCOACB/ft3/test%03d.ft3" ...
}
```

If different spectral widths have been used, the script automatically extracts the largest possible region for each dimension (e.g., 44 PPM to 70 PPM in Fig. 6). Alternatively, if a specific region is desired, users may specify the limits of the region, in PPM, for each dimension. This is particularly useful when calculating residue-specific covariance maps [20]. A single NMRPipe spectrum can be prepared, and various residue-specific regions can be extracted from it during each calculation.

The variables *extract_IJLM* and *extract_K* define the limits of extracted regions for the various dimensions that undergo element-wise multiplication. *extract_IJLM* is a 1 by 4 cell array, where each element is a 1 by 2 array representing the limits of extraction for the I, J, L, and M dimensions respectively. Be sure to verify that these limits fall within the bounds of each spectrum (*see Note 4*). If an array is empty, the script will default to the largest possible region for that dimension. If no specific regions are desired for any of the four dimensions, users can leave the entire cell array empty. *extract_K* defines the regions extracted along the covaried dimensions. It functions exactly as *extract_IJLM*, except that it is a 1 by P cell array, where P is the number of pairs of 3D spectra to be combined in the final 4D spectrum. This number should also correspond to the number of rows in the *filenames* cell array. Once again, if no specific regions are desired for any of the P different K dimensions, users can leave the entire cell array empty. The following examples are all valid inputs when P = 3:

```
extract_IJLM = {[6 10.5], [131 103], [10.5 6], [103 131]};
extract_IJLM = {[6 10.5], [], [10.5 6], []};
extract_IJLM = {[], [], [], []};
extract_IJLM = {};

extract_K = {[30 60], [15 75], [182.5 171]};
extract_K = {[], [15 75], []};
extract_K = {[], [], []};
extract_K = {};
```

The variable *filename_4D* represents the output spectrum. The spectrum may be output in one of three NMRPipe formats: a single, monolithic file; a series of 2D files; or a series of 3D files. If a monolithic file is desired, *filename_4D* should be a string specifying a single file, for example “/path/to/test.ft4.” In some cases, monolithic files may require an additional processing step after our script has run (*see Note 5*). If a series of 2D files is desired, then the string should be in the familiar NMRPipe format for 4D spectra, for example “/path/to/test_%03d_%03d.ft4.” Finally, if a series of 3D files is desired, the string should be similar to that of a 3D NMRPipe spectrum, for example “/path/to/test%03d.ft4.”

The variable *labels_4D* specifies an NMRPipe label for each dimension in the resulting 4D spectrum. It is a 1 by 4 cell array, where each element is a string of up to four characters. The order of the labels coincides with the order of the dimensions, namely I, J, L, and M. For example:

```
labels_4D = {'HN', 'N', 'HM', 'CM'};
```

The variable *lambda* corresponds to the parameter λ in the Generalized Indirect Covariance formalism [14] and reflects the power to which the covariance spectrum is taken during calculation. λ values of 1/2 were used to suppress the effects of pseudo-relay artifacts in symmetrical covariance of NOESY and TOCSY type experiments. In unsymmetrical covariance, the slope of a signal with respect to λ can be used to assess the signal’s veracity.

The Boolean variable *with_mrs* specifies whether maxima ratio scaling should be implemented. Maxima ratio scaling is a technique developed by Snyder and coworkers [15] to reduce the effects of inhomogeneous noise in covariance spectra. Because each point in a covariance spectrum is the product of two 1D slices, the noise at each point is modulated by the amplitude of the signals in the slices from which it came. This has a nonlinear effect on the distribution of the noise and can often create “ridges” of noise emanating from strong signals in the covariance spectrum. In a 4D spectrum, instead of ridges, this often manifests as varying noise intensities in each different plane of the spectrum. Turning on maxima ratio scaling can help to reduce this effect.

Finally, *downsample* is a parameter used to reduce the size of the output spectrum. Rather than limiting the size of the input spectra to reduce the size of the 4D spectrum, possibly broadening its signals, downsampling seeks to use the highest resolution input spectra possible and simply reduce the size of the 4D spectrum after the fact. It is applied to the dimensions I, J, L, and M and not to the subsumed dimension K. Downsampling reduces the number of points in a particular dimension by writing only every n^{th} point along that dimension. The variable *downsample* is a 1 by 4 cell array

with integers defining the amount of downsampling in each dimension (I, J, L, or M). A downsampling factor of one has no effect on the data. A downsampling factor of two indicates that the script should only output every 2nd point along that dimension. If downsampling is not desired in any dimension, simply leave the cell array empty. The following examples illustrate valid input:

```
downsample = {2, 2, 2, 2};
downsample = {2, 1, 2, 1};
downsample = {1, 1, 1, 1};
downsample = {};
```

While downsampling may seem like a promising solution to reduce data size, we caution users on its use. When signals are relatively sharp in the original data, some correlations may be severely reduced by the procedure. In general, we have found that it is safer to reduce the size of input spectra. Nevertheless, we have retained this option for users to experiment with.

3.3 Navigating and Interpreting 4D Covariance Spectra

Just as a 3D spectrum can be thought of as a series (line) of planes, a 4D spectrum can be thought of as a plane of planes. In a 3D spectrum, we often use a cursor to choose coordinates in two dimensions and view the corresponding 1D slice along the remaining dimension. In a 4D spectrum, choosing coordinates in two dimensions allows us to view the corresponding 2D plane along the remaining dimensions. A 4D spectrum viewer typically displays two planes of data side by side, each with its own cursor. The plane on the right is specified by the two coordinates from the cursor on the left and vice versa. Moving the cursor in one plane updates the data in the opposite plane based on the new coordinates.

Navigating a 4D spectrum is much like navigating a 3D spectrum. In a 3D spectrum, the two dimensions corresponding to a spin anchor are usually synchronized to an HSQC. For example, when viewing an HNCA, users often use an HN-HSQC to choose a signal of interest and then identify the location of its corresponding C^α chemical shift in a synchronized display of an HNCA spectrum. In a 4D spectrum, however, there are two spin anchors that can both be synchronized to HSQCs. Extending the 3D example to a 4D spectrum of (H^N, N, H^α, C^α) correlations, the H^N and N dimensions can be synchronized to an HN-HSQC while the H^α and C^α dimensions can be synchronized to an HC-HSQC. In this scenario, choosing a signal of interest in the HN-HSQC allows users to identify its corresponding signal in the HC-HSQC.

Figure 7 illustrates the navigation of a 4D covariance spectrum and demonstrates some of the complications that might arise when assigning resonances. Here, we discuss a situation exposing the limitations of pre- and postprocessing artifact suppression, and we present a solution to overcome these limitations during analysis.

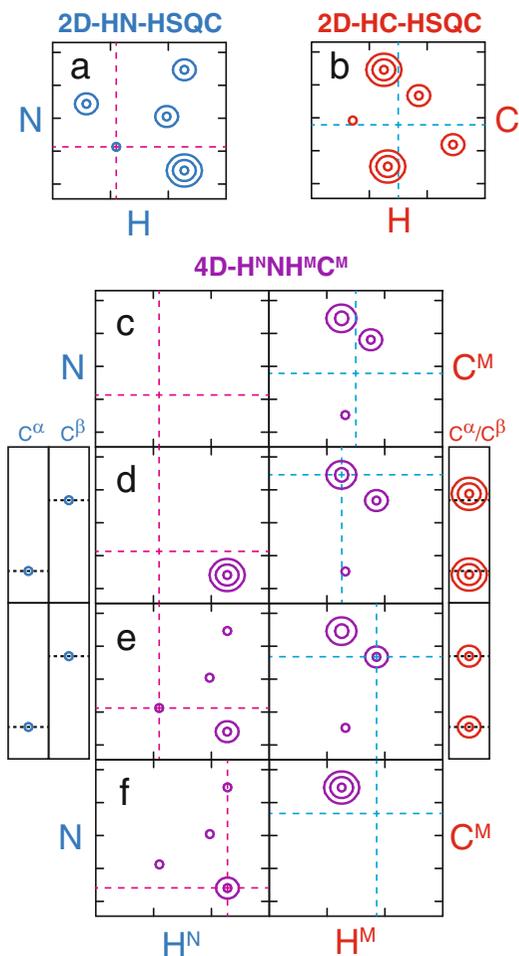


Fig. 7 Navigating a 4D spectrum. (a) Zoomed region of a 2D HN-HSQC. (b) Zoomed region of a 2D HC-HSQC. (c)–(f) Views of a 4D covariance spectrum connecting H^N spin anchors to $H^M C^M$ spin anchors through C^α and C^β . The *left* panes represent the H^N/N planes at the position of the H^M/C^M cursor on the *right* panes. The *right* panes depict the H^M/C^M planes at the position of the H^N/N cursor on the *left* panes. The panes to the *left* of panels (d) and (e) show strips from the HNCA and HN(CA)CB along the C^α and C^β dimensions at the position of the H^N/N cursor. Similarly, the panes to the *right* show strips from the HMCMBCA at the position of the H^M/C^M cursor. (c) The cursors begin synchronized to their respective positions in the 2D HSQCs. The *right* pane shows three possible methyl assignments for the amide signal selected in the HN-HSQC. No peak is selected in the HC-HSQC, so no signals are present in the *left* pane. (d) One of the assignment candidates is selected in the H^M/C^M plane revealing (H^N, N) correlations in the H^N/N plane but none at the coordinates of the cursor. The positions of signals in the strips do not match. (e) A second assignment candidate is selected in the H^M/C^M plane. Here, the positions of signals in the strips match and the methyl resonances have been assigned to the amide resonances. (f) Moving the H^N/N cursor to a peak discovered in (d) reveals its assignment candidates

The situation arises when a weak and a strong signal have nearly degenerate frequencies in the covaried dimension, and the (reduced) artifact correlation in the 4D covariance spectrum competes with the true, weak correlation. Figure 7 displays a cartoon representation of a 4D spectrum generated from HNCA, HN(CA)CB, and HMCMBCA spectra and connecting $H^N N$ spin anchors to $H^M C^M$ spin anchors in valine residues. Panels (a) and (b) display selected regions of the HN- and HC-HSQC associated with these data; note the differing peak intensities. In the following example, it will be assumed that the $H^N N$ spin anchors are fully assigned and that we are using the 4D covariance spectrum to assign their associated methyl resonances. To emphasize this strategy, we will refer to the H^N/N plane of the 4D as the fixed plane and the H^M/C^M plane as the explored plane. Panels (c) through (f) represent various states of the 4D spectrum throughout the assignment process.

Panel (c) represents a view of the 4D spectrum with the cursor positions of each plane synchronized with their respective HSQCs as defined in (a) and (b). In this initial configuration, the cursor in the fixed plane is set to investigate a weak signal observed in the HN-HSQC. As a result, we can identify several possible methyl candidates in the explored plane. On the other hand, because we have not set out to investigate a particular methyl signal, the cursor in the HC-HSQC is placed at a position containing only noise. Accordingly, the H^N/N plane has no visible signals.

Panel (d) illustrates the 4D spectrum after moving the cursor in the explored plane to its largest signal, updating the view of the fixed plane. We observe a strong amide signal in the fixed plane, but not at the H^N/N position of interest. In this scenario, the contour levels have been set according to the intense amide signal (*see Note 6*), and the signal at the (H^N, N) position of interest lies below the contour threshold. This is an immediate indicator that, in all likelihood, this is not the correct methyl assignment for the amide signal under investigation. We can verify this hypothesis by inspecting the original 3D data. The H^N/N cursor of the 4D spectrum is synchronized to those of HNCA and HN(CA)CB and the H^M/C^M cursor is synchronized to that of HMCMBCA, providing the strips shown to the left and to the right, respectively. Inspection of the carbon dimensions reveals that both C^α and C^β are nearly degenerate but clearly different for $H^N N$ and $H^M C^M$ anchors (horizontal lines). This false-positive correlation occurs in spite of the spectral derivative in the carbon dimensions because of the large amplitude of the signal in the HMCMBCA spectrum. Because such a scenario cannot be predicted, we use covariance spectra to supplement rather than supplant examination of the original data. That is, we maintain the synchronization with the original data (here HNCA, HN(CA)CB, and HMCMBCA) throughout the entire assignment procedure.

Panel (e) investigates the next-strongest signal in the explored plane. In addition to the previously discussed signal in the fixed plane, we can now observe a small signal at the position of interest. However, we also note that there are many such small signals in the fixed plane (*see Note 7*). Examination of the original data reveals identical C^α and C^β resonance frequencies in the associated strips. Thus, after only investigating two candidate resonances, we have assigned a methyl signal for this Valine residue without ever having picked a peak. Simultaneously, in panel (d), we have identified a candidate amide signal for a second methyl. This assignment is confirmed in panel (f), where the cursor in the fixed H^N/N plane has been moved to the amide candidate, revealing a single signal in the H^M/C^M explored plane. Importantly, if the explored plane is devoid of any signals, it indicates conclusively that no such correlation exists in the raw data. This is in contrast to peak picking techniques where it is not possible to distinguish between truly missing signals and unpicked peaks.

4 Notes

In the following section, we will outline some of the common problems users might experience when working with 4D covariance spectra and our script.

1. In all types of covariance NMR processing, it is very important that the dimension(s) subsumed during covariance be processed extremely carefully. In our case, we refer to the K dimension of each 3D spectrum (Fig. 4). As discussed in Sub-heading 1.2, each point in the 4D covariance spectrum is the inner product of a slice along the K dimension from each 3D spectrum. If there are any artifacts or unwanted signals in either of these slices, they will result in unwanted signals in the covariance spectrum. For instance, a baseline shift in two empty slices will create a correlation in the covariance spectrum even when no signals exist in them. Artifacts of any type in the subsumed dimension (improper phasing, baseline shifts, baseline rolls, unsuppressed water, t_1 noise, axial peaks, etc.) will all result in artifacts in the covariance spectrum. In some cases, it may be necessary to remove portions of the spectrum to recover a good covariance spectrum. For example, if the detected dimension is to be covaried, it may be necessary to remove the region containing water for the purposes of covariance. Similarly, regions selected in indirect, covaried dimensions should be devoid of axial peaks.

2. When matching digital resolutions between dimensions, it is important to note that even small differences can still give rise to errors. For example, a digital resolution of 4.00 points per PPM and a spectral width of 25 PPM would result in a spectrum of 100 points. However, a digital resolution of 4.02 points per PPM and the same spectral width of 25 PPM would result in a spectrum of approximately 100.5 points, which would be rounded up to 101 points. In such a case, users might encounter a “Resolution mismatch” error without realizing the source of the problem.
3. In Subheadings 3.1 and 3.3 we discussed issues pertaining to the potentially large file sizes of 4D spectra. While the file size is not an issue in itself, many software programs cannot handle or underperform with large files. Notably, older versions of NMRPipe may have difficulties handling large files. Users can circumvent this issue by upgrading to the latest version of NMRPipe.
4. In some cases, the nominal value for the edge of a spectrum may be slightly different from its true value. For example, if a user extracted the region from 10 to 6 PPM from the H^N dimension of a spectrum, the true right-edge of the spectrum may in fact be at 6.002 PPM. If a user then tried to extract a sub-region with our script, using the `extract_IJLM` variable, a value of 6 PPM as a limit of extraction might result in an error indicating that the limit falls outside the range of the spectrum. Simply increasing this value to 6.002 would resolve the issue.
5. Although our script can produce monolithic files for 3D and 4D NMRPipe spectra, Windows users will have to run the command “`sethdr filename.ft4 -pipeFlag 1`” after processing. We managed to avoid this constraint for Mac and Linux but not for Windows. When used in Windows, the script will issue a warning reminding users to perform this task.
6. The dynamic range in covariance spectra is larger than that of normally acquired spectra. This property results from the multiplication of signals that occur when performing covariance or when multiplying 4D maps to remove artifacts. We have taken care to scale the spectra during processing in order to cap the maximum amplitude of signals, thereby avoiding cases of floating-point “overflows” where a point exceeds $3.4e + 38$ and becomes `+Inf`. However, users should expect a large dynamic range when examining covariance spectra and may have to frequently change the contour levels as needed. Such changes in contour levels were assumed when creating the cartoon spectrum in Fig. 7.

7. When assigning correlations involving a weak resonance correlated with a strong signal in original spectra, it is often much easier to search for candidates by selecting the dimensions of the weak peak as the “fixed” plane and search for new correlations in the explored plane with dimensions of the strong signal. In effect, the amplitude of the entire explored plane (including noise) reflects the amplitude of the signal in the fixed plane. In our experience, the correct assignment is more likely to stand out when searching in this direction. In the other direction, many false-positive correlations may be of equal or greater height than the true correlation. Panel (e) in Fig. 7 demonstrates what such an example might look like.

Acknowledgments

The Frueh lab is supported by the National Institute of Health, grant R01GM104257.

References

1. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618–622. doi:10.1038/nature05512
2. Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* 125:13868–13878. doi:10.1021/ja030345s
3. Brüschweiler R, Zhang F (2004) Covariance nuclear magnetic resonance spectroscopy. *J Chem Phys* 120:5253–5260. doi:10.1063/1.1647054
4. Brüschweiler R (2004) Theory of covariance nuclear magnetic resonance spectroscopy. *J Chem Phys* 121:409–414. doi:10.1063/1.1755652
5. Trbovic N, Smirnov S, Zhang F, Brüschweiler R (2004) Covariance NMR spectroscopy by singular value decomposition. *J Magn Reson* 171:277–283. doi:10.1016/j.jmr.2004.08.007
6. Zhang F, Brüschweiler R (2004) Spectral deconvolution of chemical mixtures by covariance NMR. *ChemPhysChem* 5:794–796. doi:10.1002/cphc.200301073
7. Zhang F, Brüschweiler R (2004) Indirect covariance NMR spectroscopy. *J Am Chem Soc* 126:13180–13181. doi:10.1021/ja047241h
8. Blinov KA, Larin NI, Kvasha MP et al (2005) Analysis and elimination of artifacts in indirect covariance NMR spectra via unsymmetrical processing. *Magn Reson Chem* 43:999–1007. doi:10.1002/mrc.1674
9. Blinov KA, Larin NI, Williams AJ et al (2006) Long-range carbon-carbon connectivity via unsymmetrical indirect covariance processing of HSQC and HMBC NMR data. *Magn Reson Chem* 44:107–109. doi:10.1002/mrc.1766
10. Kupče E, Freeman R (2006) Hyperdimensional NMR spectroscopy. *J Am Chem Soc* 128:6020–6021. doi:10.1021/ja0609598
11. Lescop E, Brutscher B (2007) Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J Am Chem Soc* 129:11916–11917. doi:10.1021/ja0751577
12. Benison G, Berkholz DS, Barbar E (2007) Protein assignments without peak lists using higher-order spectra. *J Magn Reson* 189:173–181. doi:10.1016/j.jmr.2007.09.009
13. Chen K, Delaglio F, Tjandra N (2010) A practical implementation of cross-spectrum in protein backbone resonance assignment. *J Magn Reson* 203:208–212. doi:10.1016/j.jmr.2009.12.018
14. Snyder DA, Brüschweiler R (2009) Generalized indirect covariance NMR formalism for establishment of multidimensional spin correlations. *J Phys Chem A* 113:12898–12903. doi:10.1021/jp9070168
15. Snyder DA, Ghosh A, Zhang F et al (2008) Z-matrix formalism for quantitative noise assessment of covariance nuclear magnetic resonance spectra. *J Chem Phys* 129:1–9. doi:10.1063/1.2975206

16. Snyder DA, Zhang F, Brüschweiler R (2007) Covariance NMR in higher dimensions: application to 4D NOESY spectroscopy of proteins. *J Biomol NMR* 39:165–175. doi:[10.1007/s10858-007-9187-1](https://doi.org/10.1007/s10858-007-9187-1)
17. Snyder DA, Xu Y, Yang D, Brüschweiler R (2007) Resolution-enhanced 4D $^{15}\text{N}/^{13}\text{C}$ NOESY protein NMR spectroscopy by application of the covariance transform. *J Am Chem Soc* 129:14126–14127. doi: [10.1021/ja075533n](https://doi.org/10.1021/ja075533n)
18. Harden B, Nichols S, Frueh D (2014) Facilitated assignment of large protein NMR signals with covariance sequential spectra using spectral derivatives. *J Am Chem Soc* 136:13106–13109
19. Harden BJ, Mishra SH, Frueh DP (2015) Effortless assignment with 4D covariance sequential correlation maps. *J Magn Reson* 260:83–88. doi:[10.1016/j.jmr.2015.09.007](https://doi.org/10.1016/j.jmr.2015.09.007)
20. Mishra SH, Frueh D (2015) Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps. *J Biomol NMR* 62:281–290. doi:[10.1007/s10858-015-9943-6](https://doi.org/10.1007/s10858-015-9943-6)
21. Keller R (2004) The computer aided resonance assignment tutorial. Cantina Verlag, Goldau (Switzerland), p 225
22. Short T, Alzapiedi L, Brüschweiler R, Snyder D (2011) A covariance NMR toolbox for MATLAB and OCTAVE. *J Magn Reson* 209:75–78. doi:[10.1016/j.jmr.2010.11.018](https://doi.org/10.1016/j.jmr.2010.11.018)
23. The MathWorks Inc. (2014) MATLAB
24. Eaton JW, Bateman D, Hauberg S, Wehbring R (2015) GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations. ISBN 1441413006, URL <http://www.gnu.org/software/octave/doc/interpreter/>

Structures of Dynamic Protein Complexes: Hybrid Techniques to Study MAP Kinase Complexes and the ESCRT System

Wolfgang Peti, Rebecca Page, Evzen Boura, and Bartosz Różycki

Abstract

The integration of complementary molecular methods (including X-ray crystallography, NMR spectroscopy, small angle X-ray/neutron scattering, and computational techniques) is frequently required to obtain a comprehensive understanding of dynamic macromolecular complexes. In particular, these techniques are critical for studying intrinsically disordered protein regions (IDRs) or intrinsically disordered proteins (IDPs) that are part of large protein:protein complexes. Here, we explain how to prepare IDP samples suitable for study using NMR spectroscopy, and describe a novel SAXS modeling method (ensemble refinement of SAXS; EROS) that integrates the results from complementary methods, including crystal structures and NMR chemical shift perturbations, among others, to accurately model SAXS data and describe ensemble structures of dynamic macromolecular complexes.

Key words Intrinsically disordered proteins (IDP), NMR spectroscopy, SAXS, EROS, Ensemble

1 Introduction

Hybrid methods are increasingly being used to molecularly characterize the macromolecular machines that control all cellular processes. These complementary techniques include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy, small angle X-ray scattering (SAXS), small angle neutron scattering (SANS), mass spectrometry, chemical cross-linking, super-high-resolution optical microscopy, optical microscopy, fluorescence spectroscopy, and electron paramagnetic resonance (EPR) spectroscopy, among others. In particular, SAXS, a technique commonly used to characterize the shape(s) and dimensions of proteins in solution [1, 2], is increasingly used as a hybrid method in structural biology, especially as a complimentary technique with NMR spectroscopy and X-ray crystallography. This is especially true for the structural investigation of intrinsically disordered proteins (IDPs), IDP complexes, and large complexes

with large intrinsically disordered regions (IDRs), all of which are difficult or impossible to crystallize and are also difficult to study using cryo-EM.

Like most NMR measurements, SAXS experiments are performed on samples in aqueous solutions. As a result, the macromolecules that scatter X-rays in SAXS experiments are oriented randomly relative to the incident beam. This results in the spherical averaging of the signal, and as a consequence the diffraction image depends on only a single scattering angle. In this way, three-dimensional molecular structures are reduced to one-dimensional intensity profiles. Despite this loss of information, the scattering intensity profiles can be used to determine the molecular shapes and dimensions of the biomacromolecule under investigation. Because of this, NMR spectroscopy and SAXS experiments are frequently used concurrently to study IDPs, IDP:protein complexes as well as for multi-domain proteins with long IDRs. However, to model these proteins and protein complexes that exhibit considerable conformational fluctuations, it is necessary to employ methods that use multiple structural models to fit experimental data as an ensemble rather than a single structure, as commonly done for X-ray crystallography. Thus, multiple SAXS-based approaches for modeling conformational ensembles have been recently developed, including the ensemble optimization method [3], the minimal ensemble search method [4], the SAXS module in the integrative modeling platform [5], and the ensemble refinement of SAXS (EROS) method [6]. Critically, all these approaches require a vast pool of diverse protein conformations as input; i.e., extensive sampling of the ensemble/conformation pool. This is essential as a conformation pool that is not inclusive enough is likely to be incapable of accounting for the experimental data. This conformation pool can be generated based on steric exclusion [3], high-temperature molecular dynamics simulations [4], statistical potentials for protein binding [5, 6], or topology-based (Go-type) models [7]. After generating this conformation pool, heuristic algorithms are used to determine the combination of conformations that optimally fit the experimental SAXS data. Additional experimental constraints (e.g., NMR constraints, FRET, among others) can be used to either modify the ensemble pool or to determine the conformations that allow for the best SAXS data fit.

EROS uses a different strategy; namely, the pool of simulated structures is gently reweighed to improve the agreement with the SAXS data. It can also incorporate data from other methods, including chemical shift perturbations from NMR spectroscopy, FRET, and EPR distance data. Here, we will provide a detailed description of how to express and purify IDPs, collect high-quality SAXS data on these samples, and finally we describe how hybrid structures—ones that combine SAXS, NMR, and crystallographic data—are obtained using the EROS method (Fig. 1).

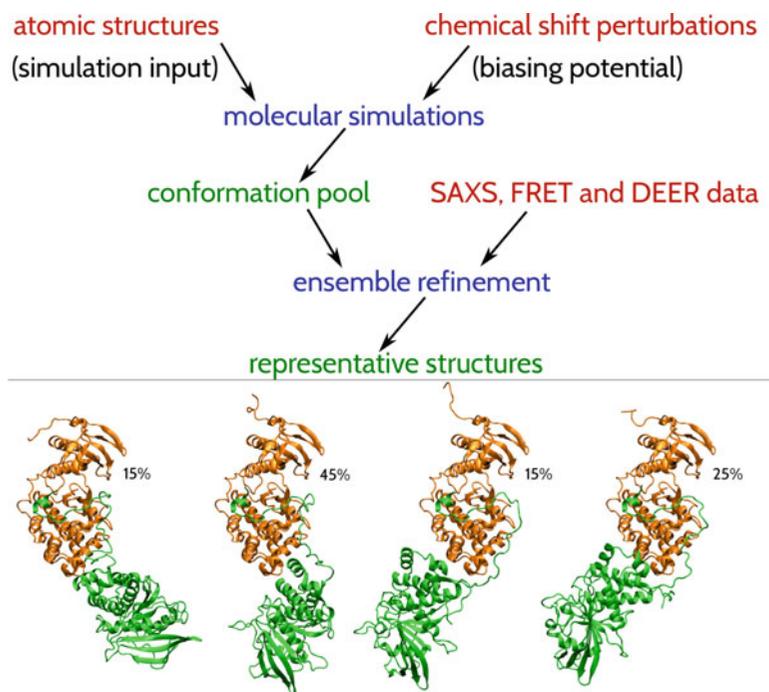


Fig. 1 Diagram showing the integration of experimental (crystallography, NMR, SAXS, FRET, EPR; marked in *red*) and computational data (molecular simulations, ensemble refinement procedures; marked in *blue*) to determine representative ensemble structures of protein complexes containing IDPs and/or IDRs in EROS

2 Materials

Detailed methods for obtaining chemical shift perturbation (CSP) data on large, dynamic complexes have recently been described [8]. Here, we focus on sample preparation and data collection for SAXS, with a focus on strategies for enhancing the expression and stability of IDPs for NMR and SAXS measurements. Finally, we described how to combine these data with NMR derived CSPs using EROS to obtain accurate models of IDP ensembles.

Prepare all solutions using ultrapure Water (Milli-Q water purification system, Millipore). Chemicals should be at least ACS grade. Prepare and store reagents/solutions at temperatures and conditions recommended by the manufacturer. Standard buffers are used for purification; when necessary, buffers are autoclaved when prepared to prevent unwanted proteolytic degradation due to the presence of trace amounts of proteases. This is particularly important for the study of IDPs, which are highly susceptible to proteolytic degradation. All buffers are stored at 4 °C and are filtered (0.22 µm PES filter, Millipore) immediately prior to use. When possible, use uniform buffers throughout all the experiments. Standard water baths and/or heat blocks that can be heated to 90 °C are required for heat purification steps.

3 Methods

3.1 IDP Expression and Purification

IDPs and IDRs are estimated to comprise more than 30% of the human genome [9]. It is now clear that IDPs play essential roles in multiple biological processes, but especially for signaling. However, their lack of a single stable structure often renders them highly susceptible to proteolytic degradation in the laboratory. This requires the implementation of additional experimental steps to ensure they can be expressed to high levels and are stable, which allows them to be studied at a molecular level using techniques such as NMR spectroscopy and SAXS (*see Note 1*).

1. *E. coli* expression plasmids: Typically, expression plasmids are used that allow IDPs to be fused to N-terminal tags that facilitate both expression (maltose-binding protein, MBP, and glutathione-S-transferase, GST) and purification (6xHis, tobacco etch virus [TEV] protease sequence) (*see Notes 2–4*).
2. Expression: Standard methods are used to express IDPs in *E. coli* [10].
3. Protease inhibition during cell lysis and purification: Standard methods are used for cell lysis and purification [10]. However, because IDPs are extremely sensitive to proteolytic degradation, additional steps are used to minimize protease exposure.
 - (a) Protease inhibitors (e.g., EDTA-free Complete tabs, Sigma-Aldrich) are added to all lysis buffers and, if needed, purification buffers.
 - (b) Buffers: Purification buffers are autoclaved prior to use.
 - (c) Columns and purification systems: All columns and purification system tubing are rigorously cleaned using 1 M NaOH (1 column volume [CV]) and 30% isopropanol/water (0.5 CV) prior to use.
 - (d) Elution collection: fraction collection tubes/blocks are autoclaved prior to use.
4. N-terminal tag cleavage: Dialyze the purified IDP with TEV protease for N-terminal tag cleavage (*see Note 5*) using standard protocols [10].
5. Heat purification: Because IDPs do not adopt a single, folded conformation, they are often heat stable. This provides a unique opportunity both for purification from folded N-terminal fusion tags (MBP/GST) and for minimizing proteolytic exposure (*see Note 6*).
 - (a) Transfer dialysate to 50 mL conical vial.
 - (b) Incubate dialysate in a water bath (65 °C) for 15 min.
 - (c) Centrifuge at $10,000 \times g$ for 15 min to separate soluble and insoluble fractions (*see Note 7*).

- (d) In a second step, incubate soluble fraction at 90 °C (*see Note 8*) for 15 min.
- (e) Repeat “**step c**” (*see Note 9*).
6. Purify the IDPs using size-exclusion chromatography (SEC) to remove any remaining contaminant proteins and aggregates (*see Note 10*).
7. In a final step, heat (90 °C) the pooled, concentrated IDP to denature any trace proteases (*see Note 11*).
8. If **steps 1–7** do not overcome IDP proteolytic degradation during purification, additional steps, including adding protease inhibitors at every step of the purification procedure and minimizing the time from lysis to final heat purification (i.e., <12 h) can also increase IDP protein yield and stability.

3.2 SAXS Experiments

The use of SAXS data in molecular modeling has a number of advantages, the most significant of which is that SAXS experiments are performed on samples in aqueous solutions. Thus, SAXS provides information about the conformations of macromolecules in their natural environment. Improper data processing can lead to errors. While this is certainly true for any method, SAXS is exceedingly sensitive to errors. For example, the SAXS intensity profile is the difference in signals between the sample and the corresponding buffer; inadequate signal subtraction can lead to significant systematic errors in the resulting profile. Furthermore, SAXS-based modeling must take into account the protein hydration shell. However, various algorithms that compute SAXS intensity profiles from atomic models, such as CRY SOL [11], FoXS [12], AXES [13], AquaSAXS [14], and SASTBX [15], differ in how they treat the hydration shell, putting additional uncertainty on SAXS-derived models.

Sample preparation for SAXS experiments requires neither crystal growth nor protein labeling. Unlike X-ray crystallography, which relies on diffracting crystals, macromolecules in solution always scatter X-rays. Similarly, unlike solution NMR techniques that have some molecular size limitations, SAXS is not limited by the molecular mass; rather larger proteins will scatter better. Furthermore, the quality of SAXS data depends neither on the size nor on the flexibility of the macromolecules under study [16].

SAXS measurements can be performed using a home X-ray source or, more typically, synchrotron radiation. They are performed on samples in a wide range of solution conditions, molecular concentrations, and temperatures. For all SAXS experiments, optimal sample preparation is essential for obtaining interpretable SAXS data. In particular, SAXS is exceptionally sensitive to aggregation, as soluble aggregates, even if they represent less than 1% of the sample, are significantly larger and thus will have a major impact on the overall measured scattering. Thus, identifying conditions

that prevent sample aggregation is essential (*see Note 12*). Detailed protocols for SAXS methods (including strategies for detecting aggregation and minimizing radiation) have been recently summarized [16]. Here, we focus on SAXS data collection for IDPs and IDR containing proteins.

1. Sample production: Purify all samples immediately prior to SAXS measurements using size-exclusion chromatography (SEC) to remove trace aggregates.
2. When possible, filter any trace aggregates immediately prior to SAXS measurements. For this, 0.02 μm syringe filters (GE Healthcare Anotop 10) are suitable (*see Note 12*).
3. Sample cells: Sample cells should be thoroughly cleaned prior to use to eliminate trace proteases (i.e., NaOH washes). Prior to measurements, the cells should subsequently be thoroughly washed with SAXS buffer (*see Note 13*).
4. Sample concentration: The optimal concentration for SAXS measurements depends on the X-ray source, the size, and assembly of the cell (flow-through or static), among other parameters. Typically, measurements are initiated using the lowest concentrations possible. Samples are then concentrated and the SAXS data collected. This is continued until the sample is concentrated as high as can be achieved before aggregation is detected (sometimes this can be as high as 30 mg/mL) (*see Note 14*).
5. SAXS data analysis: Numerous software packages are available to analyze SAXS data, with the two most widely used being ATSAS [17] and SCATTER (<https://bl1231.als.lbl.gov/scatter/>), both of which also allow for the calculation of 3D envelopes from the data; these calculations are usually performed with the highest signal/noise dataset (commonly, the highest measured concentration). This can also be done using Fast-SAXS-pro [18].
6. IDP detection: The Kratky plot, i.e., the plot of $q^2I(q)$ as a function of the momentum transfer q is used to identify IDPs (*see Note 15*).
 - (a) Convergence of the Kratky plot at high q suggests compaction, whereas a hyperbolic shape suggests flexibility [19]; the hyperbolic feature is a trademark of random coils and IDPs.
 - (b) In practice, Kratky plots may be difficult to assess if the SAXS data are noisy or truncated. Recently, analysis based on the Porod-Debye law, i.e., analysis of $q^4I(q)$ versus q^4 at intermediate q -values, has been introduced as a more robust approach to tell apart flexible molecules from rigid ones [19].

- (c) Molecular flexibility can also be presumed if SAXS data cannot be accounted for with a single model, suggesting that an ensemble of models may be required to fit the experimental data [20, 21].
7. Data processing and the need for ensemble modeling approaches: The standard approach for SAXS data analysis is to analyze the scattering intensity profile, $I(q)$, which enables the determination of the pair-distance distribution function, $P(r)$, and the corresponding molecular envelope [17, 22, 23]. Molecular envelopes provide an informative visual interpretation of the SAXS data; however, this approach holds only for rigid systems with minor ensemble fluctuations. When SAXS is used to study IDPs, this standard envelope calculation fails [24]. Furthermore, SAXS can be used to determine structures of protein complexes if atomic structures of the constituent proteins are known [19]. However, to achieve this goal with optimal accuracy, structural models of the protein complexes should be fitted directly to the experimental SAXS data; simply placing the protein models into molecular envelopes does not fully use the structural information encoded in the scattering intensity profile $I(q)$.

3.3 Ensemble Refinement of SAXS (EROS) Method

EROS was developed from the outset to combine SAXS with other spectroscopy methods, especially those that use site-directed labeling [20, 25], such as fluorescence and electron paramagnetic resonance (EPR) spectroscopy. In this way, data from various biophysical experiments can be readily combined and used simultaneously for molecular modeling. For example, by combining X-ray crystallography (which provides high-resolution structures of individual domains) with SAXS data (which provides information on the global size and shape of the molecular assembly) and NMR chemical shift, DEER (EPR) or FRET (which enables imposing local restraints on distances between selected sites) data, it is possible to obtain detailed representations of the structures in a variety of biologically important systems such as the ESCRT membrane-protein trafficking system [20, 25] and protein kinase complexes with their regulatory phosphatases [21, 26].

Different from other SAXS ensemble analysis methods, in EROS, the pool of simulated structures is only moderately reweighted to improve the agreement with the experimental SAXS data. The maximum-entropy method is used to prevent data overfitting.

1. Generating a starting model: A structural model of the protein system/complex under investigation is constructed using atomic structures (PDB files) of the constituent proteins or/and domains. If individual experimental structures are unavailable, homology models can be used. All flexible loops and inter-domain linkers, which are often missing in atomic

structures (e.g., due to the lack of electron density), must be built using programs such as MODELLER [27].

2. Generating the EROS ensemble: An ensemble of structural models is generated using molecular dynamics simulations; the structural model obtained from **step 1** (above) is used as input for the simulations.
 - (a) All-atom molecular dynamics simulations of large proteins are computationally demanding, especially when it comes to simulating large conformational fluctuations in flexible protein systems. Thus, EROS uses a more efficient, coarse-grained approach [28] to gain speed and to increase sampling of ensemble structures.
 - (b) EROS simulations of multiple protein complexes have been performed using in-house software but many coarse-grained protein simulation packages that are freely available [29, 30] can be used to generate starting ensembles (*see Note 16*).
 - (c) These protein simulations can be positively biased using experimental data [31]. For example, we used this approach for the simulations of the p38 α :HePTP complex, where NMR chemical shift perturbation constraints were incorporated into the energy function via a weak bias potential acting between the dynamic HePTP-KIS linker and residues on the surface of p38 α [21].
3. Scattering intensity profile: For each of the simulated structures obtained in **step 2**, a scattering intensity profile is computed. Different algorithms are available to compute SAXS intensity profiles on the basis of protein coarse-grained representations [32]. The EROS method uses a particularly simple approach, which assumes constant form-factors of the amino-acid beads [6], which makes this step simpler and faster compared to many other approaches.
4. Comparison with experiment: For each of the simulated structures additional parameters can be calculated and directly compared with experimental results, e.g., from FRET efficiencies or DEER dipolar evolution functions (*see Note 17*). However, for this approach it is necessary to model the fluorescence or spin labels onto the protein surface. Either rotamer libraries [33] or molecular dynamics simulations [34, 35] can be used to generate a pool of possible conformations of the fluorescence or spin labels.
5. Cluster generation: Simulation structures are sorted into clusters based on their mutual similarity.
 - (a) Standard clustering algorithms, such as k-means [36] or QT-clustering method [37], are typically sufficient.

- (b) However, it is important to choose an appropriate metric to cluster the simulated structures. Indeed, many quantities can be used as a measure of similarity between protein structures. The most common one is the root-mean-square deviation (RMSD) of atomic positions.
- (c) To compute the RMSD it is necessary to superimpose structures, which can be problematic in the case of flexible protein systems. For this reason, EROS employs the distance root-mean-square (DRMS) analysis. The DRMS between structures A and B is defined as follows

$$\text{DRMS}(A, B) = \left(\frac{1}{N_2} \sum_{n,m} \left(d_{n,m}^{(A)} - d_{n,m}^{(B)} \right)^2 \right)^{1/2} \quad \text{where } d_{n,m}^{(A)}$$

is the Cartesian distance between the amino-acid beads n and m in structure A, and N_2 is the number of bead pairs over which the sum is performed.

6. Assign measurable quantities to the clusters of the simulation structures. Use the results obtained in **steps 3–5**.

- (a) The SAXS intensity $I_k(q)$ assigned to cluster number k is the arithmetic mean of SAXS intensities resulting from all individual structures in this cluster.
- (b) By analogy, FRET efficiencies or DEER dipolar evolution functions assigned to a given cluster are arithmetic averages of FRET or DEER signals resulting from all structures in this cluster.

7. Assign statistical weights to the clusters obtained in **step 5**.

- (a) Use normalized weights, which fulfill the condition $\sum_k w_k = 1$, where w_k denotes the statistical weight of cluster number k . The average SAXS intensity profile resulting from the whole ensemble of simulation structures is now given by a weighted average over the clusters, i.e., $I_{\text{sim}}(q) = \sum_k w_k I_k(q)$, where $I_k(q)$ are the SAXS intensity profiles assigned to the individual clusters in **step 6**.

- (b) $I_{\text{sim}}(q)$ depends on the set of weights w_k assigned to the clusters. Also other ensemble-averaged quantities such as FRET efficiencies or DEER dipolar evolution functions, which should be compared directly to experimental data, depend on the cluster weights.

- (c) The discrepancy between the computed, ensemble-averaged intensity profile $I_{\text{sim}}(q)$ and the experimental SAXS data $I_{\text{exp}}(q)$ can be quantified by $\chi_{\text{SAXS}}^2 = \frac{1}{N_q} \sum_{i=1}^{N_q} \frac{(c I_{\text{sim}}(q_i) - I_{\text{exp}}(q_i))^2}{\sigma^2(q_i)}$ where the scale factor c results from the condition $\partial \chi_{\text{SAXS}}^2 / \partial c = 0$.

- (d) The discrepancy between the computed, ensemble-averaged FRET or DEER signals and the data from FRET or DEER experiments, respectively, can be quantified by analogous expressions.
- (e) The resulting model-data discrepancy $\chi^2 = \chi_{\text{SAXS}}^2 + \chi_{\text{FRET}}^2 + \chi_{\text{DEER}}^2$ is a function of the statistical weights of the clusters. Note that re-weighting the clusters can result in a decrease in the discrepancy χ^2 between simulations and experiments.
8. Fit the simulation ensemble to experimental data by optimizing the statistical weights of the clusters.
- (a) Before any refinement, the statistical weights are proportional to cluster populations. That is, if cluster number k consists of n_k structures, its initial weight is $w_k^{(0)} = n_k / \sum_k n_k$. In the course of the ensemble refinement, the cluster weights are varied to improve agreement with experimental data.
- (b) To re-weight the clusters in a controlled way, and to prevent data over-fitting, a minimum entropy method can be used. This method is based on numerical minimization of a pseudo-potential function $F = \chi^2 - S$ that consists of the model-data discrepancy function, χ^2 as introduced in **step 7**, and a cross-entropy term $S = -\frac{1}{\beta} \sum_k w_k \ln \frac{w_k}{w_k^{(0)}}$ that quantifies how far the refined ensemble is from the original simulation ensemble. Here, β is a control parameter. Including the two terms, χ^2 and S , in the pseudo-potential function F reflects our confidence in both experiments and simulations. The function F can be minimized with respect to the statistical weights w_k by using simulated annealing [6] or more advanced algorithms such as COPER [38]. As a result, the optimal weights $w_k^{(\beta)}$ are obtained for a given value of parameter β .
- (c) For sufficiently small β -values, when χ^2 is negligible in comparison to S , minimization of function F leads to only small changes in the initial weights, i.e., $w_k^{(\beta)} \approx w_k^{(0)}$ for most of the clusters. In contrast, for large values of β , when $F \approx \chi^2$, minimizing F leads to the best possible agreement with experiment but may result in data over-fitting.
- (d) Therefore, a sensible approach is to determine such a value of parameter β for which minimization of F yields $\chi^2 \approx 1$. Note that if the simulation structures correctly capture the relevant conformations of the protein system under study

then the condition $\chi^2 \approx 1$ is obtained when $w_k^{(\beta)} \approx w_k^{(0)}$ for most of the clusters.

9. Optional: Refine the structural ensemble using an alternative method and compare the outcomes [21].
 - (a) Another way of refining the simulation ensemble is the minimum ensemble method that selects the smallest possible set of clusters that accounts for experimental data [20, 21].
 - (b) In this approach, another function $G = \chi^2 + \mu N$ is minimized numerically. Here, N is the number of clusters with nonzero weights, $w_k > 0$, and μ is another control parameter, which should be fine-tuned in such a way that minimization of function G leads to $\chi^2 \approx 1$.
 - (c) The advantage of this method is that it usually produces only a small set of representative structures that can be easily inspected visually. However, by discarding a significant portion of the simulation ensemble, the minimum ensemble method does not fully use the predictive power of molecular simulations.
10. Cross-validate the structural ensemble using independent datasets excluded from refinement [20, 25].

4 Notes

1. All the steps below are suitable for IDPs and folded proteins that contain extended IDP domains, with the exception of the heat purification steps, which are appropriate *only* for IDPs (i.e., those sample *without* folded domains).
2. IDPs expressed in the absence of N-terminally fused folded proteins (i.e., just a 6xHis-TEV sequence) are often rapidly degraded by the bacterial proteolytic machinery in *E. coli*. In these cases, fusion to a large folded protein (MBP or GST) often overcomes this problem.
3. For this reason, we routinely use 1 of 3 pET-based expression plasmids that contain different N-terminal tags for IDP expression: (1) 6xHis-TEV-IDP, (2) 6xHis-MBP-TEV-IDP, or (3) 6xHis-GST-TEV-IDP.
4. TEV protease rarely exhibits nonspecific cleavage and thus is more suitable for IDP tag cleavage than other proteases, such as Factor Xa or thrombin.
5. Because the samples are IDPs, complete cleavage is often achieved in less time than for folded proteins, sometimes as soon as just a few hours. TEV cleavage time courses can be performed to minimize this step.

6. Not all IDPs can be readily heat purified; rather it must be determined experimentally.
7. We typically prepare aliquots of both the soluble and insoluble samples for SDS-PAGE gel analysis.
8. The maximum temperature use for heat purification is specific to each IDP and *must* be determined experimentally. Typically, for a new IDP purification, temperature intervals of 5 °C will be tested to identify the maximal temperature at which the IDP remains soluble (commonly from 60 to 90 °C).
9. Greater than 90% of the MBP, GST, and TEV precipitates at temperatures of ≥ 70 °C. Thus, heat purification is highly effective for removing large, folded proteins when working with IDPs.
10. Because IDPs are not globular, they sometimes elute in positions expected for “larger” proteins during the SEC step.
11. This last heat purification step maximizes the long-term stability of IDPs, as it substantially slows/prevents proteolytic degradation, which is often important for NMR spectroscopy and SAXS analysis steps.
12. Typically, we prepare 4–8 L of the appropriate buffer that is then used for all subsequent experiments. This ensures accurate buffer subtraction.
13. The SAXS intensity profile must be taken as a difference in signals between the sample and the corresponding buffer, which may lead to significant systematic errors if the signal subtraction is inadequate.
14. Concentrating samples to high concentration and then diluting the samples for SAXS measurements often results in unwanted aggregation and thus is avoided.
15. We note that in addition to the analysis of SAXS data, it is essential to assess molecular flexibility using additional biochemical or biophysical methods, such as limited proteolysis or hydrogen/deuterium exchange.
16. The ensemble of structures generated by the simulations does not need to be perfect to be useful. The molecular simulations serve merely to produce an initial pool of meaningful candidate conformations. However, part of the success of the EROS method is the use of appropriate, physics-based simulations to sample the relevant conformations of multi-domain proteins and multi-protein complexes. In fact, the transferable energy function used in the EROS simulations has been shown to correctly predict structures and binding affinities of a number of protein-protein complexes [28]. Also, in a recent study on cellulosomal proteins [39], the ensemble of simulation

structures has been found to fit experimental SAXS data without any refinement.

17. In principle, the structural ensemble can be fitted either to raw experimental data or to commensurate quantities such as SAXS-derived pair-distance distribution function or DEER-derived inter-label distance distribution. However, to avoid introducing any regularization-dependent artifacts into the ensemble refinement, the simulation structures are fitted directly to experimental data in the framework of the EROS method.

Acknowledgments

The authors thank all the members of the Page and Peti laboratory. This work was supported by NIH grant R01GM098482 to RP; NIH R01GM100910 and American Diabetes Association Pathway to the Cure 1-14-ACN-31 to WP. EB was supported by the Czech Science Foundation grant number 17-05200S, by the project InterBioMed LO1302 from the Ministry of Education of the Czech Republic and by the Academy of Sciences of the Czech Republic (RVO: 61388963). BR was supported by the National Science Centre, Poland, grant number 2016/21/B/NZ1/00006, and by the European Framework Programme VII NMP grant 604530-2 (CellulosomePlus) and co-financed by the Polish Ministry of Science and Higher Education from the resources granted for the years 2014-2017 in support of scientific projects.

References

1. Blanchet CE, Svergun DI (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu Rev Phys Chem* 64:37–54. doi:[10.1146/annurev-physchem-040412-110132](https://doi.org/10.1146/annurev-physchem-040412-110132)
2. Graewert MA, Svergun DI (2013) Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr Opin Struct Biol* 23 (5):748–754. doi:[10.1016/j.sbi.2013.06.007](https://doi.org/10.1016/j.sbi.2013.06.007)
3. Bernado P, Perez Y, Svergun DI, Pons M (2008) Structural characterization of the active and inactive states of Src kinase in solution by small-angle X-ray scattering. *J Mol Biol* 376 (2):492–505. doi:[10.1016/j.jmb.2007.11.066](https://doi.org/10.1016/j.jmb.2007.11.066)
4. Pelikan M, Hura GL, Hammel M (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 28(2):174–189
5. Alber F, Forster F, Korkin D, Topf M, Sali A (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 77:443–477. doi:[10.1146/annurev.biochem.77.060407.135530](https://doi.org/10.1146/annurev.biochem.77.060407.135530)
6. Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19 (1):109–116. doi:[10.1016/j.str.2010.10.006](https://doi.org/10.1016/j.str.2010.10.006)
7. Yang S, Blachowicz L, Makowski L, Roux B (2010) Multidomain assembled states of Hck tyrosine kinase in solution. *Proc Natl Acad Sci USA* 107(36):15757–15762. doi: [10.1073/pnas.1004569107](https://doi.org/10.1073/pnas.1004569107)
8. Peti W, Page R (2016) NMR spectroscopy to study MAP kinase binding to MAP kinase phosphatases. *Meth Mol Biol* 1447:181–196. doi: [10.1007/978-1-4939-3746-2_11](https://doi.org/10.1007/978-1-4939-3746-2_11)
9. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16(1):18–29. doi:[10.1038/nrm3920](https://doi.org/10.1038/nrm3920)

10. Peti W, Page R (2007) Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr Purif* 51 (1):1–10. doi:[10.1016/j.pep.2006.06.024](https://doi.org/10.1016/j.pep.2006.06.024)
11. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOLO - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28:768–773
12. Schneidman-Duhovny D, Hammel M, Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38(Web Server issue):W540–W544. doi: [10.1093/nar/gkq461](https://doi.org/10.1093/nar/gkq461)
13. Grishaev A, Guo L, Irving T, Bax A (2010) Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J Am Chem Soc* 132(44):15484–15486. doi:[10.1021/ja106173n](https://doi.org/10.1021/ja106173n)
14. Poitevin F, Orland H, Doniach S, Koehl P, Delarue M (2011) AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Res* 39(Web Server issue):W184–W189. doi:[10.1093/nar/gkr430](https://doi.org/10.1093/nar/gkr430)
15. Liu HG, Hexemer A, Zwart PH (2012) The small angle scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J Appl Cryst* 45:587–593
16. Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* 589(19 Pt A):2570–2577. doi:[10.1016/j.febslet.2015.08.027](https://doi.org/10.1016/j.febslet.2015.08.027)
17. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HD, Konarev PV, Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Cryst* 45(Pt 2):342–350. doi:[10.1107/S0021889812007662](https://doi.org/10.1107/S0021889812007662)
18. Ravikumar KM, Huang W, Yang S (2013) Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes. *J Chem Phys* 138 (2):024112. doi:[10.1063/1.4774148](https://doi.org/10.1063/1.4774148)
19. Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95 (8):559–571. doi:[10.1002/bip.21638](https://doi.org/10.1002/bip.21638)
20. Boura E, Rozycki B, Herrick DZ, Chung HS, Vecer J, Eaton WA, Cafiso DS, Hummer G, Hurley JH (2011) Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc Natl Acad Sci USA* 108(23):9437–9442. doi: [10.1073/pnas.1101763108](https://doi.org/10.1073/pnas.1101763108)
21. Francis DM, Rozycki B, Koveal D, Hummer G, Page R, Peti W (2011) Structural basis of p38alpha regulation by hematopoietic tyrosine phosphatase. *Nat Chem Biol* 7(12):916–924. doi:[10.1038/nchembio.707](https://doi.org/10.1038/nchembio.707)
22. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80 (6):2946–2953. doi:[10.1016/S0006-3495\(01\)76260-1](https://doi.org/10.1016/S0006-3495(01)76260-1)
23. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst* 42 (Pt 2):342–346. doi:[10.1107/S0021889809000338](https://doi.org/10.1107/S0021889809000338)
24. Rozycki B, Boura E (2014) Large, dynamic, multi-protein complexes: a challenge for structural biology. *J Phys Condens Matter* 26 (46):463103. doi:[10.1088/0953-8984/26/46/463103](https://doi.org/10.1088/0953-8984/26/46/463103)
25. Boura E, Rozycki B, Chung HS, Herrick DZ, Canagarajah B, Cafiso DS, Eaton WA, Hummer G, Hurley JH (2012) Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission. *Structure* 20(5):874–886. doi:[10.1016/j.str.2012.03.008](https://doi.org/10.1016/j.str.2012.03.008)
26. Francis DM, Rozycki B, Tortajada A, Hummer G, Peti W, Page R (2011) Resting and active states of the ERK2:HePTP complex. *J Am Chem Soc* 133(43):17138–17141. doi:[10.1021/ja2075136](https://doi.org/10.1021/ja2075136)
27. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9 (9):1753–1773. doi:[10.1110/ps.9.9.1753](https://doi.org/10.1110/ps.9.9.1753)
28. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J Mol Biol* 375(5):1416–1433. doi:[10.1016/j.jmb.2007.11.063](https://doi.org/10.1016/j.jmb.2007.11.063)
29. Kenzaki H, Koga N, Hori N, Kanada R, Li W, Okazaki K, Yao XQ, Takada S (2011) CafeMol: a coarse-grained biomolecular simulator for simulating proteins at work. *J Chem Theory Comput* 7(6):1979–1989. doi:[10.1021/ct2001045](https://doi.org/10.1021/ct2001045)
30. Liwo A, Baranowski M, Czaplowski C, Golas E, He Y, Jagiela D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, Niazdvedtski A, Oldziej S, Scheraga HA, Sieradzian AK, Slusarz R, Wirecki T, Yin Y, Zaborowski B (2014) A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J Mol Model* 20(8):2306. doi:[10.1007/s00894-014-2306-5](https://doi.org/10.1007/s00894-014-2306-5)
31. Dannenhöffer-Lafage T, White AD, Voth GA (2016) A direct method for incorporating

- experimental data into multiscale coarse-grained models. *J Chem Theory Comput* 12 (5):2144–2153. doi:[10.1021/acs.jctc.6b00043](https://doi.org/10.1021/acs.jctc.6b00043)
32. Yang S, Park S, Makowski L, Roux B (2009) A rapid coarse residue-based computational method for x-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys J* 96(11):4449–4463. doi:[10.1016/j.bpj.2009.03.036](https://doi.org/10.1016/j.bpj.2009.03.036)
33. Polyhach Y, Bordignon E, Jeschke G (2011) Rotamer libraries of spin labelled cysteines for protein studies. *Phys Chem Chem Phys* 13 (6):2356–2366. doi:[10.1039/c0cp01865a](https://doi.org/10.1039/c0cp01865a)
34. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A, Eaton WA (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc Natl Acad Sci USA* 104 (48):18964–18969. doi: [10.1073/pnas.0709567104](https://doi.org/10.1073/pnas.0709567104)
35. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci USA* 104(5):1528–1533. doi: [10.1073/pnas.0607097104](https://doi.org/10.1073/pnas.0607097104)
36. Hartigan JA, Wong MA (1979) A k-means clustering algorithm. *Appl Stat* 28:100–108
37. Heyer LJ, Kruglyak S, Yooseph S (1999) Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* 9 (11):1106–1115
38. Leung HT, Bignucolo O, Aregger R, Dames SA, Mazur A, Berneche S, Grzesiek S (2016) A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. *J Chem Theory Comput* 12(1):383–394. doi:[10.1021/acs.jctc.5b00759](https://doi.org/10.1021/acs.jctc.5b00759)
39. Rozycki B, Cieplak M, Czjzek M (2015) Large conformational fluctuations of the multi-domain xylanase Z of *Clostridium thermocellum*. *J Struct Biol* 191(1):68–75. doi:[10.1016/j.jsb.2015.05.004](https://doi.org/10.1016/j.jsb.2015.05.004)

Implementation of the NMR CHEmical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery

Stephen Boulton, Rajeevan Selvaratnam, Rashik Ahmed,
and Giuseppe Melacini

Abstract

Mapping allosteric sites is emerging as one of the central challenges in physiology, pathology, and pharmacology. Nuclear Magnetic Resonance (NMR) spectroscopy is ideally suited to map allosteric sites, given its ability to sense at atomic resolution the dynamics underlying allostery. Here, we focus specifically on the NMR CHEmical Shift Covariance Analysis (CHESCA), in which allosteric systems are interrogated through a targeted library of perturbations (e.g., mutations and/or analogs of the allosteric effector ligand). The atomic resolution readout for the response to such perturbation library is provided by NMR chemical shifts. These are then subject to statistical correlation and covariance analyses resulting in clusters of allosterically coupled residues that exhibit concerted responses to the common set of perturbations. This chapter provides a description of how each step in the CHESCA is implemented, starting from the selection of the perturbation library and ending with an overview of different clustering options.

Key words Allostery, Allosteric control, Agonism, Antagonism, Dynamics, Inhibitor, NMR, CHESCA, Structure

1 Introduction

Allostery pertains to the communication between discrete and often distant sites within a molecular system [1–5]. In proteins, allostery commonly serves as a regulatory mechanism, whereby a perturbation at one site produces structural or dynamic alterations that result in a functional change at another locus. Being able to map “signaling pathways” between allosteric and orthosteric sites in proteins is essential for understanding physiology and pathology, and also opens up new opportunities in preclinical drug discovery [1, 2, 6, 7]. By elucidating which protein sites control which functions (e.g., binding vs. activation), it becomes possible to selectively target them for the purpose of eliciting specific

pharmacological responses, e.g., inhibition without significantly reducing binding affinity. In addition, allostery is critical to shed light into the mechanism of disease-causing perturbations, such as genetic mutations, which often disrupt obligatory allosteric pathways within proteins [1, 2].

A simple but effective approach to envision allosteric pathways is through a network of interconnected residues that respond in a concerted manner to the allosteric modulators. Such networks typically include sites of structural and/or dynamical divergence among the conformational states that define the auto-inhibitory equilibria. In fact, one of the simplest rationalizations of allostery is in terms of a four-state thermodynamic cycle arising from the coupling of two-state conformational and binding equilibria (Fig. 1). This model of conformational selection, which was initially proposed by Monod, Wyman, and Changeaux over 50 years ago, describes a protein that is in equilibrium between two states, tensed (T) and relaxed (R), for which an allosteric ligand will preferentially bind to one of these two states and shift the equilibrium in its favor (Fig. 1) [8]. Depending on whether the allosteric modulator displays a greater selectivity toward binding the inactive or active states determines whether it will act as an antagonist or agonist (Fig. 1). In a perfect two-state allosteric system, each residue should sense the same conformational equilibrium and respond to the allosteric

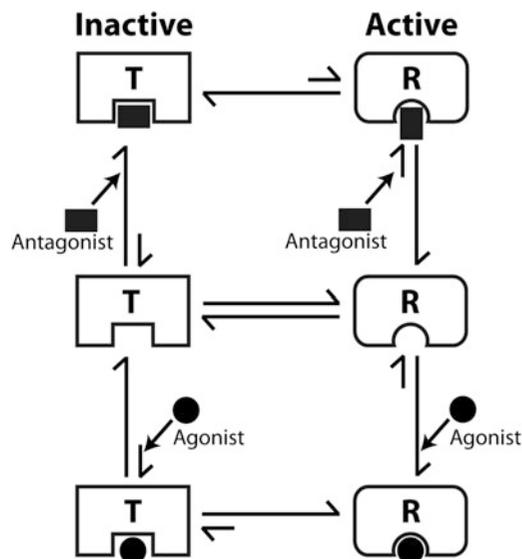


Fig. 1 Thermodynamic model of ligand-modulated conformational equilibria. Conformational selection model in which the equilibrium between two states tensed (T) and relaxed (R) is coupled with the equilibrium for ligand binding. When a ligand displays preferential binding to the inactive (T) state, it is an antagonist (*rectangle*), and when it binds more selectively to the active (R) state, it is an agonist (*circle*)

perturbations in a concerted manner. However, in reality, deviations from perfect two-state concerted conformational transitions are often observed due to the presence of intermediates resulting from weak allosteric couplings and/or from simultaneous sensing of binding and allosteric equilibria. Hence, to identify purely allosteric sites, it becomes essential to identify which residues within a protein respond in a correlated manner to allosteric modulators. This is the aim of the NMR CHEMical Shift Covariance Analysis (CHESCA) [2, 9–14].

Chemical shifts serve as excellent sensors of local chemical and structural environments and thus provide valuable information about conformational changes and interactions with ligands or other proteins [2, 15]. In particular, chemical shifts are especially useful for measuring populations of states in fast chemical exchange on the NMR chemical shift time scale, where fast exchange is defined as an equilibrium between two states with an exchange rate much larger than the chemical shift (frequency) difference between those two states. For residues in a fast exchanging two-state equilibrium, the chemical shift (δ) of a perturbed state (x) will appear as population weighted average (p) of the chemical shifts of the two pure states (R and T):

$$\delta_x = p_T \delta_T + p_R \delta_R \quad (1)$$

Equation (1) applies to any residue-specific chemical shift that is resolved in the available NMR spectra. If different residues sense the same conformational equilibrium, their average chemical shifts reflect similar populations for the T and R states utilized in this equation [2, 15]. The CHESCA analysis relies on this trait to uncover the specific residues that respond in a concerted manner to a small targeted perturbation library of ligands and/or mutations selected to modulate allosteric equilibria, as illustrated in Fig. 2. To achieve this goal, CHESCA utilizes clustering algorithms to group subsets of residues that show similar responses and assign to such clusters a function based on the functional profile of the selected perturbations obtained through biochemical assays (Fig. 2). In conclusion, CHESCA relies on a library of small molecules or local mutations to “interrogate” allosteric systems and on NMR to provide a high-resolution readout of such chemical perturbations to a biological system. In this respect, the CHESCA method reflects a genuine chemical-biologist’s approach to allostery.

2 Methods

The general overview of the CHESCA protocol is illustrated in Fig. 2. The CHESCA can be divided into three main blocks: The preparation of the perturbation library, the detection of a response

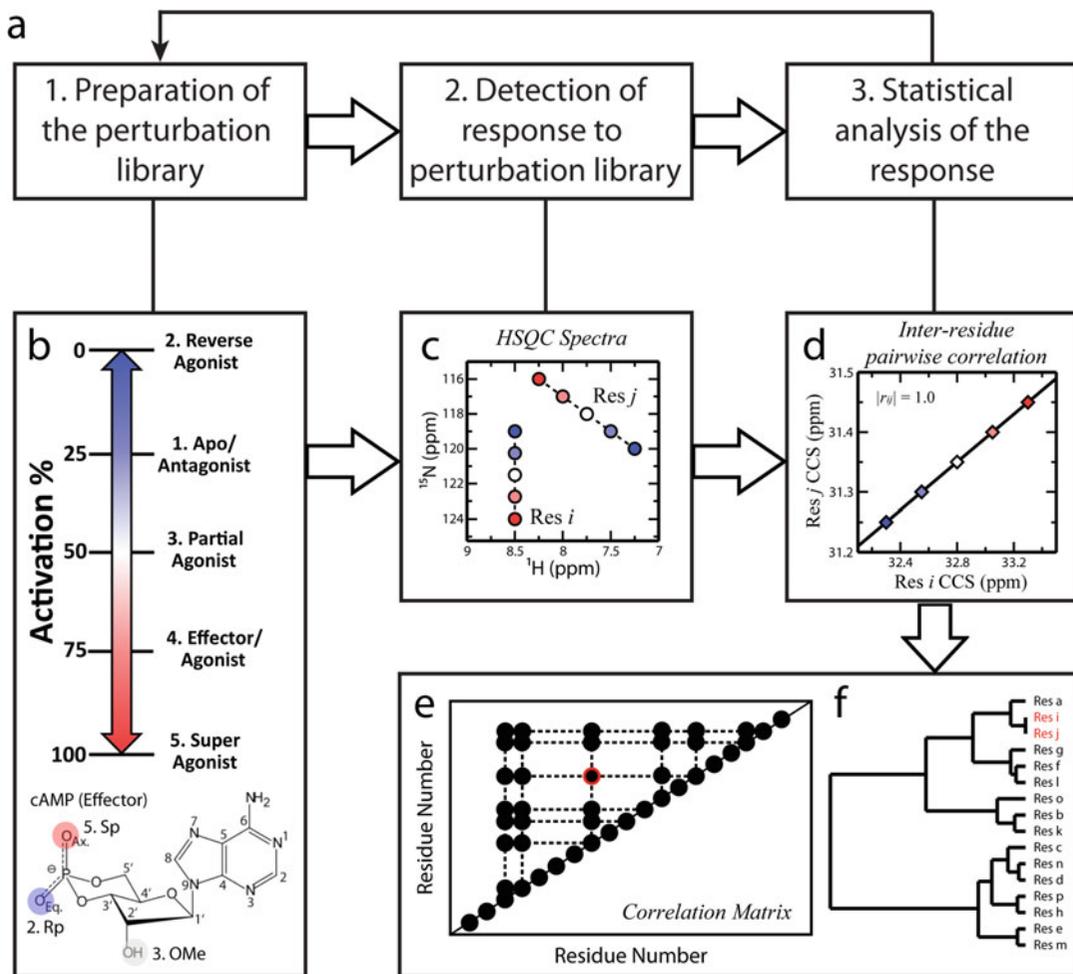


Fig. 2 (a) General scheme for the CHESCA analysis. (b) The perturbation library should span diverse activation stages. For instance, in the CHESCA analysis of protein kinase A by Akimoto et al., a perturbation library was comprised of cAMP (effector), Rp-cAMPS (reverse agonist), Sp-cAMPS (agonist), 2'-O-methyl-cAMP (partial agonist), and the apo state [10]. (c) The response to the perturbation library is then detected by HSQC spectroscopy. Finally, the compounded chemical shifts (CCS; see Eq. (2)) for each state are plotted in residue vs. residue “inter-residue pairwise correlation” (IPC) plots (d). Correlations that meet a selected cutoff (e.g., the IPC shown in panel (d)) are compiled into the correlation matrix (e) (e.g., red circle) and categorized by agglomerative clustering (f) (red residues)

to the perturbation library, and the statistical analysis of said response (Fig. 2a), including the identification of inter-residue pairwise chemical-shift correlations (IPCs) (Fig. 2d). The statistical analysis may provide insight into the mechanistic model of activation/binding, which in turn may inform the design of a 2nd generation library of CHESCA perturbations to be utilized in a 2nd CHESCA implementation. Further iterations of the CHESCA analysis may provide more refined information about the allosteric system under investigation.

2.1 Selection of Perturbation Library

Library selection is the most critical step for an informative CHESCA implementation. An ideal starting point is to select a series of perturbations spanning a wide range of activities (e.g., Fig. 2b). However, it is also important to keep the perturbations spatially constrained within a small region in the protein. This is because each perturbation introduces new substituents or interactions, which can change the local chemical environment of adjacent residues and in some cases, result in structural distortions that extend even further. These processes, which we refer to as nearest neighbor effects (NNEs), will cause de-correlations of nearby residues in the CHESCA analysis. Hence, if the placement of perturbations is widespread throughout the protein, the NNE-induced de-correlations will possibly obscure functionally relevant allosteric pathways. To circumvent this problem, it is recommended to use as CHESCA library analogs derived from a single scaffold that bind to a single site (Fig. 2b), or to use mutations constrained to a well-confined region of the 3D structure of the protein.

The suitability of a library to serve as a perturbation set for CHESCA can be easily tested through the CHEMical Shift Projection Analyses (CHESPA) analyses, as explained in references [9, 16, 17], and/or by checking the inter-residue pairwise correlations (IPCs) (Fig. 2d) obtained in the initial round of analyses. If visual inspection of the IPCs reveals any resemblance to the outlier scenarios described by the Anscombe's quartet [18] (Fig. 3), library re-design is recommended. The Anscombe's quartet describes four datasets with identical correlation coefficients, but very different appearances when graphed [18]. In particular, the scenarios described by Fig. 3b are fairly common as they develop when there is an outlier of only a single state. For instance, Fig. 3b could arise when a single state causes large structural distortions or stabilizes a new state that deviates from the two-state equilibrium, while the scenario shown in Fig. 3d can occur when the perturbation library is polarized between the opposite ends of the equilibrium (e.g., one inactive and several fully active states). However, it is also possible that both Fig. 3b, d patterns apply to only a small subset of residues because of NNEs caused by the single outlier state. If either of these scenarios comprises the majority of IPCs throughout the CHESCA analysis, it is recommended to redesign the perturbation library, but if they affect only a small subset of residues, it should not interfere with the analysis and library re-design is not strictly essential. The agglomerative clustering (AC) and singular value decomposition (SVD) analyses, which will be described later, are two ways of determining whether the majority of residues meet the criteria of Anscombe's quartet outliers. In conclusion, the ideal perturbation library for CHESCA meets three criteria: (a) spans a broad functional activity range; (b) is spatially confined to a single region of the perturbed structure; (c) does not give rise to systemic Anscombe's Quartet cases.

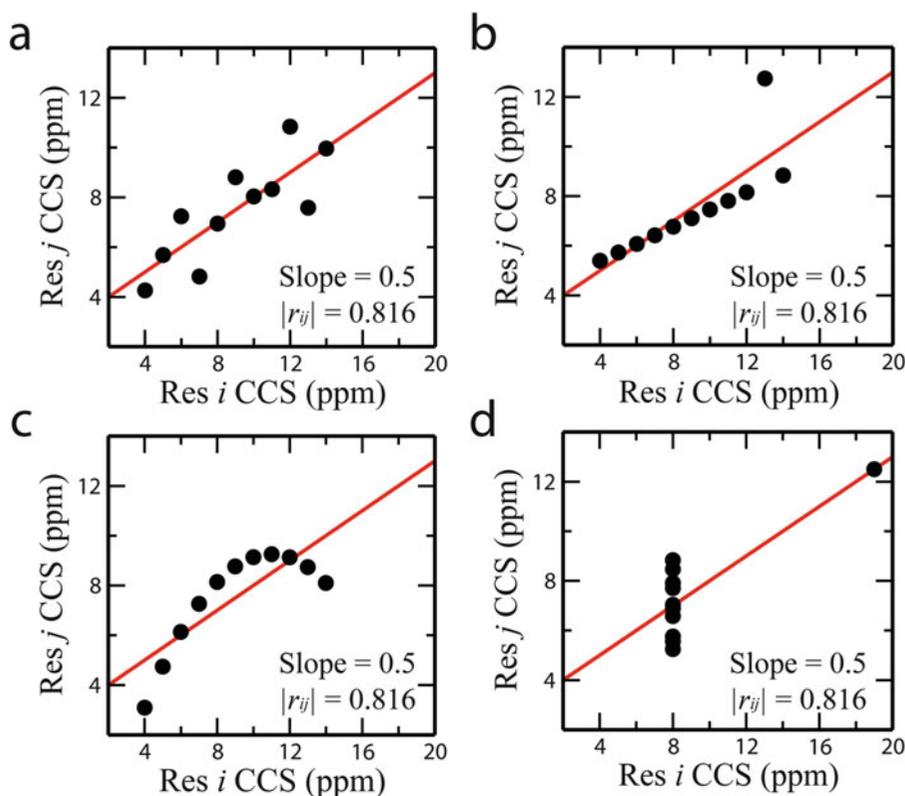


Fig. 3 Representative examples of IPCs that fit the Anscombe's quartet. The Anscombe's quartet describes four different datasets that produce identical statistics but different point distributions, as revealed by visual inspection of the correlations. The datasets for these example plots are from Chatterjee et al. [18]

When a suitable library of perturbations has been selected, CHESCA can produce a residue-resolution map illustrating how a perturbation at an allosteric site propagates to the orthosteric site and elicits a specific physiological response. This can serve as a starting point for further analyses. For instance, comparative CHESCA analyses are useful to gauge to what extent disease-causing mutations disrupt allosteric modulation and pathways. In such cases, the CHESCA can be performed for both WT and the disease-causing variant using the same library of perturbations. A comparative analysis between WT and mutant CHESCA can then reveal how the communication between the allosteric and orthosteric sites is being disrupted by the mutation (Fig. 4a). This can be seen by either the loss of pairwise correlations between the residues that are essential for signal transduction, and/or the gain of correlations between other residues, depending on which residues experience a local environment that conforms to a two-state equilibrium (Fig. 4a). Similarly, this type of comparative CHESCA analysis can be applied to Apo vs. Holo states using mutations as perturbation library. For example, Axe et al. compared the CHESCA of the apo-resting and holo-working states for the alpha-subunit of

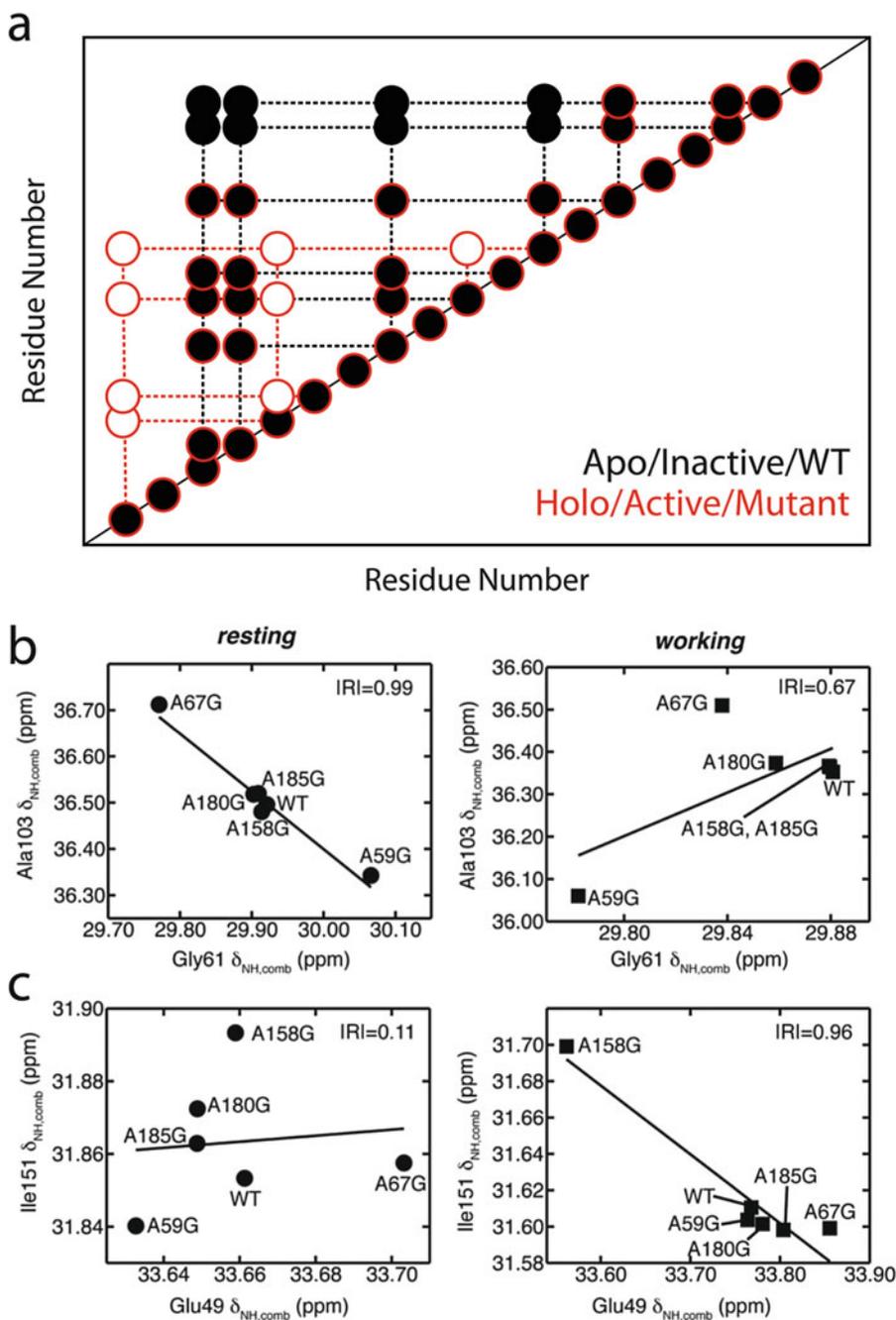


Fig. 4 Comparative analysis of CHESCA correlations. **(a)** Representative correlation matrices for which an Apo/inactive/WT protein (*black*) is compared with its Holo/active/mutant counterpart, respectively (*red*). Differences will appear as either a loss or gain of correlations for selected residue pairs, which in turn illustrate how allosteric networks are perturbed. For example, Axe et al. compared CHESCA analyses between a resting and working form of the alpha-subunit of tryptophan synthase and discovered both loss **(b)** and gain **(c)** of correlations within networks of residues [14, 19, 20]. In particular, Axe et al. discovered that selected catalytic residues exhibited more correlations in the active state. Panels B and C were adapted from Axe et al. [14] with permission from the American Chemical Society [14]

tryptophan synthase and found differing correlations (Fig. 4b, c) [14, 19, 20]. With knowledge gained from the comparative CHESCA analyses, mutations were designed to perturb surface residues belonging to the working clusters located more than 15 Å away from the active site. These mutations resulted in significant reductions to the catalytic efficiency of the enzyme [14]. This example illustrates how CHESCA dissects the functional components of proteins, revealing which regions are essential for activation and which regions confer ligand affinity and selectivity. Such a beneficial property makes CHESCA ideal for rational drug design.

2.2 Data Acquisition

Mapping the response to the perturbation library requires NMR observables that can be measured at atomic resolution with a high degree of accuracy and precision, as for example ^1H and ^{15}N chemical shifts from HSQC spectra. However, it is also possible to include ^{13}C chemical shifts as well. In either case, chemical shift to be used for CHESCA should be referenced with respect to a reliable reference compound, as variations in chemical shift referencing from one perturbation to another may generate artifactual correlations. For example, we used ^{15}N -Acetyl-Glycine. In addition, if a ligand library is utilized as perturbation, the original CHESCA implementation requires that the protein be brought to full saturation (*see Note 1*). Peak positions should be measured with sufficient digital resolution and through careful peak picking. For example, we utilized Gaussian line fitting of spectra processed with resolution-enhancing window functions.

3 Data Analysis

While it is possible to implement most of the computations for the CHESCA analysis in programs such as Microsoft Excel or Matlab, we recommend GNU Octave, which is freely available from <https://www.gnu.org/software/octave/download.html>. In addition, many of the CHESCA matrices require pre- or postprocessing, which can also be executed in any spreadsheet program or with GAWK scripts (available upon request—*see Note 2*). For the sake of simplicity, here we will focus on the correct format of each dataset that is required by the programs utilized for CHESCA processing. Agglomerative clustering is performed with Cluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/>) [21] and visualized as dendrograms using Java TreeView (<http://jtreeview.sourceforge.net/>) [22].

3.1 CHESCA Statistical Analysis: Generation of the Correlation Matrix

1. After assigning the spectra for each state from the perturbation library, transfer the peak lists to a spreadsheet like Excel and organize in the same format as Table 1. Save the peak lists for each state as separate tab delimited text files. Do not include the first row of title headings (bold font).

Table 1
Example of the format for the chemical shift peak list for one state in the perturbation library

Residue #	¹⁵ N ppm	¹ H ppm	¹⁵ N Error	¹ H Error
111	121.24	8.206	0.03	0.006
112	132.52	7.855	0.04	0.003
113	119.44	9.684	0.04	0.007

Table 2
Example of the format for the idm.txt file

Residue #	State 1	State 2	State n
111	32.156	32.165	32.188
112	34.223	34.246	34.218
113	31.256	31.226	31.208

- In a separate spreadsheet Compute the Compounded Chemical Shift (CCS) for each residue using the equation

$$\text{CCS} = \alpha * \delta_{\text{N}} + \delta_{\text{H}} \quad (2)$$

where δ_{N} and δ_{H} represent the ¹⁵N and ¹H chemical shifts, respectively, and α represents the scaling factor for the ¹⁵N chemical shifts (*see Note 3*). Organize the CCS in a matrix similar to the one shown in Table 2. Save the file as idm.txt.

- Remove the first row containing the title headings and the first column containing the residue numbers and save this file as dm.txt.
- Open Octave. In the top left window, navigate to the directory containing your files and then in the command prompt in the right window type:

```
load dm.txt
```
- To generate the correlation matrix of the transpose of the “residue x perturbation” chemical shift matrix, execute the command:

```
cdmt = corr(dm.')
```
- Save the output by typing

```
save dm.octave
```
- The dm.octave file contains the symmetric correlation matrix in which the rows and columns represent the residue numbers

and the values in the matrix represent the Pearson's correlation coefficients between the two residues. The diagonal of the matrix is always one and it depicts the autocorrelations. The dm.octave file also contains the dm matrix that was used to generate the correlation matrix. Remove the dm data from the dm.octave file as well as the text at the top of the file for the "cdmt" matrix. Save this as a new file called

rcm.txt.

8. Identify all the residue pairs in the correlation matrix with an absolute Pearson's correlation coefficient greater than 0.98 (*see* **Note 4**) and map them in a 2D matrix of residue number vs residue number as seen in Fig. 2c.
9. To graph IPCs (Fig. 2d), plot the CCS values for two different residues (obtained from the idm.txt file) as x and y coordinates. Error bars can be generated from the error propagation of ^{15}N and ^1H chemical shift standard deviations, as described in *see* **Note 1**.

3.2 CHESCA Statistics: Agglomerative Clustering (AC) of Residue Networks

In the protocol outlined below, we will focus on how to implement complete-linkage AC, which has been proposed as one of the methods reviewed so far that results in the most balanced compromise between false positives and negatives [9]. The other options, which include single linkage AC and independent clustering of ^1H and ^{15}N chemical shifts, may be preferable to minimize either the number of false negatives or false positives, respectively. The advantages and disadvantages of different clustering algorithms used in the CHESCA analysis are discussed in detail by Boulton et al. [9]. Boulton et al. also provide detailed guidelines to determine which clustering algorithm is best for your system as well as how to troubleshoot potential artifacts that arise from AC [9].

1. Open the idm.txt file in Cluster 3.0.
2. Go to the hierarchical tab and under both the "Genes" and "Arrays" boxes, check the cluster box and choose "Absolute Correlation (centered)" from the dropdown menu. Then click the Complete linkage button. This will generate two files: idm.cdt and idm.gtr. The .cdt file is for viewing the dendrogram in Java TreeView (*see* **step 3** below). The .gtr file is used for isolating the clusters containing residues that are highly correlated (*see* **step 4**).
3. Open idm.cdt in Java TreeView. There will be four main windows. The window on the left shows the complete dendrogram of correlations (similar to Fig. 2f). Clicking on a specific branch will provide the correlation coefficient for that cluster in the top window. The window on the far right shows all the residues that are within the selected branch of the dendrogram, while the second panel from the left provides a heat map of the

correlation coefficients. The panel just above it provides a dendrogram showing how the states for the currently selected branch are grouped. You can select and export specific branches to either .png or .ps files under the export tab.

4. To extract the residue clusters with similar responses to the perturbation library, open the `idm.gtr` file in a text editor such as notepad. The file will contain four columns: the last column will contain the correlation coefficient for that particular branch in the dendrogram, while the first three columns will contain a collection of “NODE” and “GENE” numbers. The NODEs correspond to branches in the dendrogram and the GENEs correspond to residues (although the GENE numbers do not necessarily reflect the actual residue number, but the position in the `idm.txt` matrix). Identify any clusters with correlation coefficients above the designated cutoff (e.g., 0.98) and three or more residues to generate dendrograms similar to Fig. 2f.
5. The functional significance of the residue clusters is determined according to how the states from the perturbation library are grouped within a specific cluster. For instance, if the largest state separations are between the inactive and active states, the function of the cluster is related to activation. If the largest separation is between the apo and holo states, the cluster could be more relevant for binding. To acquire the state dendrograms for each cluster (similarly to Fig. 3d, e in ref. 11), return to Java TreeView and scan through the branches of the residue dendrogram until the clusters from **step 4** are identified (*see Note 5*). It is recommended to export each cluster to .png or .ps once they have been discovered (*see Note 6*).
6. Plot the clusters onto the structure of the protein to see if they form a continuous surface (*see Note 7*).

3.3 CHESCA Statistics: Singular Value Decomposition (SVD)

The SVD analysis complements the correlation and clustering analyses outlined above. In addition, the loadings from the SVD analysis provide a practical means of measuring state populations for a large number of different perturbations (assuming there are at least two reference states with known populations) [12].

1. Factorize the `dm.txt` matrix by compiling relative CCS values with respect to either a reference state or through row-mean centering (i.e., Calculate the CCS difference between each state and the reference state or calculate the average CCS of each row and use that as the reference state) (*see Note 8*). This type of preprocessing can profoundly change the outcome of the SVD analysis and therefore it is advisable to repeat the SVD protocol multiple times using different types of preprocessing.

2. Column mean centers the matrix by calculating the average CCS value for each column and subtracting it from each value in that column. Save the result as
ndm.txt
3. In Octave, load the ndm.txt file by typing
load ndm.txt
and in the command line type
[u,s,v]=svd(ndm)
4. Then create the score matrix by typing
score=u*s
5. Save the results by typing
save svd.octave
and in a text editor extract the numerical values of the s, v, and score matrices and save them in separate file names: s.txt, v.txt, and score.txt, respectively.
6. Open s.txt in a spreadsheet such as Excel. Square each value and then divide by the sum of all squared values. The result is the percent variance for each of the principal components (PCs). If the sum of the first two PC variances is less than 90%, we recommend restarting with a different reference state (*see* also Fig. S2 of ref. 11 for further details and *see* **Note 8**).
7. The v.txt file contains the values for the loadings. Each column corresponds to a particular PC, while the rows correspond to the states from the perturbation library. Take the first two columns as coordinates for the PC1 and PC2 components, respectively.
8. The score.txt file contains the scores for each residue with respect to each PC. Similar to **step 7**, take the first two columns as coordinates for the PC1 and PC2 components, respectively.
9. Plot the loading and scores from **steps 7** and **8** to create the same graph as in Fig. 4 of Selvaratnam et al. [11].

4 Notes

1. If using ligands in the CHESCA perturbation library, we recommend performing titrations to ensure that saturation is reached. In addition, we suggest that you use at least three spectra acquired at saturation to calculate average chemical shifts and the respective standard deviations, which can be used as errors in the analysis of the CHESCA IPCs.
2. These scripts require a GNU compiler, which is commonly included in Linux systems. We recommend using Cygwin (<https://www.cygwin.com>), a collection of GNU and open-

source tools, which provides a Linux-like environment on Windows computers. In addition, Octave is an optional package that can be downloaded with Cygwin.

3. This equation can be modified to include ^{13}C chemical shifts instead of ^{15}N or to include both if desired. A scaling factor (α) is required for ^{13}C and ^{15}N chemical shifts since the spectral window for these nuclei is in a much higher chemical shift range and the chemical shift dispersion is much greater. Without a scaling factor for these nuclei, they would significantly outweigh the contribution from ^1H . Possible values for the scaling factor are 0.2 for ^{15}N and 0.3 for ^{13}C , but for further discussion on this topic we suggest referring to the review by Williamson [15].
4. The correlation coefficient cut-off value is dependent on the system under investigation and the error margin on the measured chemical shifts. In some cases, a strict limit of 0.98 may not be very informative. We recommend creating a heat map instead in which multiple correlation coefficients are set as cutoffs and mapped as colors in the final correlation matrix to highlight both strong and weak correlations.
5. An alternative approach to this step is to create files in the format of `idm.txt`, containing only the residues found within one particular cluster. These files can be clustered using the same protocol as in Subheading 2, step 3 and visualized in Java TreeView.
6. If using complete-linkage AC, there may be several clusters with similar state groupings (i.e., functions). This is due to the high stringency of complete-linkage AC, which can sometimes fragment clusters with similar functions. The paper of Boulton et al. [9] provides guidelines for combining “fragment” clusters with similar state groupings.
7. A comparative analysis between the protein’s structural changes and the CHESCA clusters can be very informative. Most often, the CHESCA clusters will be in agreement with local RMSD maxima [9, 11]. However, it is also possible to identify CHESCA correlations in regions devoid of any conformational change or, vice versa, to identify large conformational changes in regions without any CHESCA correlation. For instance, Selvaratnam et al. uncovered many correlations involving an N-terminal region of EPAC1 cAMP-binding domain (CBD) that displayed no significant conformational changes as seen by structural overlays and local RMSDs of the inactive and active crystal structures [11]. However, this region displayed enhanced μs -ms dynamics upon activation that were proposed to drive the release of an auto-inhibitory salt bridge between the catalytic and regulatory domains [11]. In another

example, structural changes were observed in a region of the bile acid-binding protein (BABP), which yielded no correlations through CHESCA, but inspection of X-ray crystal structures of BABP with different bound ligands revealed this region conferred selectivity toward different bile acids [23].

8. An ideal reference state would be the one that separates the two predominant functional equilibria (e.g., binding vs. activation). For instance, in the context of the thermodynamic cycle shown in Fig. 1a, an antagonist or reverse agonist might bind similarly to the effector, agonists, and partial agonists, but result in an inactive state more similar to the apo state (i.e., the antagonist binds without activation). Choosing a suitable reference state will help make the directions of maximum variance orthogonal, as required by SVD. If an antagonist or reverse agonist state is not available, we suggest trying multiple reference states or the row mean centering approach.

Acknowledgments

We thank Amir Bashiri (McMaster U.), Dr. M. Akimoto (Keio U.), Professor G. Veglia (U. Minnesota), and L.E. Kay (U. Toronto) for helpful discussions. This study received funding from Canadian Institutes of Health Research (Grant MOP-68897) to G.M. and Natural Sciences and Engineering Research Council of Canada (Grant RGPIN-2014-04514) to G.M.

References

1. Nussinov R, Tsai C (2013) Allostery in disease and in drug discovery. *Cell* 153(2):293–305
2. Boulton S, Melacini G (2016) Advances in NMR methods to map allosteric sites: from models to translation. *Chem Rev* 116(11):6267–6304
3. Hilser VJ, Wrabl JO, Motlagh HN (2012) Structural and energetic basis of allostery. *Annu Rev Biophys* 41:585–609
4. Kuriyan J, Eisenberg D (2007) The origin of protein interactions and allostery in colocalization. *Nature* 450(7172):983–990
5. Smock RG, Gierasch LM (2009) Sending signals dynamically. *Science* 324(5924):198–203
6. Nussinov R, Tsai C (2012) The different ways through which specificity works in orthosteric and allosteric drugs. *Curr Pharm Des* 18(9):1311–1316
7. Wenthur CJ, Gentry PR, Mathews TP et al (2014) Drugs for allosteric sites on receptors. *Annu Rev Pharmacol Toxicol* 54:165–184
8. Monod J, Wyman J, Changeux J (1965) On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12(1):88–118
9. Boulton S, Akimoto M, Selvaratnam R et al (2014) A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Sci Rep* 4:7306
10. Akimoto M, Selvaratnam R, McNicholl ET et al (2013) Signaling through dynamic linkers as revealed by PKA. *Proc Natl Acad Sci U S A* 110(35):14231–14236
11. Selvaratnam R, Chowdhury S, VanSchouwen B et al (2011) Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc Natl Acad Sci U S A* 108(15):6133–6138
12. Selvaratnam R, Mazhab-Jafari MT, Das R et al (2012) The auto-inhibitory role of the EPAC hinge helix as mapped by NMR. *PLoS One* 7(11):e48707
13. Cembran A, Kim J, Gao J et al (2014) NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts

- upon ligand binding. *Phys Chem Chem Phys* 16(14):6508–6518
14. Axe JM, Yezdimer EM, O'Rourke KF et al (2014) Amino acid networks in a $(\beta/\alpha)_8$ barrel enzyme change during catalytic turnover. *J Am Chem Soc* 136(19):6818–6821
 15. Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 73:1–16
 16. Selvaratnam R (2013) Probing allostery in the exchange protein activated by cAMP (EPAC) using NMR spectroscopy. Unpublished doctoral thesis, McMaster University, Hamilton, ON, Canada
 17. Selvaratnam R, VanSchouwen B, Fogolari F et al (2012) The projection analysis of NMR chemical shifts reveals extended EPAC autoinhibition determinants. *Biophys J* 102(3):630–639
 18. Chatterjee S, Firat A (2007) Generating data with identical statistics but dissimilar graphics. *Am Stat* 61(3):248–254
 19. Axe JM, O'Rourke KF, Kerstetter NE et al (2015) Severing of a hydrogen bond disrupts amino acid networks in the catalytically active state of the alpha subunit of tryptophan synthase. *Protein Sci* 24(4):484–494
 20. Axe JM, Boehr DD (2013) Long-range interactions in the alpha subunit of tryptophan synthase help to coordinate ligand binding, catalysis, and substrate channeling. *J Mol Biol* 425(9):1527–1545
 21. de Hoon MJ, Imoto S, Nolan J et al (2004) Open source clustering software. *Bioinformatics* 20(9):1453–1454
 22. Saldanha AJ (2004) Java treeview—extensible visualization of microarray data. *Bioinformatics* 20(17):3246–3248
 23. Tomaselli S, Pagano K, Boulton S et al (2015) Lipid binding protein response to a bile acid library: a combined NMR and statistical approach. *FEBS J* 282(21):4094–4113

High-Efficiency Expression of Yeast-Derived G-Protein Coupled Receptors and ^{19}F Labeling for Dynamical Studies

Libin Ye, Alexander P. Oraziotti, Aditya Pandey, and R. Scott Prosser

Abstract

We describe a detailed protocol for heterologous expression of the human adenosine A_{2A} G-protein coupled receptor (GPCR), using *Pichia pastoris*. Details are also provided for the reconstitution and functional purification steps. Yields of 2–6 mg/g membrane were obtained prior to functional purification (ligand column purification). Typically, functional purification reduced overall yields by a factor of 2–4, resulting in final functional production of 0.5–3 mg/L membrane. Yeast is an excellent protein expression system for NMR given its high tolerance for isotope-enriched solvents and its ability to grow in minimal media.

Key words *P. pastoris*, GPCR, Membrane protein, Functional purification, $A_{2A}R$

1 Introduction

In eukaryotes, signal transduction associated with vision, sensory-, autonomic-, and neurotransmitter-mediated response, inflammation, and cell homeostasis are regulated by a diverse class of membrane receptors, known as G-protein coupled receptors (GPCRs). There are over 800 human GPCRs, all of which possess a common 7-transmembrane motif [1]. Signal transduction is generally regulated by an extracellular ligand that when bound to the orthosteric binding pocket of the receptor establishes an allosteric path between the GPCR and intracellular signaling proteins such as G proteins or β -arrestins. One third of currently marketed drugs [2] target GPCRs and many drug discovery efforts are currently focused on their application to treat cancer, obesity, cardiac disease, and neurodegenerative diseases. Considering their location in the plasma membrane and their ubiquitous role in the regulation of signal transduction, there is an immense interest in understanding receptor and drug interactions.

Despite their widespread interest to the medical community, progress in understanding the relationship between GPCR structure and function has been mired by challenges associated with obtaining high yield of functionally reconstituted receptor. The past decade has witnessed a renaissance for structural biology in terms of capacity to determine high-resolution crystal structures of GPCRs in complex with specific ligands or, most recently, binding partners such as G proteins and β -arrestins.

Methodological breakthroughs in X-ray crystallography have recently provided high-resolution structures of dozens of GPCRs since 2007, spanning representative functional classes [3]. In contrast to more than 50 unique GPCR X-ray crystal structures currently available, a single structure of a chemokine receptor, CXCR1, has been obtained by solid-state NMR [4]. Current NMR initiatives have thus far focused on the identification of mutations that improve fold stability (and spectroscopic resolution) at higher temperatures while addressing issues associated with heterologous expression that allow for the facile incorporation of ^{13}C , ^{15}N , and ^2H isotopes [5–8]. Studies of related 7-transmembrane receptors suggest that the requisite resolution is at hand to allow for triple-resonance (^1H , ^{13}C , ^{15}N) experiments and subsequent structure studies by NMR [9–13]. However, the greatest challenge with each receptor is to identify a robust heterologous expression system that generates functional isotopically enriched receptors. Insect cells remain promising systems for the expression of GPCRs and many NMR initiatives are focused on the use of insect cells containing deuterated solvent and ^{15}N , ^{13}C , and ^2H -enriched amino acids derived from algal cells or yeast extract [14]. Insect cells nevertheless have a reputation for being prohibitively costly and less versatile when factoring in the cost of isotope-enriched media. The approach in our lab and others has been to make use of yeast (e.g., *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*, and *Yarrowia lipolytica*) expression systems, in part owing to their high tolerance for D_2O , amenability to the introduction of inexpensive ^{13}C and ^{15}N sources [15–19], and high cell densities. Yeast expression systems are also relatively inexpensive, can accomplish post-translational modifications, and have been shown to provide very high yields of mammalian proteins, including GPCRs [20, 21].

Despite the technical challenges, there are several key advantages that NMR brings to the study of GPCRs. First, fusion constructs that are often employed in crystallography to obtain crystals, are generally not a part of the NMR regimen, and in many cases, it is possible to make use of wild-type sequences, assuming expression is not unduly compromised. GPCRs are also reputed to undergo both slow and fast exchange between multiple conformations, associated with distinct functional states [22–25]. While this complicates NMR structure studies and backbone assignments, the

spectroscopic characterization of specific functional states and their exchange rates greatly enriches our understanding of the conformational landscape and mechanisms associated with receptor activation. In particular, ^{19}F NMR is sensitive to subtle changes in electrostatic and van der Waals environment and thus to the activation process, allowing the characterization of the conformational landscape as a function of ligand and/or interacting species [22, 23, 26]. While ^{19}F -enriched amino acids (in particular, fluoro-aromatic residues) can in principle be incorporated into GPCRs [27], the majority of ^{19}F NMR spectroscopic studies of GPCRs have been accomplished via thiol-specific tagging of optimized fluorinated derivatives [28]. Most recently, in our lab, we succeeded in high-yield expression, ^{19}F -labeling, and reconstitution of two class A GPCRs using *Pichia pastoris* [29, 30]. Herein, using the adenosine A_2 receptor ($A_{2A}R$) as an example, we provide details of our *in-house* protocol associated with screening of transformants, expression, reconstitution in detergents, and purification of functional receptor for NMR sample preparation.

2 Materials

2.1 Materials and Apparatus Commonly Used for Molecular Biology

1. pPIC9K vector (*see Note 1*).
2. *Pichia pastoris* SMD1163 strain (*see Note 2*).
3. Glycerol.
4. 1 M ice-cold sterile sorbitol.
5. Geneticin[®] 418.
6. XL 10-Gold[®] ultracompetent *E. coli* cells.
7. YPD plate: 1% yeast extract, 2% peptone, 2% glucose, and 1.5% agar.
8. YPD broth: 1% yeast extract, 2% peptone, 2% glucose.
9. YNBD plate: 1.34% yeast nitrogen base without amino acid, 1% glucose, $4 \times 10^{-5}\%$ D-Biotin, and 1.5% agar.
10. *Bam*HI-HF, *Not*I-HF, and *Pme*I-HF restriction enzymes.
11. T4 DNA ligase.
12. GenElute[™] plasmid miniprep kit (Sigma-Aldrich).
13. 100% ethanol.
14. 1.5 mL Eppendorf tubes.
15. 2 mm electroporation cuvette.
16. 14 mL round-bottom Falcon tubes.
17. UV-vis spectrometer.
18. Spectrophotometer (NanoDrop[™] 2000).
19. Electroporation system (Gene Pulser II from Bio-Rad).

2.2 Materials for Screening of Transformants

1. YPD plates with different concentrations of Geneticin[®] 418.
2. All materials for electroporation as listed in Subheading 2.1.
3. BMGY medium: 1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) YNB without amino acids, 4×10^{-5} % (w/v) D-Biotin, 1% (w/v) glycerol, 0.1 M phosphate buffer at pH 6.5.
4. BMMY medium: 1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base without amino acids, 4×10^{-5} % (w/v) D-Biotin, 0.5% (w/v) methanol, 0.1 M phosphate buffer at pH 6.5, 0.04% (w/v) histidine and 3% (v/v) DMSO, 10 mM theophylline.
5. Methanol.
6. P1 buffer: 50 mM HEPES, pH 7.4.
7. Lysis buffer P2: 50 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM EDTA, 10% glycerol, 100 U Zymolyase.
8. Anti-His6 peroxidase kit.
9. Nitrocellulose membranes.
10. Blocking buffer: 125 mM NaCl, 25 mM Tris base, pH 7.5, 0.3% Tween-20, and 3% nonfat milk.
11. Incubation buffer: Anti-His antibody diluted to 1:2000 with blocking buffer.
12. Washing buffer: 125 mM NaCl, 25 mM Tris base, pH 7.5, 0.3% Tween-20.

2.3 Materials for Functional Ligand Column Preparation

1. Xanthine amine congener $\geq 96\%$ (Sigma-Aldrich).
2. DMSO.
3. Affi-Gel[®] 10 (BioRad).
4. Isopropanol.
5. Trisaminomethane (Tris).
6. Acetic acid.
7. NaN_3 .

2.4 Materials for Cell Culture and Induced Expression

1. YPD plates.
2. YPD broth.
3. Buffered glycerol complex (BMGY) medium.
4. Buffered methanol complex (BMMY) medium.
5. Methanol.
6. Anti-foam A.

2.5 Materials for Membrane Preparation and Functional Purification

1. Buffer P1: 50 mM HEPES, pH 7.4
2. Lysis buffer P2: 50 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM EDTA, 10% glycerol, 100 U Zymolyase, EDTA-free-proteinase inhibitor (Sigma-Aldrich).
3. Ultracentrifuge.
4. Buffer P3: 50 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM theophylline, 10% glycerol, EDTA-free-protease inhibitor, solution of 1% (Lauryl Maltose Neopentyl Glycol) MNG (Anatrace), 0.2% cholesterol hemisuccinate (CHS), and 20 mM imidazole.
5. TALON[®] Metal Affinity Resin (Clontech Laboratories, Inc.).
6. Binding Buffer P4: 50 mM HEPES, pH 7.4, 100 mM NaCl, 2mM theophylline, 10% glycerol, EDTA-free protease inhibitor, solution of 0.1% MNG and 0.02% CHS, 20 mM imidazole.
7. Buffer P4+1: 50 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM theophylline, 10% glycerol, EDTA-free protease inhibitor, solution of 0.1% MNG, and 0.02% CHS.
8. Buffer P5: 50 mM HEPES, pH 7.4, 0.1% MNG, 0.02% CHS, 100 mM NaCl.
9. Buffer P6: 50 mM HEPES, pH 7.4, 100 mM NaCl, solution of 0.1% MNG and 0.02% CHS, 250 mM imidazole.
10. Buffer P7: 50 mM HEPES, pH 7.4, 0.1% MNG, 0.02% CHS.
11. Buffer P8: 50 mM HEPES, pH 7.4, 0.1% MNG, 0.02% CHS, 100 mM NaCl, 20 mM theophylline.

3 Methods

3.1 Gene Manipulation and Transformant Optimization

3.1.1 Preparation of Competent Cells

1. SMD1163 cells were streaked from a $-80\text{ }^{\circ}\text{C}$ stored glycerol stock onto YPD plates and incubated for 3–5 days at $30\text{ }^{\circ}\text{C}$ before making competent cells (*see Note 3*).
2. A single colony from streaked YPD plates was inoculated into 10 mL YPD broth, and incubated with shaking at 275 rpm at $30\text{ }^{\circ}\text{C}$ until an OD_{600} of ~ 3.0 was reached.
3. The 10 mL cell broth was then re-inoculated into 500 mL YPD broth and the OD_{600} was continuously monitored until such time that an OD_{600} of 1.0–1.3 was reached (typically requiring 20–24 h).
4. The cell pellets were collected into pre-autoclaved centrifugation tubes, upon centrifugation at $1500 \times g$ for 5 min at $4\text{ }^{\circ}\text{C}$.
5. The cell pellets were washed twice with $10\times$ autoclaved ice-cold distilled H_2O (*see Note 4*).



Gene fragment for Flag-tag: GACTACAAGGACGATGACGATAAAG

Gene fragment for α -Factor:

ATGAGATTTCTTCAATTTTACTGCAGTTTTATTTCGAGCATCCT
 CCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACG
 GCACAAATTCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGA
 AGGGGATTCGATGTTGCTGTTTTGCCATTTCCAACAGCACAA
 ATAAGGGTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCT
 AAAGAAGAAGGGGTATCTCTCGAAAAAAGA

Fig. 1 Gene fragment inserted into the pPIC9K vector, containing the target gene, surrounded by α -Factor signal peptide, FLAG-tag, and His₆-tag

6. The final cell pellets were resuspended in 1 mL 1 M sterile ice-cold sorbitol (*see Note 5*).
7. 80 μ L of resuspended cells were aliquoted into 1.5 mL pre-chilled sterile Eppendorf™ tubes, for electro-transformation (*see Note 6*).

3.1.2 Preparation of Linearized Plasmid Containing Target Gene

Prior to electroporation, the plasmid should be ready in accordance with the following procedure.

1. The target gene fragment (Fig. 1) including the target adenosine A_{2A} (2-317) receptor [23] gene, a fragment of *Saccharomyces cerevisiae* derived from N-terminal α -Factor signal peptide (α -Factor), an N-terminal FLAG-tag, and a C-terminal His₁₀-tag gene fragment were ligated by multi-PCR reactions.
2. Both the products from the PCR reaction and pPIC9K plasmid were digested for 2 h at 37 °C by restriction enzymes *Bam*HI-HF and *Not*I-HF.
3. DNA gel electrophoresis was applied to both digested PCR products and the pPIC9K plasmid and the resulting products were purified via a QIAquick® Gel Extraction Kit.
4. The extracted PCR products and empty pPIC9K were then ligated by T4 DNA ligase.
5. The plasmid thus obtained was amplified in XL 10-Gold® ultracompetent cells and extracted by GenElute™ plasmid miniprep kit.
6. The concentration of the extracted plasmid was next measured using a NanoDrop™ 2000 spectrophotometer.
7. In accordance with the concentration of the extracted plasmid, *Pme*I-HF was added and incubated for 2 h to fully linearize the plasmid.

8. *PmeI*-HF enzyme was denatured at 80 °C for 30 min (Note: This step is OPTIONAL).
9. The linearized plasmid was then precipitated by 70% ethanol (*see* **Note 7**).
10. The precipitated DNA was then micro-centrifuged for 1 min at 13,000 × *g*.
11. The supernatant was discarded and the precipitated linearized DNA was dried under a fume hood or vacuum for 20–30 min.
12. The linearized DNA was dissolved in distilled water to a final concentration of 500–1000 ng/μL.

3.1.3 Electroporation

The electroporation method was used to achieve multi-copy insertion of the plasmid.

1. 5–10 μg (about 10 μL) linearized plasmid was gently mixed with 80 μL freshly prepared *P. pastoris* competent cells and kept on ice for 5 min before electroporation.
2. The mixture of plasmids and competent cells was transferred to a prechilled 2 mm electroporation cuvette.
3. Electroporation was performed on a Gene Pulser II.
4. The electroporation conditions were 1500–2000 V charging voltage, 25 μF capacitance, and 400 Ω resistance (*see* **Note 8**).
5. 1 mL ice-cold 1 M sorbitol was immediately added to the transformation cuvette.
6. The sample was then transferred to a sterile 14 mL round-bottom tube and incubated for 2–3 h at 30 °C without shaking.
7. 200 μL of medium containing transformed plasmids was spread on YNBD plates, and incubated for 3–5 days until colonies appeared.

3.2 High-Yield Transformant Screening

The screening to obtain a desirable transformant was carried out in two stages, that is, high-copy transformants screening and high-yield expression transformants screening. The main procedure for screening high-copy transformants was in accordance with the previous publication [31] in addition to a secondary screening cycle to increase the probability of high-copy transformants.

3.2.1 Two-Stage Screening Approach for High-Copy Transformants

1. After 3–5 days incubation at 30 °C, the transformants typically appeared on histidine-deficient YNBD plates with the number of colony-forming units (CFU) varying from hundreds to thousands.
2. The transformants were then inoculated individually onto YPD plates containing 0.1 mg/mL G418 for an additional 3–5 days incubation at 30 °C.

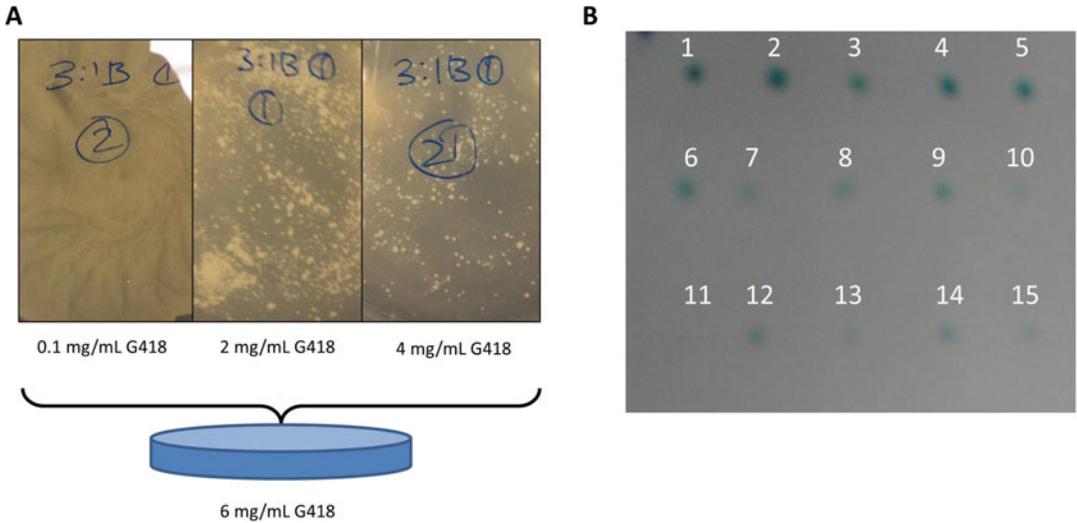


Fig. 2 The secondary screening plates containing different concentrations of G418 (a) and immunoblot images for crude expression screening of colonies chosen from 6 mg/mL G418 plates (b)

3. The colonies were subsequently transferred to YPD plates with 2 mg/mL G418 for an additional incubation period of 5–7 days; transformants were then transferred onto YPD plates containing 4 mg/mL G418 and incubated for an additional 5–7 days (*see Note 9*).
4. The colonies with high-copy target-gene integration can be easily judged from colony size/morphology.
5. The colonies resistant to high concentrations of G418 were selected to make competent cells for second-cycle screening (*see Note 10*).
6. The secondary electro-transformation was performed using a high-copy transformant from the first cycle screening to make competent cells.
7. After transformation, the secondary transformants were directly spread onto YPD plates with different G418 concentrations (2 mg/mL and 4 mg/mL), rather than YNBD plates (*see Note 11*).
8. The colonies that grew on 4 mg/mL G418 YPD plates were further transferred onto 6 mg/mL G418 YPD plates for an additional 5–7 day incubation period.
9. 10–15 high-copy colonies were picked from YPD plates containing 6 mg/mL G418 for further expression to screen high-yield transformants (Fig. 2).

3.2.2 High-Yield Transformants Screening via Immunoblotting

1. 10–15 different single colonies from YPD plates containing 6 mg/mL G418 were inoculated into 5 mL BMGY medium in 14 mL falcon tubes, and were cultured at 30 °C for at least 24 h with shaking (275 rpm) until an OD₆₀₀ between 2 and 6 was obtained.
2. The medium was then transferred into 250 mL BMMY media in Erlenmeyer flasks and was further cultured at 22 °C for 60 h (275 rpm) (*see Note 12*).
3. During cell growth, 0.5% methanol was added every 18 h to induce receptor expression.
4. At the end of induction, the cell pellets were collected after centrifugation at 4000 × *g* for 15 min.
5. The cell pellets were washed once using buffer P1.
6. The cell pellets were resuspended into lysis buffer P2 and left at room temperature for 1 h (*see Note 13*).
7. The cell pellet suspensions were then vortexed at 2000 rpm for 2 h at 4 °C in the presence of a slurry of 5 mm glass beads.
8. The disrupted cell pellets were centrifuged at 8000 × *g* for 30 min at 4 °C to discard unbroken cells and cellular debris.
9. The supernatant containing cell membrane was collected and directly applied to immunoblotting membrane.
10. Immunoblotting was performed by anti-His₆-peroxidase.
11. 1 μL of the supernatant was blotted on nitrocellulose membrane and allowed to dry (*see Note 14*).
12. The membrane was placed in blocking buffer for 1 h at room temperature.
13. The membrane was then transferred to incubation buffer containing anti-His₆-peroxidase antibody (1:2000) at room temperature for 2 h.
14. The membrane was then washed three times with washing buffer, followed by distilled water.
15. The membrane was visualized by reacting with BM Blue POD substrate (*see Note 15*).

3.3 Ligand Column Synthesis

Obtaining functional receptor is critical for NMR studies. For this purpose, affinity purification using a column with immobilized ligand is the key to removing misfolded receptor. The synthesis of an XAC ligand (xanthine amine congener, a functionalized antagonist) column for the adenosine A_{2A} receptor was closely followed from the protocols proposed by Nakata et al. and Grisshammer et al. [32, 33] with slight modification, as described below:

1. 0.01 mmol XAC (4.285 mg) was dissolved in 10 mL of DMSO (*see Note 16*).

2. Prior to XAC addition, Affi-Gel[®] 10 was washed extensively (at least 20×) with ice-cold isopropanol and then briefly with DMSO (*see Note 17*).
3. The moist gel cake was immediately resuspended in 120 mL DMSO containing 51 mg XAC.
4. The gel suspension was incubated at room temperature overnight with slow continuous rotation/shaking.
5. The reaction was stopped by washing the gel with DMSO extensively.
6. The gel was further washed sequentially with distilled H₂O, 1 M Tris, and distilled H₂O.
7. The washed gel was incubated with 200 mM Tris acetate buffer, pH 8.0, for 24 h at 4 °C.
8. The gel was then washed with distilled water extensively.
9. Finally, the XAC-agarose was stored at 4 °C in 0.02% NaN₃.
10. The amount of covalently bound XAC was estimated by monitoring the absorbance at 310 nm in 0.01 N HCl in comparison to that of the starting XAC solution.

3.4 Cell Culture and Induced Expression

1. The high yield transformant was streaked from a glycerol stock onto YPD agar plates and was incubated for 3–5 days until clear single colonies appeared on the plates.
2. A single colony was inoculated in 4 mL YPD broth at 30 °C with shaking at 275 rpm overnight (usually 12–14 h).
3. The cells were then transferred into the 200 mL BMGY medium and cultured at 30 °C for about 20–24 h until the OD₆₀₀ reached 2–6.
4. The cell pellets were spun down at 4000 × *g* for 15 min, and were resuspended in the 1 L BMMY medium to begin expression under induction with methanol at 20–22 °C.
5. 0.5% methanol was added every 18 h to continue induction of receptor expression.
6. After 60 h methanol induction, cell pellets were harvested for receptor purification.

3.5 Membrane Preparation, ¹⁹F Labeling, and Functional Purification

1. The cell pellets were washed one time with buffer P1.
2. Harvested cells were suspended in 4 mL of Lysis buffer P2 at a ratio of 1:4 (*see Note 18*) and left at room temperature (20 °C) for 1 h.
3. Yeast cell walls were further disrupted by the addition of 5 mm glass beads to the resuspended cells and vortexing at 4 °C for 2 h.

4. Intact cells and cell debris were separated from the membrane suspension by low-speed centrifugation ($8,000 \times g$, 30 min, 4°C).
5. The supernatant was collected and centrifuged at $100,000 \times g$ for 1 h.
6. The cell pellet was dissolved in 25 mL buffer P3 under continuous shaking for about 1–2 h, at 4°C , until complete resuspension was achieved.
7. TALON[®] Metal Affinity Resin was washed with distilled water twice and with binding buffer P4 once, then combined with the resuspended membrane under gentle shaking for at least 2 h at 4°C .
8. The suspension was then centrifuged at 4°C to collect A_{2A}R conjugated TALON[®] Metal Affinity Resin at $500 \times g$ for 2 min (*see Note 19*).
9. The A_{2A}R-conjugated resin was washed extensively (at least $5\times$) with buffer P4+1 to remove nonconjugated receptor and impurities in solution.
10. The mixture of TALON[®] resin and supernatant was incubated at 4°C on a shaker for 30 min with $100\ \mu\text{M}$ Tris (2-carboxyethyl)phosphine hydrochloride (TCEP).
11. The TCEP was washed out immediately with buffer P4+1 twice, and the receptor was resuspended in P4+1 buffer.
12. $200\ \mu\text{M}$ ¹⁹F tag (BTFMA) was added into the solution, and incubated at 4°C on a shaker for 6 h under nitrogen.
13. The solution was washed once with buffer P4+1 and another aliquot of ¹⁹F tag was added. The solution was then incubated for an additional 6 h.
14. A_{2A}R-conjugated Talon resin was loaded on an empty column, washed with $10\times$ bed volumes of buffer P5, and the apo A_{2A}R was eluted with at least $5\times$ bed volume of buffer P6.
15. The eluate containing apo-state A_{2A}R was concentrated to 5 mL.
16. The NaCl and imidazole concentrations in the solution were decreased by the addition of 15 mL buffer P7.
17. The XAC-agarose gel was washed with distilled water twice, and equilibrated with buffer P7.
18. The receptor was then incubated with XAC-agarose gel for at least 1 h with gentle shaking at 4°C to allow for A_{2A}R binding to XAC agarose.
19. A_{2A}R-conjugated XAC agarose was packed on a disposable column.
20. The column was washed with buffer P7 for $2\times$ column volumes to remove non-bound receptor.

21. A_{2A}R was then eluted with 10× bed volumes of buffer P8.
22. The eluted samples were concentrated into 30–40 mL.
23. See **Note 20** for alternatives to the steps described below.
24. The sample was combined with TALON[®] resin which was then incubated for 2 h with the functional purified A_{2A}R.
25. The functional A_{2A}R-conjugated resin was packed on a column and washed extensively with buffer P5 (at least 10× column volumes) to remove all theophylline (see **Note 21**).
26. Functional apo A_{2A}R was then eluted out by buffer P6.
27. Functional apo A_{2A}R was concentrated into 1–2 mL and dialyzed twice against Buffer P5 with dilution factor of 1×10^6 .
28. Functional apo-A_{2A}R can now be prepared for NMR samples or other biophysical measurements.

4 Notes

1. The modified vector has a *Bam*HI restriction site in front of the α -Factor, and a *Not*I restrict site following his-tag; refer to Fig. 1.
2. A methylotrophic yeast *P. pastoris* protease-deficient strain—SMD1163 ($\Delta his4$, $\Delta pep4$, $\Delta prb1$), which is deficient in both proteinases A (encoded by *pep4*) and B (encoded by *prb1*) genes, was used as an expression host system.
3. The strain was kept in 15% glycerol stock. The normal cultured cells should emit a fragrant smell.
4. Complete cell dispersal can be achieved by gently pipetting up and down during wash. Cell pellets should always be kept on ice.
5. At this point, the cell solution is very viscous and quite difficult to pipette by a micropipettor. However, this viscous condition is needed. Cells should not be diluted by adding more ice-cold sorbitol.
6. To reach high efficiency, always make fresh *P. pastoris* competent cells for electro-transformation.
7. At this point, linearized DNA can usually be seen by visual inspection after ethanol addition.
8. During electroporation, quickly wipe off moisture from the cuvette and insert into the electroporation device.
9. In this process, the same colonies were re-transferred onto grid plates with different concentrations of G418. For instance, the same colony will be simultaneously transferred onto plates containing 2 mg/mL and 4 mg/mL G418, respectively. As

different colonies will have unique copies of the gene of interest, the different survival probabilities for distinct colonies will be easily used for judging the copy number of colonies. It should be noted the procedures of making the secondary competent cells and electroporation were the same as in the first electroporation step.

10. Note the procedures for making the secondary competent cells and electroporation are unchanged from the first electroporation step.
11. Usually, hundreds of CFUs will appear on plates containing 4 mg/mL G418 at this stage, in contrast to the first-cycle screening where directly spreading the medium onto plates with 4 mL/mL G418 produces no visible colony.
12. Keep the temperature at 20–22 °C to increase the yield of the functional receptor.
13. Ensure the sample is left at room temperature to optimize Zymolyase activity and ensure cell walls are disrupted. Alternatively, you can use lyticase instead of Zymolyase.
14. Nitrocellulose membranes are essential as PVDF membranes are not suitable for immunoblotting with supernatant applied directly onto the membrane.
15. Different transformants will give rise to different intensities for the immunoblots and the transformant showing strongest intensity should be used for further expression.
16. MW of XAC is 428.5. The XAC concentration is 1 mM when dissolved in 10 mL of DMSO. For 20 mL of Affi-Gel[®] 10, 51 mg XAC must be dissolved in 120 mL DMSO. Note 12 mL XAC in DMSO for 1 mL of packed gel.
17. Keep the temperature at 10 °C; otherwise, DMSO and Affi-Gel[®] 10 will be freeze below this temperature.
18. 1 g cell pellets were resuspended in 4 mL Lysis buffer.
19. Do not exceed 1000 × *g*; otherwise the resin will be crushed. Refer to the manual for different resins from different suppliers.
20. The following steps can be alternatively replaced by extensive dialysis against buffer P6.
21. This and subsequent steps are required to return the receptor to an apo state.

References

1. Lv X, Liu J, Shi Q, Tan Q, Wu D, Skinner JJ, Walker AL, Zhao L, Gu X, Chen N, Xue L, Si P, Zhang L, Wang Z, Katritch V, Liu ZJ, Stevens RC (2016) In vitro expression and analysis of the 826 human G protein-coupled receptors. *Protein Cell* 7(5):325–337. doi:[10.1007/s13238-016-0263-8](https://doi.org/10.1007/s13238-016-0263-8)

2. Lundstrom K (2016) New winds in GPCR-based drug discovery. *Future Med Chem* 8 (6):605–608. doi:[10.4155/fmc-2016-0008](https://doi.org/10.4155/fmc-2016-0008)
3. Katritch V, Cherezov V, Stevens RC (2013) Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol* 53(1):531–556. doi:[10.1146/annurev-pharmtox-032112-135923](https://doi.org/10.1146/annurev-pharmtox-032112-135923)
4. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM, Opella SJ (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491 (7426):779–783. doi:[10.1038/nature11580](https://doi.org/10.1038/nature11580)
5. Scott DJ, Plückthun A (2013) Direct molecular evolution of detergent-stable G protein-coupled receptors using polymer encapsulated cells. *J Mol Biol* 425(3):662–677. doi:[10.1016/j.jmb.2012.11.015](https://doi.org/10.1016/j.jmb.2012.11.015)
6. Klenk C, Ehrenmann J, Schütz M, Plückthun A (2016) A generic selection system for improved expression and thermostability of G protein-coupled receptors by directed evolution. *Sci Rep* 6:21294. doi:[10.1038/srep21294](https://doi.org/10.1038/srep21294)
7. Schütz M, Schöppe J, Sedlák E, Hillenbrand M, Nagy-Davidescu G, Ehrenmann J, Klenk C, Egloff P, Kummer L, Plückthun A (2016) Directed evolution of G protein-coupled receptors in yeast for higher functional production in eukaryotic expression hosts. *Sci Rep* 6:21508. doi:[10.1038/srep21508](https://doi.org/10.1038/srep21508)
8. Isogai S, Deupi X, Opitz C, Heydenreich FM, Tsai C-J, Brueckner F, Schertler GFX, Veprintsev DB, Grzesiek S (2016) Backbone NMR reveals allosteric signal transduction networks in the β_1 -adrenergic receptor. *Nature* 530 (7589):237–241. doi:[10.1038/nature16577](https://doi.org/10.1038/nature16577)
9. Etzkorn M, Raschle T, Hagn F, Gelev V, Rice AJ, Walz T, Wagner G (2013) Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: biophysical properties and NMR accessibility. *Structure* 21(3):394–401. doi:[10.1016/j.str.2013.01.005](https://doi.org/10.1016/j.str.2013.01.005)
10. Gautier A, Kirkpatrick JP, Nietlispach D (2008) Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed Engl* 47 (38):7297–7300. doi:[10.1002/anie.200802783](https://doi.org/10.1002/anie.200802783)
11. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D (2010) Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol* 17(6):768–774. doi:[10.1038/nsmb.1807](https://doi.org/10.1038/nsmb.1807)
12. Nietlispach D, Gautier A (2011) Solution NMR studies of polytopic alpha-helical membrane proteins. *Curr Opin Struct Biol* 21 (4):497–508. doi:[10.1016/j.sbi.2011.06.009](https://doi.org/10.1016/j.sbi.2011.06.009)
13. Nisius L, Rogowski M, Vangelista L, Grzesiek S (2008) Large-scale expression and purification of the major HIV-1 coreceptor CCR5 and characterization of its interaction with RANTES. *Protein Expr Purif* 61(2):155–162. doi:[10.1016/j.pep.2008.06.001](https://doi.org/10.1016/j.pep.2008.06.001)
14. Opitz C, Isogai S, Grzesiek S (2015) An economic approach to efficient isotope labeling in insect cells using homemade ^{15}N -, ^{13}C - and ^2H -labeled yeast extracts. *J Biomol NMR* 62 (3):373–385. doi:[10.1007/s10858-015-9954-3](https://doi.org/10.1007/s10858-015-9954-3)
15. Fan Y, Emami S, Munro R, Ladizhansky V, Brown LS (2015) Isotope labeling of eukaryotic membrane proteins in yeast for solid-state NMR. *Methods Enzymol* 565:193–212. doi:[10.1016/bs.mie.2015.05.010](https://doi.org/10.1016/bs.mie.2015.05.010)
16. Morgan WD, Kragt A, Feeney J (2000) Expression of deuterium-isotope-labelled protein in the yeast *Pichia pastoris* for NMR studies. *J Biomol NMR* 17(4):337–347
17. Fan Y, Shi L, Ladizhansky V, Brown LS (2011) Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment. *J Biomol NMR* 49(2):151–161. doi:[10.1007/s10858-011-9473-9](https://doi.org/10.1007/s10858-011-9473-9)
18. Sugiki T, Shimada I, Takahashi H (2008) Stable isotope labeling of protein by *Kluyveromyces lactis* for NMR study. *J Biomol NMR* 42 (3):159–162. doi:[10.1007/s10858-008-9276-9](https://doi.org/10.1007/s10858-008-9276-9)
19. Wood MJ, Komives EA (1999) Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. *J Biomol NMR* 13(2):149–159
20. Niebauer RT, Robinson AS (2006) Exceptional total and functional yields of the human adenosine A2a receptor expressed in the yeast *Saccharomyces cerevisiae*. *Protein Expr Purif* 46 (2):204–211. doi:[10.1016/j.pep.2005.09.020](https://doi.org/10.1016/j.pep.2005.09.020)
21. O'Malley MA, Lazarova T, Britton ZT, Robinson AS (2007) High-level expression in *Saccharomyces cerevisiae* enables isolation and spectroscopic characterization of functional human adenosine A_{2a} receptor. *J Struct Biol* 159(2):166–178. doi:[10.1016/j.jsb.2007.05.001](https://doi.org/10.1016/j.jsb.2007.05.001)
22. Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012) Biased signaling pathways in β_2 -adrenergic receptor characterized by ^{19}F -NMR. *Science* 335(6072):1106–1110. doi:[10.1126/science.1215802](https://doi.org/10.1126/science.1215802)

23. Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS (2016) Activation of the A_{2A} adenosine G-protein-coupled receptor by conformational selection. *Nature* 533(7602):265–268. doi:[10.1038/nature17668](https://doi.org/10.1038/nature17668)
24. Horst R, Stanczak P, Stevens RC, Wüthrich K (2012) β_2 -adrenergic receptor solutions for structural biology analyzed with microscale NMR diffusion measurements. *Angew Chem Int Ed* 52(1):331–335. doi:[10.1002/anie.201205474](https://doi.org/10.1002/anie.201205474)
25. Horst R, Liu JJ, Stevens RC, Wüthrich K (2013) β_2 -adrenergic receptor activation by agonists studied with ¹⁹F NMR spectroscopy. *Angew Chem Int Ed* 52(41):10762–10765. doi:[10.1002/anie.201305286](https://doi.org/10.1002/anie.201305286)
26. Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK (2015) Structural insights into the dynamic process of β_2 -adrenergic receptor signaling. *Cell* 161(5):1101–1111. doi:[10.1016/j.cell.2015.04.043](https://doi.org/10.1016/j.cell.2015.04.043)
27. Kitevski-LeBlanc JL, Prosser RS (2011) Current applications of ¹⁹F NMR to studies of protein structure and dynamics. *Prog Nucl Magn Reson Spectrosc* 62:1–33. doi:[10.1016/j.pnmrs.2011.06.003](https://doi.org/10.1016/j.pnmrs.2011.06.003)
28. Ye L, Larda ST, Li YFF, Manglik A, Prosser RS (2015) A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in ¹⁹F NMR studies of proteins. *J Biomol NMR* 62(1):97–103. doi:[10.1007/s10858-015-9922-y](https://doi.org/10.1007/s10858-015-9922-y)
29. Asada H, Uemura T, Yurugi-Kobayashi T, Shiroishi M, Shimamura T, Tsujimoto H, Ito K, Sugawara T, Nakane T, Nomura N, Murata T, Haga T, Iwata S, Kobayashi T (2011) Evaluation of the *Pichia pastoris* expression system for the production of GPCRs for structural analysis. *Microb Cell Fact* 10:24. doi:[10.1186/1475-2859-10-24](https://doi.org/10.1186/1475-2859-10-24)
30. Yurugi-Kobayashi T, Asada H, Shiroishi M, Shimamura T, Funamoto S, Katsuta N, Ito K, Sugawara T, Tokuda N, Tsujimoto H, Murata T, Nomura N, Haga K, Haga T, Iwata S, Kobayashi T (2009) Comparison of functional non-glycosylated GPCRs expression in *Pichia pastoris*. *Biochem Biophys Res Commun* 380(2):271–276. doi:[10.1016/j.bbrc.2009.01.053](https://doi.org/10.1016/j.bbrc.2009.01.053)
31. Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K (1994) Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *Nat Biotechnol* 12:181–184
32. Weiss HM, Grisshammer R (2002) Purification and characterization of the human adenosine A_{2a} receptor functionally expressed in *Escherichia coli*. *Eur J Biochem* 269(1):82–92
33. Nakata H (1989) Purification of A₁ adenosine receptor from rat brain membranes. *J Biol Chem* 264(28):16545–16551

Quantitative Determination of Interacting Protein Surfaces in Prokaryotes and Eukaryotes by Using In-Cell NMR Spectroscopy

David S. Burz, Christopher M. DeMott, Asma Aldousary, Stephen Dansereau, and Alexander Shekhtman

Abstract

This paper describes three protocols for identifying interacting surfaces on ^{15}N -labeled target proteins of known structure by using in-cell NMR spectroscopy. The first protocol describes how to identify protein quinary structure interaction surfaces in prokaryotes by using cross-relaxation-induced polarization transfer, CRIPT, based in-cell NMR. The second protocol describes how to introduce labeled protein into eukaryotic (HeLa) cells via electroporation for CRIPT-based in-cell studies. The third protocol describes how to quantitatively map protein interacting surfaces by utilizing singular value decomposition, SVD, analysis of SStructural INTeractions by in-cell NMR, STINT-NMR, data.

Key words In-cell NMR spectroscopy, Quinary structure, Electrophoresis, Singular value decomposition (SVD), Isotopic labeling, Protein-protein interactions, HeLa cells

1 Introduction

The ultimate goal of structural and biochemical research is to understand how macromolecular interactions give rise to and regulate biological activity in living cells. The challenge is formidable due to the complexity that arises not only from the number of proteins (gene products) expressed by the organism, but also from the combinatorial interactions between them [1, 2]. Despite ongoing efforts to decipher the complex nature of protein interactions, new methods for structurally characterizing protein complexes are needed to fully understand molecular networks [3]. With the onset of in-cell NMR spectroscopy [4] molecular structures can be studied under physiological conditions shedding light on the structural underpinning of biological activity.

The interior of a living cell is an extremely crowded environment and NMR-active nuclei in biological macromolecules are

extremely sensitive to this environment and changes in its composition. In a cell the concentrations of macromolecules can exceed 400 mg/mL [5], hydrophobic and hydrophilic interactions are perturbed, there is limited bulk water, cytosolic water forms solvent shells on protein surfaces [6], and the distances between macromolecules are less than the characteristic Debye screening length of ion charges [7]. In addition, specific and nonspecific binding interactions with ions, small effector ligands and other macromolecules, as well as changes due to biochemical modifications, alter the chemical environment. As a result, in-cell NMR spectra are inherently noisier than spectra acquired in vitro.

The low signal-to-noise ratio arises from the myriad of interactions between the target protein and components of the cytosol. While these interactions do not result in large differences in the chemical shift between spectra acquired in-cell versus in vitro, they do result in slower tumbling times, a dramatic increase in the apparent molecular weight of the target, and extensive broadening of many or all of the spectral peaks [8–11]. For example, Thioredoxin, a 12 kDa protein, exhibits an apparent molecular weight of ~1 MDa in-cell. In particular, due to the high intracellular concentration, ribosomal and messenger RNA interact extensively with target proteins, altering the structure and activity in unpredictable ways [11–13]. These inherently transient interactions constitute a fifth level of protein structural organization, quinary structure [14].

Protein quinary interactions are refractory to biochemical analyses [15]. Studying these interactions requires in-cell methodologies because isolation techniques will disrupt protein complexes with limited thermodynamic stability [16]. Two-dimensional in-cell NMR spectra acquired by using ^{15}N heteronuclear single quantum coherence, ^1H - ^{15}N HSQC, experiments [17] do not resolve signals from target proteins engaged in quinary interactions due to broadened peaks. Transverse proton relaxation of the in-cell NMR signal, T_2 , is so short for high molecular weight complexes that only proteins that are unstructured, such as the prokaryotic ubiquitin-like protein, Pup, and thus have a longer T_2 due to internal flexibility or that interact extremely weakly with the cytosol, have been studied by using ^1H - ^{15}N HSQC.

To overcome the inability to detect high molecular weight species in-cell, relaxation optimized ^{15}N -edited cross-relaxation enhanced polarization transfer, CRINEPT, heteronuclear multiple quantum coherence, HMQC, transverse relaxation optimized spectroscopy, TROSY, ^1H - ^{15}N CRINEPT-HMQC-TROSY experiments are attractive for in-cell NMR studies due to superior sensitivity to NMR signals and relative insensitivity to unavoidable magnetic field inhomogeneity [18]. In vitro solution NMR performed on extensively deuterated supramolecular structures successfully used CRINEPT NMR spectroscopy to study complexes with molecular weights greater than 100 kDa [19]. For in-cell

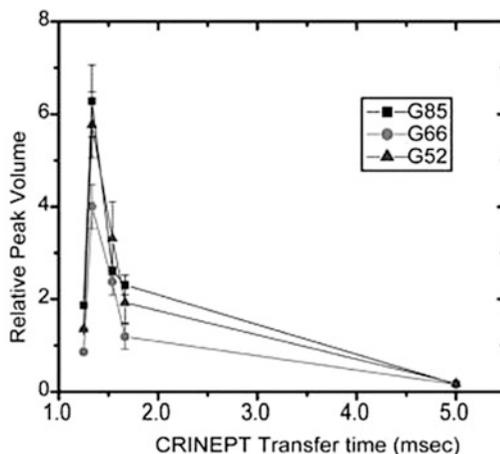


Fig. 1 The relative volumes of the G52, G66, and G85 crosspeaks in the in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectra of Trx are plotted against the CRINEPT transfer delay times. An endogenous tryptophan indole amide peak in the in-cell spectra is used as a reference. The optimum CRINEPT transfer delay for Trx is ~ 1.3 ms, which corresponds to an apparent molecular weight of 1.1 MDa. Reproduced from [11, 37] with permission from the American Chemical Society

experiments, REDuced PROton density (REDPRO) labeling is used to exchange α and β protons of amino acids for deuterons to minimize proton relaxation [20, 21], and the optimal CRINEPT transfer time is calibrated for the target protein in-cell to maximize the NMR peak intensities [11] (Fig. 1). The resulting in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum thus reveals the structure of the protein engaged in quinary interactions (Fig. 2a).

To acquire an in-cell NMR spectrum, the concentration of labeled nuclei must be high enough to provide well-resolved resonances for unambiguous identification. In prokaryotes and some eukaryotes overexpression of target protein in a labeled medium is used to generate an intracellular concentration of target protein sufficiently high to acquire NMR spectra [22–25]. The in-cell NMR spectrum of a fully expressed protein is compared with the spectrum of free protein either purified or from cell lysates. Residues exhibiting the greatest reduction in intensity are presumed to contribute to the quinary interacting surface.

In eukaryotic cells overexpression of target proteins is often too weak to produce an intracellular concentration of target protein sufficient to acquire in-cell NMR spectra. Several methods have been developed to introduce ^{15}N labeled target proteins into cells including cell-penetrating peptides [26], toxins [27], micro injections [28, 29], and electroporation [11, 30]. Each method has its advantages and disadvantages, but all have proven effective in delivering sufficient quantities of labeled target protein to acquire reproducible in-cell NMR spectra (Fig. 3a).

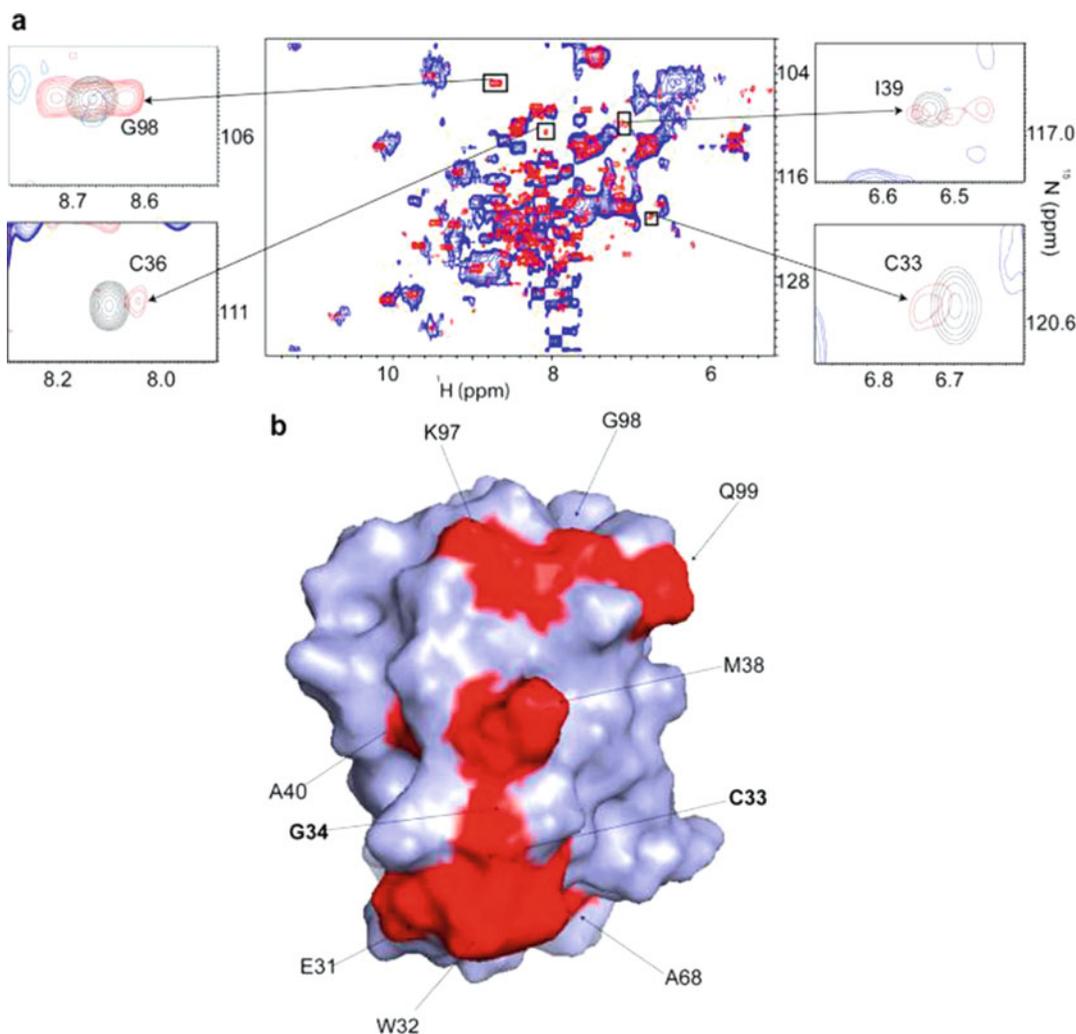


Fig. 2 (a) Overlay of the in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum of REDPRO labeled Trx (*blue*) and that of the cellular lysate (*red*). The intensities of C33, C36, I39, and G98 peaks, from the residues involved in the quinary interactions, are broadened out. The insets show overlays of the boxed regions of the in-cell spectrum (*blue*) and the corresponding regions of the ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum of lysate (*red*), and ^1H - ^{15}N HSQC spectrum of purified Trx in 10 mM potassium phosphate buffer at pH 6.5 (*black*). (b) Residues involved in the quinary interactions (*red*) are mapped onto the molecular surface of Trx (PDB code 1X0B); active site residues, C33 and G34, are in bold. Reproduced from [11] with permission from the American Chemical Society

Following the in-cell experiment, cells are lysed and the ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum is re-acquired on free protein. The difference between the normalized peak intensities of the in-cell and lysate spectra is calculated for each residue (Fig. 4). Residues exhibiting the greatest change in intensity are mapped onto the surface of the target protein (Figs. 2b and 3b). Such analyses can provide insights into the effect of protein quinary structure on the regulation of biological activity in living cells [11, 31].

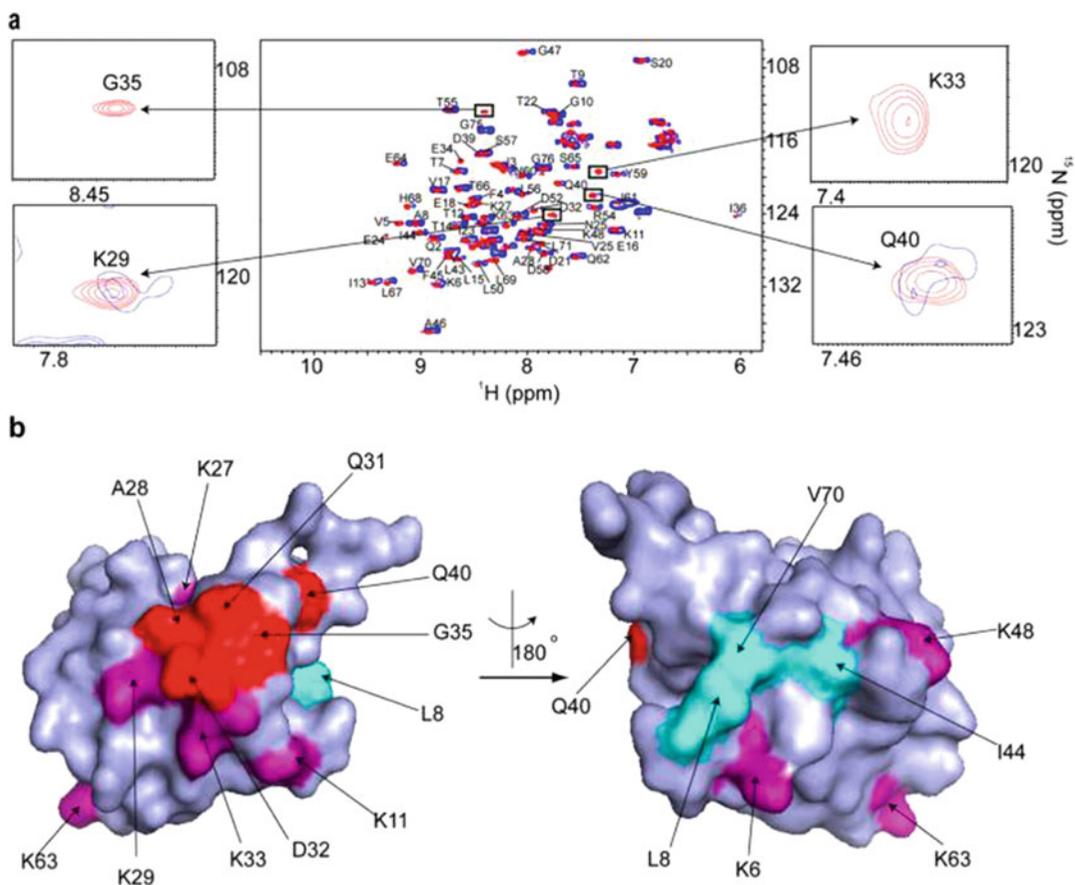


Fig. 3 (a) Overlay of the in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum of REDPRO labeled Ubiquitin electroporated into HeLa cells (blue) and the ^1H - ^{15}N HSQC spectrum of the cell lysate. NMR peaks corresponding to K29, K33, G35, and Q40 (insets) are broadened in the in-cell spectrum, suggesting that Ubiquitin is involved in transient interactions with components of the cytosol. (b) Residues (shown in red), whose NMR peaks are broadened, form a contiguous interaction surface involved in Ubiquitin quinary interactions (PDB code 1D3Z). The seven lysines of Ubiquitin, which are used for ubiquitylation, are in purple. K27, K29, and K33 are a part of the interaction surface and affected by the quinary interactions. The canonical I44 hydrophobic patch of Ubiquitin (shown in cyan) spanning L8, I44, and V70 is unperturbed by quinary interactions. Reproduced from [11] with permission from the American Chemical Society

Unlike quinary interactions, which are of comparatively low (micromolar or weaker) affinity, specific high (sub-micromolar) affinity interactions between a target and interactor protein are studied by using STINT-NMR [32, 33], which elucidates structural interactions between proteins within their native environment by using in-cell NMR [22–25, 33]. In its simplest form, STINT-NMR can identify the interacting surface of a target protein when a single interactor protein (ligand) binds to it. This is accomplished in prokaryotes by overexpressing the target protein in [U - ^{15}N] medium, then sequentially overexpressing interactor protein in unlabeled medium. Samples are collected at various times

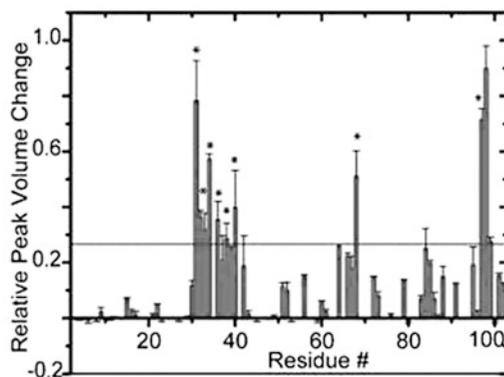


Fig. 4 Bar plot showing the relative changes in in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY peak intensities of Trx residues due to quinary interactions. Residues E31, W32, C33, C36, M38, A40, A68, and Q99, annotated with (*), are also affected in total RNA bound Trx. The horizontal line differentiates residues whose NMR peaks undergo significant broadening. Reproduced from [11] with permission from the American Chemical Society

and the in-cell NMR spectrum of the target protein is acquired. As the concentration of the interactor increases, the spectrum of the target changes to reflect the different chemical environment of the residues that have been affected by the binding interaction. The resulting NMR data provide a complete titration of the interaction and define structural details of the interacting surfaces at atomic resolution.

Conventional analyses of interacting proteins tend to incorrectly estimate the number of residues involved in the interaction because of the widespread signal broadening associated with the formation of a stoichiometric complex. The process of distinguishing which spectral changes are due to specific binding generally considers only the difference between the spectrum of free target protein and the final in-cell target spectrum following full over-expression of the interactor to assess the change in intensity of a given peak resonance [34]. This absolute difference is used to infer whether or not the corresponding amino acid contributes to the principal binding mode of the target. Time-dependent degradation of the target protein inside the cell or differences in sample preparations can lead to changes in the NMR spectra. A rigorous objective analysis of spectral changes is needed to unambiguously differentiate between signals that result from concentration-dependent and concentration-independent processes. This problem can be overcome by analyzing in-cell NMR data by using Singular Value Decomposition, SVD.

SVD is a mathematical technique used to identify the principal components of an arbitrary matrix that contribute maximally to the variance of its elements [35]. Over the course of a STINT-NMR

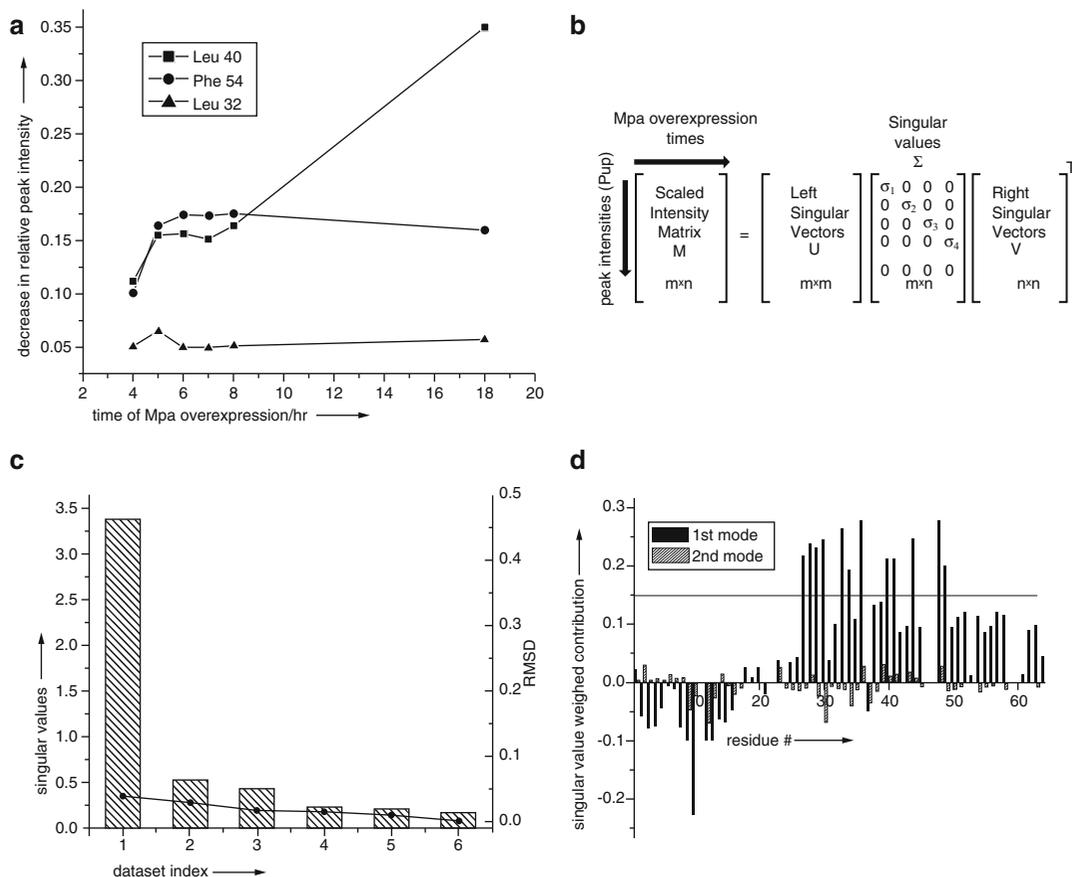


Fig. 5 (a) The intensity of individual cross peaks of Pup change at different rates during interactor protein, Mpa, overexpression. SVD analysis evaluates the magnitude of the contribution of an intensity change to the NMR data over the experimental time course to identify concentration-dependent interactions. (b) SVD of the experimental data matrix M of size $m \times n$ yields the matrices U , Σ , and V^T , with sizes $m \times m$, $m \times n$, and $n \times n$, respectively, where m is the number of target protein amino acid residues used in the NMR analysis, n is the number of time-course NMR data sets, and V^T is the transpose of matrix V . (c) The Scree plot shows the distribution of singular values for each data-set index (binding mode) from 1 to 6. The root mean square deviation, RMSD, values between respective components, and the complete dataset are indicated by solid circles. (d) The weighted contribution of each Pup amino acid residue to the Mpa principal binding modes, calculated as a product of a corresponding singular value and left singular vector, is shown for the 1st (black) and 2nd (hatched) binding modes. The threshold of 0.14 is chosen to highlight the 12 amino acids that exhibit the largest singular value weighted. Negative values are due to spectral overlap between the target protein and cellular metabolites. Reproduced from [37] with permission from Wiley and Sons

titration, a series of in-cell NMR spectra are collected, and a matrix, M , is created that contains the changes in target protein peak intensities versus the expression time of the unlabeled binding partner (Fig. 5a). SVD analysis of matrix M discriminates between changes in the in-cell NMR spectrum of a target protein due to specific and nonspecific interactions and changes due to the presence of the complex cellular environment over the time course of

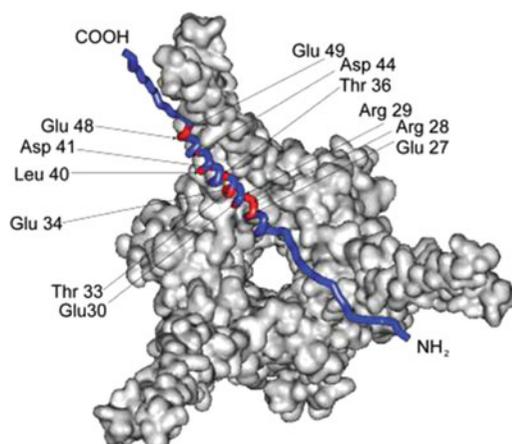


Fig. 6 Residues comprising the principle-binding mode (*red*) between Pup and Mpa are mapped onto a Pup-Mpa complex (PDB code 3M9D) [38]. Reproduced from [37] with permission from Wiley and Sons

interactor overexpression (Fig. 5b). A Scree plot (Fig. 5c) shows the distribution of singular values that define the relative contribution of each binding mode to the change in chemical shift or intensity. An abrupt drop in the singular values following the first binding mode indicates the presence of a single principal binding mode; a gradual decrease in singular values corresponds to random binding of target protein to components of the cytosol. The analysis identifies the amino acid residues involved in the principal binding mode of a target protein with its interactor (Figs. 5d and 6).

This paper describes three protocols for acquiring in-cell NMR spectra on ^{15}N -labeled target proteins. The first protocol describes how to identify protein quinary structure interaction surfaces in prokaryotes by using CRIPT-based in-cell NMR. The second protocol describes how to introduce labeled protein into eukaryotic (HeLa) cells via electroporation for CRIPT-based in-cell NMR. The third protocol describes how to quantitatively map protein interacting surfaces by utilizing SVD analysis of STINT-NMR data.

2 Materials

All the NMR experiments were performed at room temperature using a 700 MHz Bruker Avance spectrometer equipped with a TXI 2-gradient cryoprobe. All data were processed with Topspin 2.1 and analyzed by using CARA software. Interacting residues were mapped onto surface models of proteins using SWISS-PDB Viewer.

**2.1 Determining
Protein Quinary
Interaction Surfaces in
Prokaryotes**

1. pRSF-Trx: Plasmid used to overexpress bacterial Thioredoxin [11].
2. BL21(DE3) codon + (Novagen): *E. coli* strain used to overexpress plasmids.
3. Luria Broth (LB medium): Dissolve 20 g of LB Broth Lennox in 1 L of distilled water and autoclave.
4. Kanamycin (Kn) 1000× stock: Dissolve 0.35 g of kanamycin sulfate in 10 mL of distilled water, sterile filter and store at -20°C .
5. LB-Kn: LB medium supplemented with 35 $\mu\text{g}/\text{mL}$ of kanamycin.
6. LB-Kn agar plates: Dissolve 20 g of LB Broth Lennox in 1 L of distilled water, add 15 g of agar and autoclave. Monitor the temperature as it cools. When the temperature reaches $50\text{--}55^{\circ}\text{C}$, add 1 mL of kanamycin stock solution. Pour 20–30 mL per plate into 100×10 mm Petri dishes. Let the agar solidify and the plates dry on the benchtop for several days. Store dried plates in a plastic bag at 4°C .
7. 1 M CaCl_2 : Dissolve 5.55 g of CaCl_2 in 50 mL of distilled water and sterile filter.
8. M9 salts: Dissolve 6 g of Na_2HPO_4 , 3 g of KH_2PO_4 , and 0.5 g of NaCl in 1 L of distilled water. Adjust the pH to 7–7.4 and add 10 μL of 1 M CaCl_2 . Autoclave and store for up to 1 month at 4°C . Autoclaving after adding CaCl_2 will ensure that any calcium phosphate precipitates will be solubilized.
9. 1 M Mg_2SO_4 : Dissolve 6.02 g of MgSO_4 in 50 mL of distilled water and sterile filter.
10. Solid NH_4Cl and $^{15}\text{NH}_4\text{Cl}$.
11. 1 mg/mL thiamine HCl.
12. 20% D-glucose: Dissolve 20 g of D-glucose in 80 mL of distilled water, sterile filter and store at 4°C .
13. Glycerol.
14. M9 labeling medium: Add 1 g [$U\text{-}^{15}\text{N}$]- NH_4Cl , 2 mL of 1 M Mg_2SO_4 , 1 mL of 1 mg/mL thiamine HCL and 20 mL of 20% (w/v) D-glucose, or 4 mL of glycerol to 1 L of M9 salts.
15. D_2O .
16. Deuterated M9-Kn labeling medium: M9 labeling medium prepared using D_2O and supplemented with 35 $\mu\text{g}/\text{mL}$ of kanamycin.
17. 1 M IPTG: Dissolve 2.383 g IPTG in 10 mL distilled water and store in -20°C freezer.
18. NMR buffer: 10 mM potassium phosphate, pH 6.5.
19. 5 mm standard sample tubes.

**2.2 Determining
Protein Quinary
Interaction Surfaces in
Eukaryotes**

1. 150 cm² cell culture flasks.
2. Low glucose Dulbeccos's Modified Eagle Medium (DMEM).
3. Fetal bovine serum (FBS).
4. Complete medium: DMEM supplemented with 10% FBS.
5. 10× 0.5% Trypsin/5.3 mM EDTA stock (we use Gibco).
6. Balanced salt solution without calcium and magnesium: 5 mM KCl, 440 μM KH₂PO₄, 4.2 mM NaHCO₃, 137 mM NaCl, 340 μM Na₂PO₄·7H₂O.
7. Phosphate-buffered saline (PBS): Dissolve 8 g NaCl, 0.2 g of KCl, 1.44 g of Na₂HPO₄, and 0.24 g of KH₂PO₄ in 1 L of distilled water, adjust the pH to 7.4 and. Store PBS at room temperature.
8. Purified REDPRO-labeled [*U*-¹⁵N]-Ubiquitin [11, 20].
9. Electroporation 2× salts stock: 100 mM sodium phosphate pH 7.2, 5 mM KCl, 15 mM MgCl₂, 15 mM Hepes, 5 mM ATP, 5 mM reduced glutathione.
10. Electroporation buffer: 50% electroporation 2× salts stock and 50% Amaxa Nucleofector Solution R (Lonza).
11. We use a Amaxa™ Nucleofector 2b electroporator (Bio-rad). Any suitable electroporator may be used.
12. Electroporation cuvettes.
13. Culture dishes (Falcon).
14. NMR buffer (*see* Subheading 2.1).
15. D₂O.
16. 5 mm Shigemi sample tubes.

**2.3 Using SVD
Analysis to Determine
Protein-Protein
Structural Interactions**

1. pASK-Pup-GGQ: Plasmid used to overexpress Pup-GGQ [36].
2. pRSF-*Msm* Mpa: Plasmid used to overexpress *Mycobacterium smegmatis*, (*Msm*) mycobacterial proteasome ATPase, Mpa [36].
3. BL21(DE3) codon + (*see* Subheading 2.1).
4. Kanamycin (Kn) 1000× stock (*see* Subheading 2.1).
5. Ampicillin (Ap) 1000× stock: Dissolve 1.00 g of kanamycin sulfate in 10 mL of distilled water, sterile filter and store at −20 °C.
6. LB-Ap-Kn: LB medium (*see* Subheading 2.1) supplemented with 35 μg/mL of kanamycin and 100 μg/mL of ampicillin.
7. LB-Ap-Kn agar plates: Dissolve 20 g of LB Broth Lennox in 1 L of distilled water, add 15 g of agar and autoclave. Monitor the temperature as it cools. When the temperature reaches 50–55 °C, add 1 mL of kanamycin and 1 mL of ampicillin

stock solution. Pour 20–30 mL per plate into 100 × 10 mm Petri dishes. Let the agar solidify and the plates dry on the benchtop for several days. Store dried plates in a plastic bag at 4 °C.

8. M9 salts (*see* Subheading 2.1).
9. M9 labeling medium (*see* Subheading 2.1).
10. M9-Ap-Kn labeling medium: M9 labeling medium supplemented with 35 µg/mL of kanamycin and 100 µg/mL of ampicillin.
11. Dimethylformamide.
12. 2 mg/mL anhydrotetracycline stock: dissolve 20 mg of anhydrotetracycline in dimethylformamide, store at –20 °C.
13. NMR buffer (*see* Subheading 2.1).
14. Glycerol.
15. 1 M IPTG (*see* Subheading 2.1).
16. D₂O.
17. Matlab 2009b or Octave 4.0.1 software.

3 Methods

3.1 Determining Protein Quinary Interaction Surfaces in Prokaryotes

This protocol requires an inducible plasmid that overexpresses the target protein and NMR peak assignments for the target protein. We will use Thioredoxin (Trx) as a specific example of the target protein, although any appropriately cloned target should work. Trx was cloned into expression vector pRSF-1b to yield pRSF-Trx [11]. pRSF-Trx confers kanamycin resistance, codes for lac repressor, and expresses N-terminal His-tagged bacterial Trx from the T7 promoter/lac operon, which is induced by IPTG. *E. coli* strain BL21 (DE3) codon + was transformed with pRSF-Trx using standard techniques and the transformed cells were selected on LB-Kn agar plates. NMR peak assignments for the target protein are required to complete the analysis.

Control experiments are performed to monitor the comparative viability of the transformed cells before and after acquisition of the in-cell spectrum by using cell growth plate assays. Additional controls test for cell leakage by acquiring a spectrum on the sample supernatant.

3.1.1 Target Protein Labeling and Overexpression

1. Inoculate 5 mL of LB-Kn (35 µg/mL) with a single colony of BL21(DE3) codon + containing pRSF-Trx.
2. Grow the culture overnight at 37 °C with shaking at 250 rpm.
3. Dilute the overnight culture to OD₆₀₀ ~0.07 in 50 mL of LB-Kn.

4. Grow the cells at 37 °C, 250 rpm to OD₆₀₀ ~0.9–1.0.
5. Centrifuge the cells at 4500 × *g* for 15 min at room temperature.
6. Wash the cells twice with a total of 50 mL of M9 salts.
7. Recentrifuge the cells and resuspend in 50 mL of deuterated M9-Kn labeling medium.
8. Incubate the culture at 37 °C, 275 rpm for 20 min.
9. Induce overexpression of Trx by adding 0.1% culture volume of 1 M IPTG.
10. Overexpress REDPRO-labeled target protein for up to 36 h at 37 °C, 275 rpm.

3.1.2 Sample Preparation

1. Centrifuge 50 mL of culture at 4500 × *g* for 15 min at room temperature.
2. Wash the cells twice with 50 mL of NMR buffer.
3. Resuspend the cell pellet in 0.45 mL of NMR buffer and 0.05 mL of D₂O. Remove 10 μL for cell viability assays. Transfer the remaining sample to a standard 5 mm NMR tube.
4. Prepare 10⁻⁴, 10⁻⁵, and 10⁻⁶ serial dilutions of the 10 μL cell sample. Plate 100 μL of each dilution onto LB-Kn plates in triplicate. Incubate the plates overnight at 37 °C and count the resulting colonies.

3.1.3 NMR Spectroscopy

All of the experiments are performed at room temperature (RT) using a standard 5 mm NMR tube in a 700 MHz Bruker Avance II NMR spectrometer equipped with TXI *z*-gradient cryoprobe.

1. Shim the sample manually using the *z*, *z*₂, and *z*₃ shims.
2. Select the [¹H-¹⁵N]-CRINEPT-HMQC-[¹H]-TROSY experiment as described by Riek et al. [19] (Fig. 3b in [19]).
3. Apply water selective pulses to align water magnetization along the +*z* axis to achieve water suppression.
4. Set the recycle delay between the transients to 300 ms.
5. Set the number of transients to 64.
6. Set the spectral widths to 12 ppm and 30 ppm in the ¹H and ¹⁵N dimensions respectively.
7. Set 1024 and 1 points in the ¹H and ¹⁵N dimensions, respectively to optimize the CRINEPT transfer delay time. Increase the CRINEPT transfer time by 0.2–0.3 ms starting from 1 ms. Monitor the backbone amide envelope and tryptophan indole proton peaks to determine the CRINEPT transfer time that produces maximum peak volume, i.e., T_{opt} (Fig. 1c). Set T_{opt} to 1.3 ms for Trx.

8. Set 1024 and 128 points in the ^1H and ^{15}N dimensions, respectively.
9. Collect the $[^1\text{H}-^{15}\text{N}]$ -CRINEPT-HMQC- $[^1\text{H}]$ -TROSY spectrum of the sample.
10. When data collection is completed, recover the sample. Remove a 10 μL aliquot for cell viability assays. Prepare 10^{-4} , 10^{-5} , and 10^{-6} serial dilutions of the 10 μL cell sample. Plate 100 μL of each dilution onto LB-Kn plates in triplicate. Incubate the plates overnight at 37 °C and count the resulting colonies.
11. Centrifuge the remaining sample at $4500 \times g$ for 10 min at RT.
12. Transfer the supernatant to a standard 5 mm NMR tube and acquire a $[^1\text{H}-^{15}\text{N}]$ -HSQC spectrum to test for cell leakage.
13. Resuspend the cell pellet in 0.3 mL of NMR buffer.
14. Lyse the remaining cells by freeze-thawing (-80 °C-RT) five times.
15. Centrifuge the sample at $18,000 \times g$ for 15 min at RT.
16. Collect a $[^1\text{H}-^{15}\text{N}]$ -CRINEPT-HMQC- $[^1\text{H}]$ -TROSY spectrum on the clarified lysate using the optimized value for T_{opt} .

3.1.4 Data Processing and Analysis

1. Process data with Topspin 2.1 and analyze by using CARA software. Alternative offline data processing software may also be used.
2. Disregard residues with chemical shifts that differ by >0.1 ppm between the lysate and purified protein.
3. Normalize peak intensities for the in-cell and lysate spectra using an amino proton peak from either a glutamine or an asparagine sidechain that does not change chemical shift between in-cell and lysate spectra.
4. Calculate changes in protein peak intensities, ΔI , due to quinary complexation as

$$\Delta I = (I_{\text{lysate}} - I_{\text{in-cell}}) / I_{\text{in-cell}}$$

where I_{lysate} and $I_{\text{in-cell}}$ are the normalized, integrated peak intensities from the lysate and in-cell spectra.

5. Plot ΔI versus protein residue (bar graph). Positive values correspond to peak broadening.
6. Draw a threshold line that distinguishes peaks with extreme broadening. Corresponding residues exceeding the threshold contribute to the quinary interaction surface.
7. Map broadened residues onto the protein surface based on PDB file(s) using SWISS-PDB Viewer.

3.2 Determining Protein Quinary Interaction Surfaces in Eukaryotes

Overexpressing labeled target protein in eukaryotic cells frequently does not generate an intracellular concentration that is sufficiently high to provide interpretable in-cell spectra. To overcome this problem, purified labeled target protein is introduced into HeLa cells by using electroporation. The protocol is a modification of that presented in Theillet et al. [30]. We will use Ubiquitin (Ubq) as a specific example of the target protein. This protocol requires purified REDPRO-labeled [U - ^{15}N]-Ubq [11, 20], and NMR peak assignments for the target protein to complete the analysis.

3.2.1 Preparation of HeLa Cells

1. Seed $\sim 2 \times 10^6$ cells/flask into six 150 cm² cell culture flasks.
2. Culture the cells for 3 days in 15 mL of complete medium, until 80% confluence is reached ($0.8\text{--}1 \times 10^7$ cells/flask).
3. Remove the growth medium by sterile aspiration.
4. Dilute $10\times$ 0.5% Trypsin/5.3 mM EDTA stock to $1\times$ using balanced salt solution without calcium and magnesium.
5. To harvest the cells add 3 mL of 0.05% Trypsin/530 μM EDTA and incubate at 37 °C for 10 min.
6. Add 12 mL of complete medium to neutralize the Trypsin.
7. Centrifuge the cells at $200 \times g$ for 10 min at RT.
8. Wash the cells with 15 mL of PBS pre-warmed to 37 °C and re-centrifuge.
9. Resuspend the cells to 2×10^6 cells/mL using PBS.
10. Transfer 1 mL of resuspended cells into Eppendorf tubes.
11. Centrifuge the cells at $200 \times g$ for 10 min at RT, discard the supernatant.

3.2.2 Electroporation of HeLa Cells

1. Dilute purified REDPRO labeled Ubq to a final concentration of 0.8–1 mM with sterile filtered, freshly prepared electroporation buffer.
2. Prepare a protein electroporation (EP) sample by adding 100 μL of Ubq to the cell pellet and gently mixing.
3. Transfer the EP sample to an electroporation cuvette.
4. Electroporate the cells at RT using an Amaxa™ Nucleofector 2b electroporator (or any other suitable electroporator), pulse program B-28.
5. Pulse cells two to three times, gently mixing the sample between pulses.
6. Immediately add 1 mL of complete medium, pre-warmed to 37 °C, to each cuvette, and transfer the contents to pre-warmed 15 cm diameter cell culture dishes. Apply three to four samples to each dish.
7. Incubate the culture dishes in 5% CO₂ incubators at 37 °C for 2 h to recover.

3.2.3 Sample Preparation

1. Add 15 mL of complete medium to each culture dish and incubate for 2–3 h.
2. Remove non-adherent cells by sterile aspiration.
3. Wash each culture dish three times with 15 mL of PBS.
4. Harvest the cells by treatment with Trypsin/EDTA (*see* Subheading 3.2.1, steps 4–7).
5. Wash the cells once with 1 mL of PBS and twice with 1 mL of NMR buffer.
6. Resuspend the cells in 250 μ L of NMR buffer containing 10% D₂O.
7. Transfer the cell suspensions into 5 mm diameter Shigemi tubes for NMR analysis.

3.2.4 NMR Spectroscopy

The same procedure is followed to collect in-cell NMR spectra from eukaryotic cells as is used for prokaryotic cells. The major difference being that due to the instability of transfected HeLa cells, the CRIPT transfer delay is not optimized. All of the experiments are performed at RT using a standard 5 mm NMR tube in a 700 MHz Bruker Avance II NMR spectrometer equipped with TXI z-gradient cryoprobe.

1. Shim the sample manually using the z, z2, and z3 shims.
2. Select the [¹H-¹⁵N]-CRINEPT-HMQC-[¹H]-TROSY experiment as described by Riek et al. [19] (Fig. 3b in [19]).
3. Apply water selective pulses to align water magnetization along the +z axis to achieve water suppression.
4. Set the recycle delay between the transients to 300 ms.
5. Set the number of transients to 64.
6. Set the spectral widths to 12 ppm and 30 ppm in the ¹H and ¹⁵N dimensions, respectively.
7. Set T_{opt} to 1.4 ms for Ubq.
8. Set 1024 and 128 points in the ¹H and ¹⁵N dimensions, respectively.
9. Collect the [¹H-¹⁵N]-CRINEPT-HMQC-[¹H]-TROSY spectrum of the sample.
10. When data collection is completed, recover the sample. Centrifuge the cells at 200 × g for 10 min at RT.
11. Transfer the supernatant to a standard 5 mm NMR tube and acquire a [¹H-¹⁵N]-HSQC spectrum to test for cell leakage.
12. Resuspend the cell pellet in 0.3 mL of NMR buffer.
13. Lyse the remaining cells by freeze-thawing (−80 °C-RT) five times.

14. Centrifuge the sample at $18,000 \times g$ for 15 min at RT.
15. Collect a $[^1\text{H}-^{15}\text{N}]$ -CRINEPT-HMQC- $[^1\text{H}]$ -TROSY spectrum on the clarified lysate using the optimized value for T_{opt} .

3.2.5 Data Processing and Analysis

1. Process data with Topspin 2.1 and analyze by using CARA software.
2. Disregard residues with chemical shifts that differ by >0.1 ppm between the lysate and purified protein.
3. Normalize peak intensities for the in-cell and lysate spectra using an amino proton peak from either a glutamine or an asparagine sidechain that does not change chemical shift between in-cell and lysate spectra.
4. Calculate changes in protein peak intensities, ΔI , due to quinary complexation as

$$\Delta I = (I_{\text{lysate}} - I_{\text{in-cell}}) / I_{\text{in-cell}}$$

where I_{lysate} and $I_{\text{in-cell}}$ are the normalized, integrated peak intensities from the lysate and in-cell spectra.

5. Plot ΔI versus protein residue (bar graph). Positive values correspond to peak broadening.
6. Draw a threshold line that distinguishes peaks with extreme broadening. Corresponding residues exceeding the threshold contribute to the quinary interaction surface.
7. Map the broadened residues onto the protein surface based on PDB file(s) using SWISS-PDB Viewer (or any other structure viewer).

3.3 Using SVD Analysis to Determine Protein-Protein Structural Interactions

All of the experiments are performed at RT using a standard 5 mm NMR tube in a 700 MHz Bruker Avance II NMR spectrometer equipped with TXI z-gradient cryoprobe. This protocol requires two expression plasmids that contain compatible origins of replication, different antibiotic resistance, and overexpress the target and interactor proteins from two distinct inducible promoters. NMR peak assignments for the target protein are also required to complete the analysis. We will use the prokaryotic Ubiquitin-like protein, Pup, as a specific example of the target protein, and mycobacterial proteasome ATPase, Mpa, from *Mycobacterium smegmatis*, *Msm*, as a specific example of the interactor protein.

Pup-GGQ, a precursor to Pup-GGE, was cloned into expression vector pASK3+ to yield pASK-Pup-GGQ [36]. pASK-Pup-GGQ contains an fl origin, confers ampicillin resistance, codes for Tet repressor, and expresses Pup-GGQ from the tet promoter/operator, which is induced by tetracycline or anhydrotetracycline. Mpa was cloned into expression vector pRSF-1b (Novagen) to yield pRSF-*Msm* Mpa [36]. pRSF-*Msm* Mpa contains an RSF origin, confers kanamycin resistance, codes for lac repressor and expresses

N-terminal His-tagged *Msm* Mpa from the T7 promoter/lac operon, which is induced by IPTG. *E. coli* strain BL21(DE3) codon + was co-transformed with pASK-Pup-GGQ and pRSF-*Msm* Mpa using standard techniques and the transformed cells were selected on LB-Ap-Kn agar plates.

Pup-GGQ does not engage in extensive quinary interactions, therefore unlike the previous protocols that utilized CRINEPT-HMQC-TROSY, Pup-GGQ is easily visualized in-cell by using $^1\text{H}\{^{15}\text{N}\}$ -HSQC.

3.3.1 Target Protein Labeling and Overexpression

1. Inoculate 50 mL of LB-Ap-Kn medium with a single colony of BL21(DE3) codon + containing pASK-Pup-GGQ and pRSF-*Msm* Mpa.
2. Grow the culture overnight at 37 °C with shaking at 250 rpm.
3. Centrifuge the cells at $4500 \times g$ for 15 min at room temperature.
4. Wash the cells once with 50 mL of M9 salts.
5. Resuspend the cells to $\text{OD}_{600} \sim 0.5$ in 500 mL of M9-Ap-Kn labeling.
6. Incubate the culture at 37 °C, 250 rpm for 15 min.
7. Induce overexpression of Pup-GGQ by adding 2 mg/mL of anhydrotetracycline in dimethylformamide to a final concentration of 0.2 $\mu\text{g}/\text{mL}$.
8. Allow the induction to proceed for 4 h at 37 °C and 250 rpm.
9. Centrifuge 100 mL of culture at $4500 \times g$ for 15 min at RT. Reserve the remaining culture for interactor overexpression (*see* Subheading 3.3.2).
10. Wash the cells twice with 50 mL of NMR buffer.
11. Resuspend the final cell pellet in 1 mL of NMR buffer containing 10% glycerol and store the sample at -80 °C for subsequent NMR analyses. This control sample provides the basis spectrum for the target protein in the absence of interactor protein.

3.3.2 Interactor Protein Overexpression

1. Centrifuge the remaining culture at $4500 \times g$ for 15 min at RT.
2. Wash the cells once with 500 mL of M9 salts.
3. Resuspend the cells to $\text{OD}_{600} \sim 0.5$ in 1000 mL of LB-Ap-Kn.
4. Incubate the culture at 37 °C, 250 rpm for 15 min.
5. Induce overexpression of *Msm* Mpa by adding 0.1% culture volume of 1 M IPTG. Allow the induction to proceed at 37 °C and 250 rpm.
6. Remove 100 mL of culture at ~ 4 h, 5 h, 6 h, 7 h, 8 h, and 18 h post-induction.

7. Centrifuge each sample at $4500 \times g$ for 15 min at RT.
8. Wash the cells twice with 50 mL of NMR buffer.
9. Resuspend the final cell pellet in 1 mL of NMR buffer containing 10% glycerol and store each sample at $-80\text{ }^{\circ}\text{C}$ for subsequent NMR analyses.

3.3.3 Sample Preparation

1. Thaw the sample and centrifuge at $4500 \times g$ for 15 min at RT.
2. Wash the cells once with 10 mL of NMR buffer.
3. Resuspend the cell pellet in 0.45 mL of NMR buffer and 0.05 mL of D_2O . Remove 10 μL for a cell viability assay. Transfer the remaining sample to a standard 5 mm NMR tube.
4. Prepare 10^{-4} , 10^{-5} , and 10^{-6} serial dilutions of the 10 μL cell sample. Plate 100 μL of each dilution onto LB-Ap-Kn plates in triplicate. Incubate the plates overnight at $37\text{ }^{\circ}\text{C}$ and count the resulting colonies.

3.3.4 NMR Spectroscopy

1. Shim the sample manually using the z, z2, and z3 shims.
2. Select the Watergate version of an $^1\text{H}\{^{15}\text{N}\}$ -edited HSQC experiment.
3. Set the number of transients to 32.
4. Set the spectral widths to 12 ppm and 30 ppm in the ^1H and ^{15}N dimensions respectively.
5. Set 512 and 64 points in the ^1H and ^{15}N dimensions, respectively, apodized with a squared cosine-bell window function and zero-filled to 1024 and 128 points, respectively, prior to Fourier transformation.
6. Collect the $^1\text{H}\{^{15}\text{N}\}$ -edited HSQC spectrum of the sample.
7. When data collection is completed, recover the sample. Remove a 10 μL aliquot for a cell viability assay. Prepare 10^{-4} , 10^{-5} , and 10^{-6} serial dilutions of the 10 μL cell sample. Plate 100 μL of each dilution onto LB-Ap-Kn plates in triplicate. Incubate the plates overnight at $37\text{ }^{\circ}\text{C}$ and count the resulting colonies.
8. Centrifuge the remaining sample at $4500 \times g$ for 10 min at RT.
9. Transfer the supernatant to a standard 5 mm NMR tube and acquire a $^1\text{H}\{^{15}\text{N}\}$ -edited HSQC spectrum to test for cell leakage.
10. Resuspend the cell pellet in 0.3 mL of NMR buffer.
11. Lyse the remaining cells by freeze-thawing ($-80\text{ }^{\circ}\text{C}$ -RT) five times.
12. Centrifuge the sample at $18,000 \times g$ for 15 min at RT.
13. Collect a $^1\text{H}\{^{15}\text{N}\}$ -edited HSQC spectrum on the clarified lysate using the optimized value for T_{opt} .

3.3.5 Data Processing and Analysis

This section describes how to perform an SVD analysis on a hypothetical file named ChemShiftSVD.txt. MATLAB commands are indicated in **bold Calibri font**.

1. Process data with Topspin 2.1 and analyze by using CARA software.
2. Calculate changes in protein peak intensities, ΔI , as

$$\Delta I = (I/I_{\text{ref}})_{\text{bound}} - (I/I_{\text{ref}})_{\text{free}}$$

where $(I/I_{\text{ref}})_{\text{free}}$ is the scaled intensity of individual peaks in the in-cell spectrum of free Pup-GGQ and $(I/I_{\text{ref}})_{\text{bound}}$ is the scaled intensity of individual peaks in the in-cell spectrum of the Pup-*Msm* Mpa complex, and I_{ref} is a glutamine peak at 7.45 ppm and 112.5 ppm in the proton and nitrogen dimensions respectively, that does not shift during titration.

3. Assemble the data matrix in Excel in order of residue numbers (*see Note 1*).
4. Save the data as an ASCII text file, ChemShiftSVD.txt.
5. Start MATLAB (*see Note 2*) and enter **clear all**.
6. Use the file browser to locate the ChemShiftSVD.txt file.
7. Enter **load ChemShiftSVD.txt** to load the data for analysis.
8. Enter **M = ChemShiftSVD;** to name the matrix.
9. Enter **[U,S,V] = svd(M)** to perform SVD analysis (*see Note 3*).
10. Enter **save ('S.txt', 'S','-ascii')** to save the S matrix.
11. Enter **B1 = S(1,1)*U(:,1)** to determine the first binding mode (*see Note 4*).
12. Enter **save ('B1.txt', 'B1','-ascii')** to save the B1 matrix.
13. Enter **B2 = S(2,2)*U(:,2)** to determine the second binding mode (*see Note 5*).
14. Enter **save ('B2.txt', 'B2','-ascii')** to save the B2 matrix.
15. Divide the square of the singular value of the first binding mode by the sum of the squares of all of the singular values (the square of the Frobenius norm of S) and multiply by 100 to compute the percentage of change represented in the first binding mode (*see Note 6*).
16. Plot the singular values (bar plot) against the binding mode number (dataset index) to generate a Scree plot.
17. Plot (bar plot) the first and the second binding modes against the residue numbers.
18. Identify interacting residues by comparing the difference in the magnitude of the first vs the second binding modes. The

highest value of the second binding mode is a threshold value that separates residues affected and not affected by binding.

19. Enter **clear all** before proceeding to analyze new data.
20. Map the interacting residues onto the protein surface based on PDB file(s) using SWISS-PDB Viewer or any other suitable structure viewer.

4 Notes

1. Enter either the chemical shift or the intensity changes for each residue measured at each time point in a single column. Do not enter residue numbers; i.e., a data set containing three chemical shift changes determined at three time points will have three columns of data in the matrix file.
2. Matlab 2009b or Octave 4.0.1 were used; however, later versions are acceptable.
3. Three main outputs are “U=”, left singular vectors, “S=”, singular values, and “V=”, right singular vectors. There are the same number of columns in the “S=” output as there are in the original data file. The singular values will appear as nonzero numbers on the diagonal of the matrix. E.g., for ChemShiftSVD, the “S=” output should appear as:

4.1582	0	0
0	0.3947	0
0	0	0.3023

where the first binding mode has a singular value of 4.1582, the second 0.3947, and the third 0.3023.

4. The larger absolute values represent greater contribution of the residue to the binding mode.
5. To calculate the values for other binding modes, use the command $\mathbf{B}\# = \mathbf{S}(\#, \#) * \mathbf{U}(:, \#)$ where # is the number of the binding mode. Do not exceed the dimensions of the matrix.
6. Using the hypothetical singular values for ChemShiftSVD, 98.59% of the changes occur in the first binding mode. Usually, true binding modes should represent more than 90% of the square of the Frobenius norm of S.

References

1. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P (2001) The protein-protein interaction map of helicobacter pylori. *Nature* 409 (6817):211–215

2. Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. *Nature* 422(6928):216–225
3. Gerstein M, Lan N, Jansen R (2002) Proteomics. Integrating interactomes. *Science* 295(5553):284–287
4. Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dotsch V (2001) High-resolution macromolecular NMR spectroscopy inside living cells. *J Am Chem Soc* 123(10):2446–2447
5. Goodsell D (2009) *The machinery of life*. Springer, New York
6. Persson E, Halle B (2008) Cell water dynamics on multiple time scales. *Proc Natl Acad Sci USA* 105(17):6266–6271
7. Benedek GB, Villars FMH (2000) *Physics: with illustrative examples from medicine and biology: electricity and magnetism*. Springer-Verlag, New York
8. Maldonado AY, Burz DS, Shekhtman A (2011) In-cell NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* 59(3):197–212
9. Li C, Wang GF, Wang Y, Creager-Allen R, Lutz EA, Scronce H, Slade KM, Ruf RA, Mehl RA, Pielak GJ (2010) Protein ¹⁹F NMR in *Escherichia coli*. *J Am Chem Soc* 132(1):321–327
10. Crowley PB, Chow E, Papkovskaia T (2011) Protein interactions in the *Escherichia coli* cytosol: an impediment to in-cell NMR spectroscopy. *ChemBioChem* 12(7):1043–1048
11. Majumder S, Xue J, DeMott CM, Reverdatto S, Burz DS, Shekhtman A (2015) Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* 54(17):2727–2738
12. Barbieri L, Luchinat E, Banci L (2015) Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep* 5:14456
13. Kyne C, Ruhle B, Gautier VW, Crowley PB (2015) Specific ion effects on macromolecular interactions in *Escherichia coli* extracts. *Protein Sci* 24:310–318
14. McConkey EH (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc Natl Acad Sci USA* 79(10):3236–3240
15. Wirth AJ, Gruebele M (2013) Quinary protein structure and the consequences of crowding in living cells: leaving the test-tube behind. *Bioessays* 35(11):984–993
16. Sarkar M, Smith AE, Pielak GJ (2013) Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci USA* 110(48):19342–19347
17. Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2007) *Protein NMR spectroscopy*. Academic Press, New York
18. Wider G (2005) NMR techniques used with very large biological macromolecules in solution. *Meth Enzymol* 394:382–398
19. Riek R, Wider G, Pervushin K, Wuthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci USA* 96(9):4918–4923
20. Shekhtman A, Ghose R, Goger M, Cowburn D (2002) NMR structure determination and investigation using a reduced proton (RED-PRO) labeling strategy for proteins. *FEBS Lett* 524(1–3):177–182
21. Ferrage F, Dutta K, Shekhtman A, Cowburn D (2010) Structural determination of biomolecular interfaces by nuclear magnetic resonance of proteins with reduced proton density. *J Biomol NMR* 47(1):41–54
22. Serber Z, Dotsch V (2001) In-cell NMR spectroscopy. *Biochemistry* 40(48):14317–14323
23. Banci L, Barbieri L, Bertini I, Luchinat E, Secci E, Zhao Y, Aricescu AR (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol* 9(5):297–299
24. Bertrand K, Reverdatto S, Burz DS, Zitomer R, Shekhtman A (2012) Structure of proteins in eukaryotic compartments. *J Am Chem Soc* 134(30):12798–12806
25. Hamatsu J, O'Donovan D, Tanaka T, Shirai T, Hourai Y, Mikawa T, Ikeya T, Mishima M, Boucher W, Smith BO, Laue ED, Shirakawa M, Ito Y (2013) High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc* 135(5):1688–1691
26. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 458(7234):106–109
27. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I (2009) Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, Streptolysin O. *J Am Chem Soc* 131(31):10834–10835
28. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis*

- egg extracts and intact oocytes. *Proc Natl Acad Sci USA* 103(32):11904–11909
29. Sakai T, Tochio H, Tenno T, Ito Y, Kokubo T, Hiroaki H, Shirakawa M (2006) In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. *J Biomol NMR* 36(3):179–188
 30. Theillet FX, Binolfi A, Bekei B, Martorana A, Rose HM, Stuiver M, Verzini S, Lorenz D, van Rossum M, Goldfarb D, Selenko P (2016) Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* 530(7588):45–50
 31. Majumder S, DeMott CM, Reverdatto S, Burz DS, Shekhtman A (2016) Total cellular RNA modulates protein activity. *Biochemistry* 55(32):4568–4573
 32. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat Methods* 3(2):91–93
 33. Burz DS, Shekhtman A (2010) The STINT-NMR method for studying in-cell protein-protein interactions. *Curr Protoc Protein Sci*. Chapter 17:Unit 17 11
 34. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) In-cell NMR for protein-protein interactions (STINT-NMR). *Nat Protoc* 1(1):146–152
 35. Golub GH, Van Loan CF (2012) Matrix computations, 4th edn. The Johns Hopkins University Press, Baltimore, USA
 36. Maldonado AY, Burz DS, Reverdatto S, Shekhtman A (2013) Fate of pup inside the mycobacterium proteasome studied by in-cell NMR. *PLoS One* 8(9):e74576
 37. Majumder S, DeMott CM, Burz DS, Shekhtman A (2014) Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *ChemBioChem* 15(7):929–933
 38. Wang T, Darwin KH, Li H (2010) Binding-induced folding of prokaryotic ubiquitin-like protein on the mycobacterium proteasomal ATPase targets substrates for degradation. *Nat Struct Mol Biol* 17(11):1352–1357

INDEX

A

Agonism 391
 Allostery and allosteric control 391
 Antagonism 391

B

Bacteriophages 8, 19, 67–95,
 139, 141
 Biomaterials 22, 23, 25

C

Chemical exchange 188, 207, 211, 219, 224,
 238, 393
 CHEMical Shift Covariance Analysis (CHESCA) 391
 Chemical shifts 3, 5, 7, 10, 13–15, 58–60,
 64, 83, 85, 88, 90, 91, 113–117, 126, 127, 163,
 165, 172, 173, 176, 178, 179, 207, 214, 224,
 230, 232–234, 237, 238, 247, 257–308, 364,
 365, 376, 377, 380, 382, 391–404, 424, 430,
 435, 438, 442
 Conformational dynamics
 millisecond-to-microsecond (μ s-ms) 206
 picosecond-to-nanosecond (ps-ns) 170, 173, 206
 Covariance 353–372, 391

D

Dark states 243
 Dynamic nuclear polarization (DNP)
 in solids 133–152, 156,
 158, 162, 165
 in solution 139, 142, 143, 145,
 147–149, 156–159, 161, 162, 164

E

Encounter complex 243, 244
 Ensemble refinement of SAXS (EROS) 376, 377,
 380, 382–387

G

G protein-coupled receptors (GPCRs) 407–419

H

High-resolution relaxometry 169–198
 Hybrid methods 375
 Hyperpolarization 155–158, 160, 161, 163, 164

I

Interactions
 protein–DNA 70, 83, 89
 protein–protein 386, 438–442
 Intransient interactions 427
 Intrinsically disordered proteins (IDPs) 5, 19, 23,
 173, 356, 375–382, 386
 Isotopic labeling 20, 68, 251

L

Lowly populated states 243

M

Magic-angle spinning (MAS) 7, 9, 11, 13–17,
 19–24, 37–51, 56–59, 62, 95, 135, 137, 139,
 141, 145, 147, 151, 152
 Maximum entropy reconstruction 9, 350
 Membrane curvature 99, 101
 Membrane proteins 5, 8, 19, 21, 22, 37, 39,
 40, 45, 55, 57, 59, 64, 134, 139, 141, 145
 Multidrug resistance 57
 Multiple acquisitions 51
 Multiple receiver detection 61

N

NMR_{Fx} 308
 NMRViewJ 308
 Non-uniform sampling (NUS) 8, 9, 13, 23,
 56, 262, 341–351
 Nuclear magnetic resonance (NMR)
 in cell 2, 55, 134, 160, 362, 378, 408
 in solids 3, 37, 55, 101, 111, 133,
 169, 348, 408
 in solution 3, 38, 56, 102, 139,
 156, 169, 223, 243, 299, 342, 363, 375, 411

Nuclear spin relaxation	
Carr Purcell Meiboom Gill (CPMG).....	163, 210, 223–239
model-free analysis of.....	173, 174, 188, 197, 200
relaxation dispersion	206, 211, 217, 236, 237
relaxation in the rotating frame ($R_{1\rho}$).....	219
P	
Protein dynamics. <i>See</i> Conformational dynamics	
Proton detection in the solid-state	11
Q	
Quadrupolar nuclei.....	2, 9, 15, 24
Quinary structure.....	424, 426, 430
R	
Radicals	15, 133, 135, 136, 138, 139, 146, 147, 149, 151, 156–159, 161, 248, 251, 253
Real-time NMR.....	160, 163
Resonance assignment	9–11, 21, 38, 85, 141, 365
Rosetta	68, 69, 88, 90, 91
S	
Sensitivity enhancement	8, 10, 51, 155–165, 349
Singular value decomposition (SVD)	395, 401, 402, 404, 428–430, 438–442
Site-directed spin-labeling	380
Small angle X-ray scattering (SAXS)	250, 375–383, 386, 387
Spectral	
analysis	40, 111, 127, 258, 344
derivative.....	369
prediction.....	127
reconstruction	341, 347
Spherical nanoparticle supported lipid bilayers (SSLBs).....	99–108
Spin correlations.....	164
Structure determination	7–9, 11, 13, 14, 19, 21–23, 38, 95, 239
T	
Transient interactions	253, 424
Transverse relaxation optimized spectroscopy (TROSY).....	10, 11, 15, 212, 214, 229, 236, 247, 424, 425, 428, 434, 435, 437, 439
U	
Ultrahigh magnetic fields	3–18, 23
X	
XPLOR-NIH	246, 249, 250