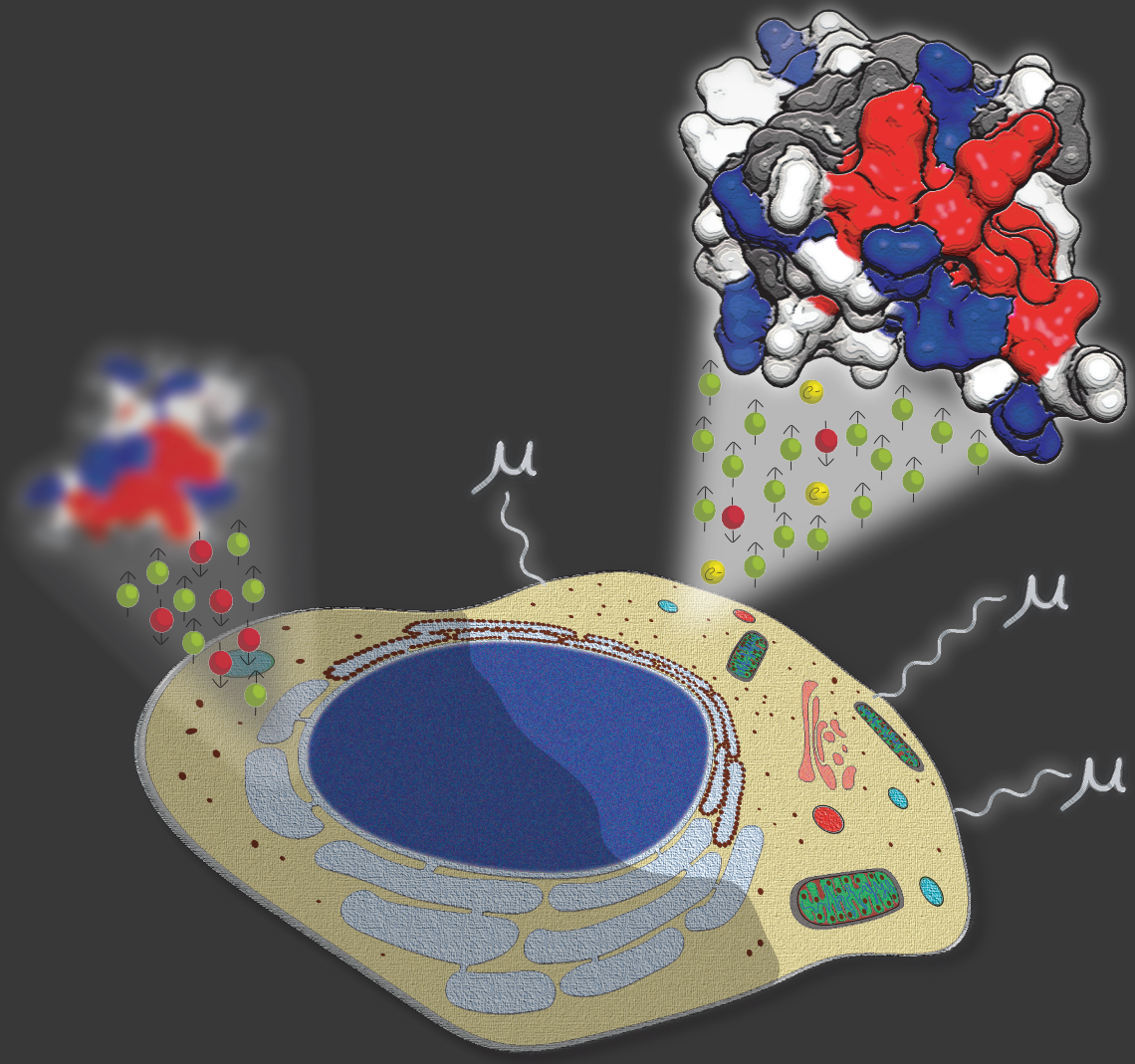
Sensitivity Enhanced Solid-State NMR of Soluble Proteins in Cells



Siddarth Narasimhan

Sensitivity Enhanced Solid-State NMR of Soluble Proteins in Cells

Siddarth Narasimhan

ISBN 978-94-6380-801-9

Doctoral Thesis
Sensitivity Enhanced Solid-State NMR of Soluble Proteins in Cells

Siddarth Narasimhan
NMR Spectroscopy Research Group,
Bijvoet Center for Biomolecular Research
Department of Chemistry,
Utrecht University, The Netherlands
May 2020

Cover and lay-out design: Siddarth Narasimhan
Printed in the Netherlands by Proefschriftmaken

Sensitivity Enhanced Solid-State NMR of Soluble Proteins in Cells

Vaste-stof NMR met verhoogde gevoeligheid van oplosbare eiwitten in cellen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
woensdag 13 mei 2020 des middags te 2.30 uur

door

Siddarth Narasimhan

geboren op 27 augustus 1992
te Haiderabad, India

Promotor:
Prof. dr. M. Baldus

Copromotor:
Dr. G.E. Folkers

Beoordelingscommissie:

Prof. dr. C.R. Berkers
Prof. dr. P.C.A. Bruijnincx
Prof. dr. V. Dötsch
Prof. dr. A.P.M. Kentgens
Prof. dr. E. Mastrobattista
Prof. dr. J.G. Roelfes
Prof. dr. H.A.B. Wösten

This work was supported by funding from NWO, for
the Netherlands' Magnetic Resonance Research School
(NMARRS) (022.005.029)

Table of Contents

Chapter 1:
General Introduction

Second part of this chapter has been adapted and used for the following review article:
"*When small becomes too big: Expanding the use of in-cell solid-state NMR spectroscopy*"
Siddarth Narasimhan, Gert E. Folkers, Marc Baldus
DOI: 10.1002/cplu.202000167 ChemPlusChem. *in press*

Part 1- Biomolecular Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is a non-invasive spectroscopic technique, developed in the mid-1940's[1,2] whose early applications were limited to chemical characterization and determination of structures of small molecules. It exploits an inherent property of nuclei, known as spin, inherent to nuclei with a non-zero spin quantum number. Nuclei with a non-zero spin, possess a magnetic moment (μ), which is determined by the gyromagnetic ratio (γ) and the spin angular momentum (I) (eq. 1). In the presence of an external magnetic field, nuclear spins can experience a torque, leading to precession at a frequency known as the Larmor frequency (ω) which depends on the external magnetic field strength (B_o) and the gyromagnetic ratio (γ) (eq. 2).

$$\vec{\mu} = \gamma\vec{I} \quad (1)$$

$$\omega = \gamma B_0 \quad (2)$$

The nuclei can orient themselves in different orientations with respect to the magnetic field, manifested by the Zeeman splitting. Nuclei with a spin quantum number ½ (diamagnetic), which are commonly encountered in biomolecular NMR can exist in two states (spin up and down). The energy difference of the two levels is given as a multiple of the Planck's constant ($\hbar$), the gyromagnetic ratio (γ) and the external magnetic field (B_o) (eq. 3). This difference in the energy is also proportional to the Larmor frequency given in eq. 2.

$$\Delta E = \gamma\hbar B_0 = \hbar\omega \quad (3)$$

At a given temperature (T), the ratio of the population between the two spin states can be determined by eq. 4 where k is the Boltzmann constant.

$$N^+/N^- = e^{\left(\frac{-h\nu}{kT}\right)} \quad (4)$$

In NMR, a resonance condition is achieved by matching the energy difference between the states. Assuming that the magnetization exists in a 3-dimensional space, at equilibrium, in an external magnetic field oriented along the Z-axis, the net magnetization is oriented in the Z-direction (Figure 1A). When a radio frequency (RF) pulse at the Larmor frequency of a certain length (duration) is applied, the net magnetization can be rotated away from the equilibrium, i.e., the Z-direction.

Depending on the RF pulse phase and the duration, the net magnetization begins to evolve in the XY- plane (Figure 1B). During this process, the nutation of the nuclear spins induces an electric field in the NMR coil, which can be recorded. In the

absence of an NMR pulse, this magnetization eventually returns to equilibrium (Z), through a phenomenon called relaxation (Figure 1C). This decaying magnetization is defined as Free Induction Decay (FID), which is the output of an NMR experiment (Figure 1D). To interpret the FID, it is digitally processed and converted to frequency units by applying a Fourier Transformation to obtain the spectrum. The resulting NMR spectrum contains resonance frequencies observed in the sample, which provide information on the chemical environment surrounding the nuclei. The resonant frequencies (in Hz) obtained in the NMR measurements are normalized to chemical shift units (δ) given in parts per million (ppm) with respect to the corresponding Larmor Frequency of the nuclei in each magnetic field (in MHz) (see eq. 5).

$$\delta_{(\text{in ppm})} = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}\,(\text{in MHz})} \quad (5)$$

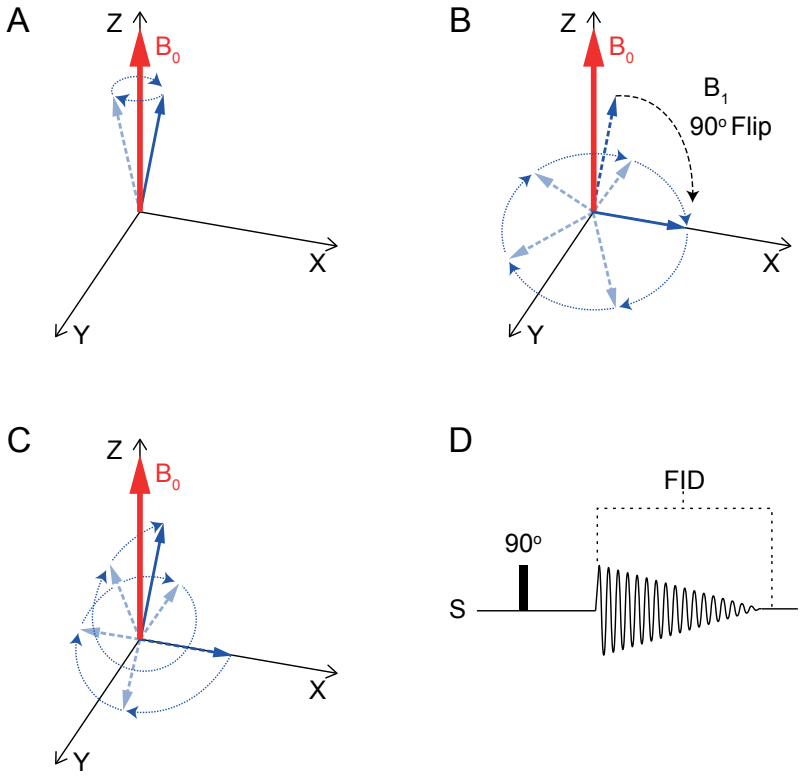


Figure 1: In the presence of an external magnetic field, the net magnetization is aligned along the Z- direction (A). Upon a 90° pulse using an RF pulse of an amplitude B_1 at Larmor frequency, the magnetization flips and evolves in the XY plane (B). The magnetization eventually "relaxes" back to into equilibrium (C) as depicted in panel A. The process is summarized in panel D, where the pulse is given, and the FID is recorded.

1.1 Sensitivity and resolution

The quality of an NMR spectrum is determined by two factors, i.e., spectroscopic sensitivity and resolution. As with any other biophysical technique, sufficient signal needs to be obtained to analyze the spectra. The sensitivity of the signal in the NMR experiment is determined by the difference in the population between the spin up and down states. In practical terms, this difference is influenced by the amount of sample, the temperature (T), the gyromagnetic ratio (γ) of the nucleus under detection, and the strength of the external magnetic field (B_o) that affects the Larmor frequency (ν) (see also eq 4). The gyromagnetic ratio is a critical factor for the NMR sensitivity. Therefore, the detection of ^{1}H spins is the most obvious choice for biomolecular NMR spectroscopists due to its high gyromagnetic ratio, the near- 100% isotopic abundance and the proton density in biomolecules, in particular, proteins (Table 1).

Table 1: Gyromagnetic ratios (γ) of commonly observed nuclei in biomolecular NMR expressed in 10^6. rad.s^{-1}.T^{-1} and their isotopic natural abundance.

Nucleus	γ	Natural Abund.
^{1}H	267.52	~99.9%
^{13}C	67.28	~1.1%
^{15}N	-27.11	~0.4%
^{31}P	108.29	100%

The spectral resolution of NMR spectra of biomolecules depends on a range of factors such as, the number of distinguishable chemical species in the sample, relaxation properties of the spins (discussed in the following section), and finally the homogeneity, stability & strength of the external magnetic field (B_0). Improving NMR resolution is an active field of research for both manufacturers and users of the NMR spectrometers. A powerful means to increase spectral resolution is simply by conducting experiments in different spectral dimensions (a.k.a. multidimensional NMR). In such experiments, the NMR signals of different nuclei are, for example, correlated via through-space or through-bond interactions in several spectral dimensions. Due to a high proton density in biomolecules, it is almost always necessary to conduct heteronuclear multidimensional ^{1}H NMR experiments. The most obvious nuclei chosen for heteronuclear biomolecular NMR spectroscopy are ^{13}C and ^{15}N which have a very low natural abundance (Table 1). Proteins are thus often enriched with ^{13}C and/or ^{15}N spins, which not only speeds up NMR data acquisition but also eventually enables complete resonance assignment and 3D structure determination. Adding spectral dimensions generally leads to longer experimental times. Hence obtaining optimal NMR data is often a compromise between adequate NMR sensitivity and resolution.

1.2 Relaxation in NMR

A full description of mechanisms underlying relaxation in NMR is beyond the scope of this thesis. Instead, we will focus here on the influence of relaxation properties on the NMR sensitivity and resolution. Furthermore, factors affecting the relaxation rates are discussed. When net magnetization is brought to the transverse plane, relaxation to equilibrium follows an evolution period as depicted in Figure 1C. Relaxation to equilibrium (Z) occurs in both the transverse (described by the transverse relaxation time T_2) and the longitudinal (denoted by T_1) planes simultaneously, with usually T_1 relaxation being slower than T_2.

In a typical NMR experiment, multiple scans are added to increase the signal

to noise ratio. To consecutively record NMR experiments, a short time delay must be given for the magnetization to relax back to the equilibrium and to ensure that the hardware is not affected due to heat generated while pulsing. This waiting period determines the so-called duty cycle. Since the T_1 relaxation rate is the slower than T_2, T_1 determines how quickly NMR experiments (scans) can be repeated. T_2 relaxation rates on the other hand are strongly correlated with the linewidths of the acquired signal, thereby affecting the resolution of the signals observed. Both relaxation rates are sensitive to molecular motions and the fluctuations in the magnetic field (both local and external). In their landmark work of 1948, Bloembergen *et al.* observed that both T_1 and T_2 relaxation rates for a molecule in solution are inversely related to the global molecular motions given by the correlation coefficient (τ_c) of the tumbling molecules[3]. While T_2 relaxation rate scales inversely with the molecular tumbling rate, the T_1 relaxation rate is fastest for intermediate τ_c which is approximately equal to the inverse of the Larmor frequency and increases with larger molecular weight.

Just as in the field of Magnetic Resonance Imaging (MRI), studying relaxation patterns can be beneficial in biomolecular NMR too. For example, relaxation rates for particular spins such as amide protons or the $^{13}C_\alpha$ carbons in proteins can also be used to calculate their internal dynamics[4] in both solution- and solid-state NMR. Additionally, the presence of paramagnetic effect (that are caused by electron spin fluctuations which affect nearby NMR-active nuclei) can influence relaxation rates causing broadening (increasing T_2 rate) and by decreasing experimental times (increasing T_1 rate) as well as by inducing so called paramagnetic shifts. In fact, paramagnetic relaxation enhancement NMR (PRE-NMR) is widely used to obtain information on protein structure, dynamics and long-range distances[5,6]. In the chapter 5 of this thesis, paramagnetic effects are observed in NMR experiments that probe the assembly of a metalloenzyme in *E. Coli* cells.

1.3 Magic Angle Spinning Solid-State NMR (MAS- ssNMR)

Biomolecular NMR initially focused on applications that probe molecules freely tumbling in solution. As mentioned before in this chapter, NMR signals are acquired during longitudinal and transverse (T_1 and T_2) relaxation processes. If molecules are present in a viscous environment or tumble slowly due to their size, transverse relaxation times (T_2) decrease, leading to enhanced NMR line broadening and thereby complicating the spectral analysis. Therefore, solid biological samples such as fibrils or membrane proteins, cannot be measured using solution NMR approaches. The lack of any tumbling in solids also leads to anisotropic chemical shift interactions referred to as chemical shift anisotropy (CSA) that reflects the anisotropic charge distribution around the nuclei leading to chemical shift inhomogeneity. NMR

spectra of solids are the superposition of these interactions in a randomly oriented sample, and thus they tend to be broad and featureless. As depicted in Figure 2, CSA makes the spectra complex to interpret as a single nuclear species would give rise to multiple chemical shifts, due to the different orientations of the molecules.

There are four major interactions seen in diamagnetic spins when placed in an external magnetic field (eq. 6). These are characterized as Zeeman interaction (H_z), the Chemical Shielding (H_{CS}), the J- Coupling (H_J) and the Dipolar (H_D) interaction (eq. 6). By far, the strongest contribution is that of the Zeeman component, which depends on the external magnetic field (B_o). The other components are the so called "internal factors" inherent to the nuclei and their molecular environment under consideration, where the chemical shielding (CS) component describes the chemical shift. And additionally, J- and Dipolar couplings mediated through chemical bonds and space respectively, describe interactions between nuclei.

$$H = H_z + H_{CS} + H_J + H_D \quad (6)$$

J- couplings, also known as scalar couplings, between nuclei can be probed in NMR spectra, provided that they are up to four bonds away from each other. J- couplings are usually 1-2 orders of magnitude weaker than dipolar couplings. Dipolar couplings are very important to consider, as they directly influence spin-spin relaxation (T_2) and can lead to homogeneous line broadening. The dipolar coupling term (H_D) can be further expanded as follows:

$$D_{ij} = \frac{\mu_0}{4\pi}\frac{\gamma_i\gamma_j\hbar}{r^2}\frac{1}{2}(3\cos^2\Theta_{i,j} - 1) \quad (7)$$

where r is the distance between the spins i & j, and $\Theta_{i,j}$ the angle at which the dipolar vector connecting the two spins is oriented with respect to the external magnetic field (B_o). In a molecule freely tumbling in the solution, the effective dipolar coupling averages out to zero. This is due to the fast time dependence of $\Theta_{i,j}$ (and in the case of dynamic molecules: *r)* values that would be sampled within the timescale of the NMR experiment. This effect does not occur in static or solid samples or in samples which exhibit a high viscosity.

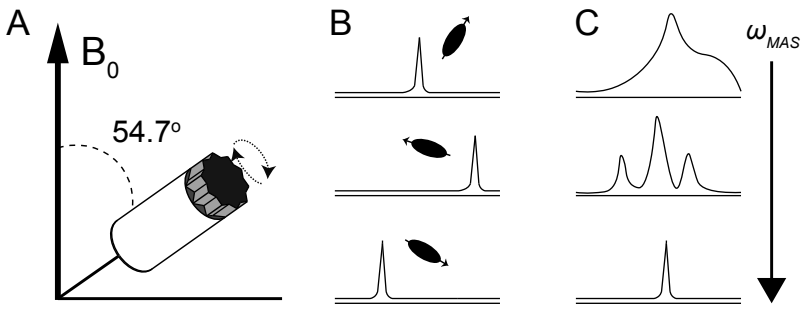


Figure 2: A) A description of magic angle spinning (MAS), where the rotor is spun at high frequencies, at the magic angle (54.7°) with respect to the external magnetic field B_0. B) Different orientations of the spins with respect to B_0 lead to different chemical shifts in the NMR spectrum C) Top: typical "powder" pattern seen in solids, when there is no MAS. As the MAS frequency is increased above a certain threshold, an isotropic peak can be obtained (bottom).

At a specific angle (~54.7° with respect to the B_0), the dipolar coupling term amounts to zero (eq. 7). This orientation is known as the "magic angle", where the effect of dipolar couplings is minimized. By simply placing the solid sample in a magnetic field at the magic angle, one would not fully eliminate the dipolar couplings to obtain a resolved spectrum. Therefore, physical spinning of the sample helps to mimic the random tumbling seen for molecules in a solution. This leads to a better resolution in the spectra as the dipolar couplings begin to average out (Figure 2). Furthermore, magic angle spinning (MAS) would also lead to averaging out of the CSAs (Figure 2) due to reduced anisotropic interactions[7]. In general, the spinning frequencies must be higher than the strength of the dipolar couplings (calculated in frequency units). Hence, this condition depends on the nuclei of choice, as dipolar couplings also scale with the gyromagnetic ratios (γ) of the interacting nuclei as well as their internuclear distance (r) (eq. 7). For this reason, dipolar couplings of nearby protons are generally the strongest among nuclei in biomolecular solid-state NMR. Note that in an anisotropic solution, dipolar couplings are only partially averaged out leading to the so called residual dipolar couplings (RDCs).

Considering the above, to average out dipolar couplings corresponding to ^{1}H-^{13}C that are directly bonded, MAS should be >20 kHz while one bond ^{13}C-^{13}C can be averaged out by MAS of >2 kHz. MAS is practically achieved by using special cylindrical ssNMR sample containers known as rotors (Figure 2A) which are made of strong materials like Zirconia or Sapphire. Because mechanical spinning would not be feasible to achieve very high spinning speeds, the rotors are spun using a combination of bearing and drive pressures that drive rotor caps via an air turbine

mechanism. The rotors need to be made adequately small to ensure high spinning rates. Notably smaller rotors (fast/ ultra-fast MAS rotors) hold smaller sample volumes, which can impact ssNMR sensitivity.

1.4 Sensitivity enhancement in MAS-ssNMR

Despite its potential to study large biomolecules, ssNMR method development has been devoted for the last two decades to simultaneously improve NMR sensitivity and resolution. Summarizing the previous section, higher resolution in solids can be achieved using higher spinning speeds or higher field magnets. In the following, we discuss three complementary strategies to further increase spectroscopic sensitivity, i.e., Cross-Polarization, ^{1}H detection and Dynamic Nuclear Polarization.

1.4.1 Cross-Polarization (CP): Unlike in solution-state NMR, ^{1}Hs were not the most obvious nuclei to be detected in ssNMR since ^{1}H-^{1}H dipolar couplings are among the strongest among nuclei, leading to broad lines and thus to conditions that are far from high-resolution ssNMR spectroscopy. For biomolecular applications, the most sensitive nucleus after ^{1}H is ^{13}C, whose strongest (^{1}H-^{13}C) dipolar couplings can be averaged out by merely spinning at speeds of ~20 kHz. ^{13}C detection however can have other limitations than low sensitivity such as long T_1 relaxation times even at ambient temperature. ^{13}C detection is hence often preceded by a cross-polarization (CP) step[8], where the polarization of ^{1}H is first transferred to ^{13}C, via dipolar couplings between ^{1}H and ^{13}C spins. Transfer is possible when the Hartmann-Hahn matching condition[9] is met by adjusting RF pulses on the ^{1}H and ^{13}C nuclei as indicated in (eq. 8 where ω_H and ω_C are the RF pulses applied on ^{1}H and ^{13}C channels respectively while ω_{MAS} is the MAS frequency in Hz and n is an arbitrary integer).

$$\omega_H = \omega_C \pm n.\omega_{MAS} \quad (8)$$

In the actual pulse program (Figure 3), the two channels are spin-locked at the matching condition, to enable CP.

The CP technique was a breakthrough in the field of ssNMR, as it could reduce experimental times by an order of magnitude.

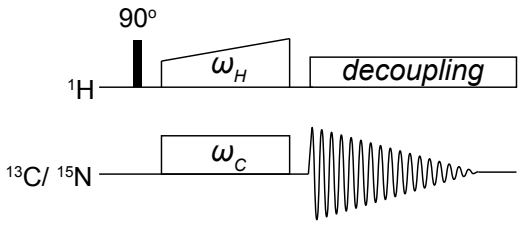


Figure 3: A typical pulse program for cross polarization (CP) in solids where ω_H and ω_C are matched to an integral multiple of the MAS frequency (ω_{MAS}). ω_H is usually applied as a ramped pulse, to ensure that a range of frequencies in the ^{1}H spins are excited during CP.

1.4.2 ^{1}H Detection: ^{1}H detected NMR spectroscopy is the norm in solution-state NMR, since ^{1}H nuclei are not only the most sensitive nucleus available, but they are also sensitive to the change in the surrounding environment. As mentioned above, directly detecting ^{1}H nuclei in solids has been a great challenge, as biomolecules generally contain a large amount of ^{1}H spins and due to the lack of motion, the dipolar couplings are dominating. As many of these spins are close in space, it is impossible to average out all the dipolar couplings by merely spinning at ~20 kHz. Partial/selective deuteration of the biomolecules has been shown to be help in partially obliterating dipolar couplings for small proteins at low spinning speeds[10]. However, this would not be beneficial for proteins that suffer from decreased spectral dispersion owing to their large sizes. MAS-rotors used for ^{1}H detected ssNMR thus need to ideally be able to spin at very high frequencies[11,12]. They are generally smaller in size, holding less sample volume. The decrease in the sample volume is often outweighed by the sensitivity gain by ^{1}H detection. Currently 1.3- and 0.7-mm rotors are very popular due to the ultra-high MAS frequencies (~40 kHz to ~130 kHz) they allow for. It is very common these days to achieve proton linewidths below < 0.1 ppm in solid-state NMR using proton detection[13,14]. Combined with other probehead development efforts, ^{1}H detection is emerging as the norm in biomolecular ssNMR technique at ambient temperatures.

1.4.3 (Cross-Effect) Dynamic Nuclear Polarization (DNP): DNP is a recent, and a revolutionary addition to NMR and MRI with particular benefits for the biomolecular ssNMR world. DNP exploits the fact that unpaired electron spins have the highest gyromagnetic ratio (600 times that of ^{1}H) of all the diamagnetic species. Therefore, when properly exploited, electron polarization can offer the highest signal-to-noise enhancements in NMR experiments. For this purpose, unpaired electrons are polarized using microwave irradiation, which is then transferred to the nuclei by one of the three modes: Overhauser, solid or cross-effect DNP. Although, each of these processes have their own merits, only cross-effect DNP is described here and

any further references to DNP in this thesis is synonymous to cross-effect DNP.

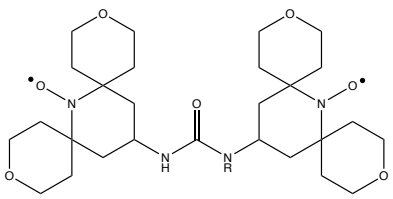
To establish cross-effect DNP, bi-radicals containing two unpaired electrons at an appropriate distance are used. The distance within the biradical can be modified by using different "chemical handles" that harbor the electrons. Efficient cross-effect occurs under MAS conditions, when the sum or the difference between the Larmor frequencies of the two electrons (ω_{e1} & ω_{e2}) matches that of the Larmor frequency of the nucleus (ω_N) to which it is coupled via electron-nuclear hyperfine interactions (eq. 9).

$$\omega_N = \omega_{e1} \pm \omega_{e2} \quad (9)$$

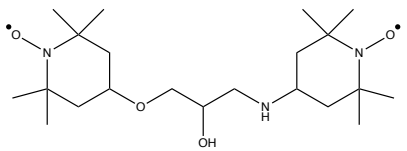
1.5 Biomolecular DNP-ssNMR

DNP for biomolecular applications would not have been possible without the conception of water-soluble biradicals. Several such radicals like AMUPol and PyPol[15] that are used in the chapters of this thesis (Figure 4), have been developed for the purpose of biomolecular DNP-ssNMR. Due to the very short T_1 relaxation times of electrons at ambient temperatures (~293K), the experiments are performed at very low temperatures (100-110K) for efficient transfer between the electrons and nuclei to occur.

In practical terms, a microwave radiation source in the form of a gyrotron magnet is connected to the probe of the spectrometer via a wave guide. The microwaves are then directly focused on the cold sapphire rotor, which is transparent to microwaves, and holds a glass matrix composed of the molecule(s) of interest and the biradicals (Figure 5A). A DNP setup and a typical pulse program for a CP (as depicted in Figure 3) is shown in Figure 5.



AMUPol: R = $(CH_2CH_2O)_4Me$
PyPol: R = H



TOTAPOL

Figure 4: Examples of water-soluble biradicals used in cross-effect DNP, including AMUPol and PyPol which are used in the work described in the following chapters.

Due to the low temperatures and the strong electron-nuclear hyperfine couplings, the NMR linewidths observed under low-temperature DNP conditions are usually broader than seen at ambient temperatures, leading to reduced spectral

resolution. The radicals shown in Figure 4, provide best signal enhancement are relatively low B_o fields (i.e., 400-600 MHz). Therefore, any ssNMR spectra acquired under such conditions are usually of multidimensional character. Current efforts in DNP method development focus on increasing spectral resolution and DNP performance at ultra-high magnetic fields, as this seems to be the only bottleneck before the technique is applied on a larger scale. For example, on the chemical synthesis end, water-soluble trityl-nitroxide based biradicals are currently being developed which are tuned to function at higher B_o fields, along with showing superior DNP performance at higher MAS rates at cryogenic temperatures. Secondly, on the DNP hardware end, electron decoupling is being developed, which would reduce the hyperfine couplings during the acquisition period. Lastly, DNP methods that perform also at higher (i.e. ambient) temperatures are currently being pursued.

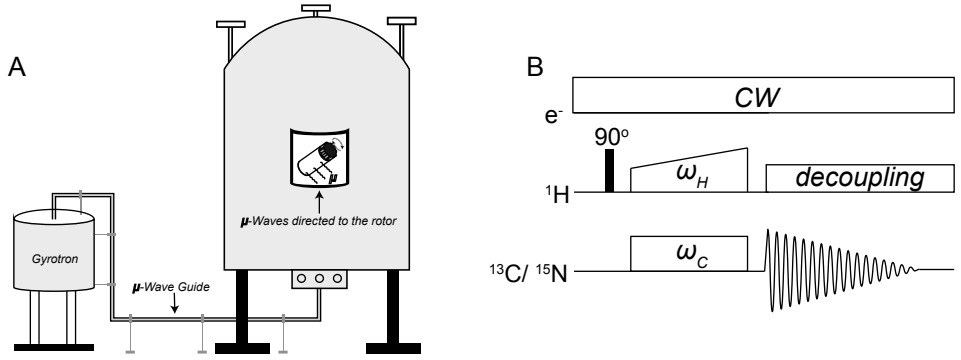


Figure 5: A) DNP-ssNMR setup where a microwave source (gyrotron) is coupled to a ssNMR probe via a wave guide. B) Typical CP pulse program in a DNP-ssNMR experiment.

Part-2 Cellular Structural Studies Using NMR

An increasing trend in biology has been to merge the fields of cell and structural biology. Structural analysis through decades has largely neglected the cellular context, due the lack of methods available that would provide atomistic details in the cellular environment. Robust approaches were not developed until recently when structural biologists began to realize the need to validate their in-vitro findings with correlative studies on the biomolecule of interest in native environments. Such efforts are motivated by the notion that the structure solved in-vitro, in controlled buffer conditions *may* not always reflect the true state of the protein in its native environment. The native environment thus remained the biggest "blind spot" in structural biology.

The benefits of conducting "*cellular structural biology*" has been fully realized in the last two decades. The following sections outline the historical development of structural studies that has led us to systematically eliminate the proverbial "blind spots" poised by in-vitro structural studies. Although several approaches have their own merits in cellular structural studies, the major focus in the following sections will be on what the NMR community contributed to tackling these challenges. In particular, the review presented in the following sections discusses the premise for the work presented in this thesis which attempts to further develop and to demonstrate the use of DNP enhanced solid-state NMR for biomolecular studies in cells for soluble proteins.

Evolution of NMR in structural biology

The emergence of the field of structural biology in the early stages can be almost entirely accredited to X-ray crystallography. Parallel developments in recombinant DNA (rDNA) technology and advanced purification techniques made it easier to obtain pure protein samples. The late 90's thus saw a strong increase in the number of crystal structures deposited in PDB solved using crystallography (Figure 6) leading to a vast amount of reliable high-resolution structures every year. Yet, crystallization can still be a very strenuous process which can inexplicably fail to produce high-quality and stable crystals. There are also limitations in terms of the types of proteins compatible with crystallization. For example, very small (< 5kDa) or very large proteins and complexes, as well as insoluble and highly dynamic proteins remain difficult or even impossible to crystallize. Some of these proteins, which include membrane or membrane associated proteins, fibrils and intrinsically disordered proteins are very potent drug candidates. Membrane proteins often play the gate-keeping role, mediating signal transduction as well as molecular transport across cell membrane, while intrinsically disordered proteins can adopt

different structures at variable timescales, mediating different cellular processes. In conclusion, a significant set of potent and druggable targets could not be easily studied using X-ray crystallography and continues to be a challenge till date[16].

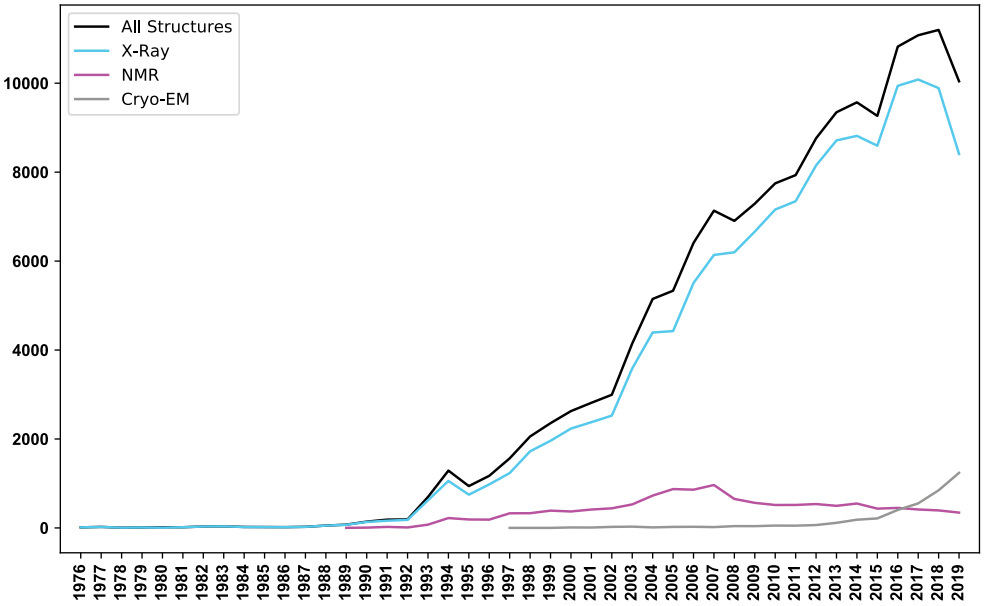


Figure 6: Overview of the number of structures from different sources deposited every year in the protein data bank (PDB).

Along with the widespread use of rDNA technology, another groundbreaking phenomenon in the early 90's was the emergence and the development of biomolecular solution-state Nuclear Magnetic Resonance (NMR). At this point and as discussed before, NMR as a technique had existed for almost half a century [1,2]. But the development of high-field and multidimensional NMR brought about a revolution for applications in chemistry and, in particular, biochemistry in this period. NMR developed into a reliable means to study small, dynamic and intrinsically disordered proteins (IDPs), and to some extent membrane proteins solubilized in organic solvents or detergent micelles. Owing to the availability of whole genome sequences, biomolecular solution-state NMR and X-ray crystallography jointly led us into the structural genomics revolution with the final goal to determine the structure of every biomolecule in the genome. The major paradigm shift that NMR brought about was the ability to study the dynamics of the proteins[4], which contribute significantly to mediating molecular recognition and interactions. The temporal dimension added to the structures allowed for real-time kinetic studies of biomolecular interactions,

small molecule binding, or the detection of time-dependent changes in the protein structure. This is particularly important as many proteins and biomolecules are dynamic and can exist in a multitude of conformations rather than a single state[17]. For example, IDPs[18] often sample a multitude of secondary structures that can be directly probed by NMR[19].

After complete resonance assignment and protein structure determination, NMR data can be funneled into rapid high throughput drug screening and design. In fact, completely automated structure-based drug screening can be performed on a library of drugs, where a wealth of structural information can be compiled and collated almost effortlessly[20,21]. However, the biggest bottleneck of solution-state NMR, similar to X-ray crystallography is that it is limited by molecular size of about 30 kDa (see Part 1). This can be partially circumvented, by focusing on just the methyl groups which undergo fast local motion, largely irrespective of protein size[22]. However, this strategy would fail entirely for membrane proteins, in their physiologically relevant lipid environments and for insoluble fibrillar proteins. By early 2000's, there were still these two classes of very important proteins that remained (almost) completely elusive to high-resolution structural studies by NMR. Therefore, solid-state NMR emerged as a complementary method to solution-state NMR.

Biomolecular solid-state NMR

ssNMR, unlike solution-state NMR is not theoretically limited by size (see section 1.3). Since the early-mid 2000's, ssNMR combined with MAS (≤ 20 kHz as mentioned in the previous sections) has offered an increasingly powerful alternative to study large and/or insoluble proteins that are inaccessible to solution-state NMR or crystallography. In the beginning, it suffered from sensitivity and resolution related issues, both of which is being progressively circumvented by the development of high magnetic fields, DNP & ^{1}H detected ssNMR methods combined with ultra-fast MAS (60-180 kHz). Three distinct types of biomolecules for which ssNMR is uniquely suited for are described below:

1) Fibrils: Protein misfolding is related to many diseases including Alzheimer's or Parkinson's disease. In contrast to expectations, misfolded proteins can sometimes assemble from disordered monomers into higher order aggregates[23–25]. Often, they are known to be highly insoluble, as the hydrophobic amino acids tend to point outward from the core of the protein. There are other types of insoluble peptides that can form pathogenic aggregates known as amyloids, which are found in mammals. ssNMR has provided information on the fibrillar structure, including insights into the organization of individual units up to the atomic level[26–28]. In addition, structures

of fibrils formed by larger proteins such as alpha-synuclein[29] and tau K19[30], have been studied using ssNMR.

2) Membrane proteins: Structures of a few solubilized membrane proteins have been solved using crystallography or solution-state NMR. These structures have been acquired in a membrane mimetic environment using detergents or organic solvents, and thus are not always relevant to the protein's physiological state. Membrane proteins function differently in different lipid environments, and certain types of lipids can be indispensable to their function. Purification and reconstitution in model membranes has enabled structure determination of several membrane proteins at atomic resolution. ssNMR has greatly contributed to conducting structural studies on membrane proteins in their functional lipid environments[13,31–37].

3) Large protein machines: Despite the size limitations, solving structures of very large proteins or protein machinery is complicated by the higher spectral crowding. Nevertheless, ssNMR has provided very valuable structural insights into large protein machineries. For example, segmental labelling approaches in combination with intricate sample preparation methods have allowed for studies on large beta-barrel proteins including complexes like the beta-barrel assembly machinery[38–40] (BAM) and on secretion systems such as the type III and IV secretion systems (T3SS, T4SS)[41].

Cryo-EM and the resolution revolution

Cryo-electron microscopy (Cryo-EM), is the most recent entrant into structural biology. It evolved from electron microscopy, which was widely used to obtain high-resolution images of cells and organelles albeit, not at atomic level. The so called "*resolution-revolution*[42]" that made cryo-EM popular among structural biologists, occurred in the current decade owing to innovative sample freezing techniques[43,44], better sample holding grids[45], improved detectors and perhaps most importantly the development of better data interpretation and analysis methods[46]. It has evolved to be truly complementary to NMR and X-ray crystallography to solve structures of large molecules and systems which were impossible in the past[47,48]. In contrast to other techniques, the number of structures being deposited in the PDB solved using cryo-EM is on the rise every year. While cryo-EM has no upper limit, it can be difficult to apply it to small proteins[49].

Adding the cellular context to proteins

Despite the rise of Cryo-EM, the number of structures being solved every year is now approaching a plateau (Figure 6). A simple explanation to this is that there are only a finite number of important and unique structures that are of biological relevance. In fact, as per the statistics collected by the CATH database[50],

no new unique topologies have been deposited in the PDB since the year 2012. In the proverbial race to solve all biomolecular structures in isolation, the biological relevance of these structures in their native cellular contexts is often overlooked, or, until recently ignored. A future where the whole cell can be studied at an atomic resolution using structural methods is perhaps a *utopian* idea. There are however recent indications[51–53] that we are progressing in this direction using combinations of in-vitro and in-vivo hybrid approaches. This goal can never be achieved by a single approach, as every technique has its own niche as well as limitations. As of today, a combined and correlative approach to cellular structural studies seems to be the only feasible way forward. For instance, cryo-EM tomography is currently being honed to obtain better resolution in complex environments such as on an organelle scale or even in the whole cell. Efforts in this direction has undoubtedly provided new insights into biomolecular organization and functioning in cells. However, atomistic information has remained elusive and is often supplemented by simply adding the high-resolution structural information obtained in-vitro[51]. In parallel, missing information can be supplemented using advanced light microscopy methods via correlative light-electron microscopy (CLEM). As live-cell imaging is being made more accessible, CLEM is often performed with live cell imaging to obtain dynamic information on molecular processes that is lost due to the freezing[54].

Although NMR can never be used to acquire the information that cryo-EM or CLEM is currently able to provide[54], it can play a decisive role in providing atomistic details in complex cellular settings. The ability to focus on a single molecule (and a partner) that has been made spectroscopically distinguishable in a highly complex environment is something that other techniques cannot readily offer. The emphasis with in-cell NMR studies is on the effect of cellular components on molecule itself, rather than the cellular milieu that is found in. While cryo-EM or CLEM can visualize the over-all "picture", molecular structures of known molecules can be added from in-vitro results, and NMR could be used to refine the high-resolution details of the molecule as a result of the environment it is placed in. Finally, there can be significant contributions from several other techniques such as electron paramagnetic resonance[55,56] (EPR) or fluorescence resonance energy transfer[57] (FRET) that offer specific distance related information in cellular studies.

in-situ NMR spectroscopy 60's to 80's

During the early days of crystallography, in which structures of several biomolecules were yet to be solved in-vitro, structural biologists often devoted their lives to solving a single structure. Detailing these efforts would be out of the scope of this review, but some of these challenges include, obtaining reliable

crystals, solving the phase problem and data processing with limited computational resources[58]. After the advent of biomolecular NMR spectroscopy, solving structures without crystallization was already seen as an innovative leap towards probing a more functional state of macromolecules. The logical, albeit non-trivial next phase in biomolecular NMR then was to enable spectral acquisition or even the ability of solving molecular structures within the cellular environment, denoted as "*in-cell* NMR".

The ability to solve complete structures of most biologically relevant proteins in cells is indeed a difficult endeavor to pursue. Structural biologists, and in particular, NMR spectroscopists are still significantly far away from achieving this goal for most proteins. Clearly, cryo-EM based approaches that have emerged in the last couple of years will play a much significant role in this respect, provided that the molecules are large enough. Incidentally, in 2001, Serber et al., the early pioneers of in-cell NMR, who arguably even coined the term, state the following in their landmark contribution[59]:

> "*The power of in-cell NMR spectroscopy lies not in determining structures de novo, but in observing changes in the structures of biological macromolecules and in their interaction with other cellular components.*"

This holds true to date in 2019, despite several groups having attempted structure determination with limited success except for proteins that are small, and preferably inert in the cellular environment[60,61].

Prior to early 90's, structure determination was never the only goal of most NMR spectroscopists in the context of biological applications. As it will be highlighted below and in Figure 7, they were much rather interested in uncovering the (bio) chemical underpinnings of biological processes in-situ, by exploiting the non-invasive nature of NMR. This holds especially true magnetic resonance imaging (MRI), which employs magnetic resonance for visualizing complex systems such as vertebrates and plants non-invasively. In a similar spirit, early in-situ NMR (not excluding in-cell) was focused on acquiring very precise chemical information at the atomic scale that complemented imaging at millimeter resolution. The first in-situ experiments, including in intact cells in fact predate the conception of multidimensional NMR. As early as 1967, the lack of mobility of Na+ ions were confirmed in eukaryotic tissue samples, which was scientifically disputed beforehand[62].

The following decade (1970's) saw increased application of NMR metabolomic studies, which were elegantly designed to trace metabolic processes over time. Given the proton density in biological samples, and the non-existence of

multidimensional experiments, ^{31}P 1D- NMR was used for detection as phosphorus was much less abundant in cells as compared to ^{1}H. ^{31}P is an obvious choice as it is isotopically 100% abundant and many phosphorus containing biomolecules display unique ^{31}P chemical shifts. ^{31}P NMR was, as a consequence used in studying a range of complex samples such as bacterial cells[63], yeast cells[64], tumor cells[65,66], cultured eukaryotic cells[66], rat organs[63] and to track metabolic changes during the development of *Xenopus laevis* embryos[67]. Next to ^{31}P, ^{13}C emerged as another nuclei of interest, where isotope enriched glucose and glycerol were introduced and tracked inside the cells[63]. Similar approaches have evolved over time and continue to be in use now, as a recent study was reported where fully automated methods were used to track metabolism in real time in viable patient samples over long periods of time[68].

In the late 70's through 80's, a series of three landmark innovations were made in the in-situ NMR studies of complex systems:

1) Bioreactor in NMR: The development of a continuous flow reactor which allowed for tracking metabolism in live cells upon application of stress and stimuli[69,70]. This concept has been improved further by other groups, to an extent that it is now possible to purchase a continuous flow setup compatible with a probe of choice, along with software support for analysis from Bruker.

2)Multidimensional NMR: Development of multidimensional NMR[71] experiments single handedly set precedence for structural biology, enabling protein structure (& dynamics) determination[72,73]. Pulse schemes relevant for such studies include Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Correlation (HSQC), Transverse-Relaxation Optimized Spectroscopy (TROSY) for probing large proteins, Nuclear Overhauser Effect Spectroscopy (NOESY) etc.. Truly ground braking high-resolution in-cell/ in-situ NMR experiments which followed much later would not have been possible without the development of such multidimensional NMR Methods.

3) MAS-ssNMR in Biomolecules: CP enabled MAS-ssNMR that was used to improve resolution and sensitivity in solids, found applications in metabolic studies on immobile parts of the cells such as membranes and/or the peptidoglycan layer[74].

in-cell solution-state NMR in structural biology (21st century)

Should one attempt to carry out multidimensional NMR studies in cells, the inherently low sensitivity would require longer experimental acquisition times, which would make it impossible to capture rapid biochemical changes. The most exciting challenges for NMR spectroscopists during 90's lay in determining high-resolution structures of proteins (Figure 6) due to the advent of NMR in structural genomics studies. The in-vitro structures determined then, later went on to serve as references

for *cellular* studies. An early reported NMR study in the cellular environment was in 1996, where an approach was designed to check the integrity of recombinantly overexpressed proteins prior to arduous purification on cell lysates[75]. Later, in the 2000's, the prospect of structural studies on biomolecules in-cell began to emerge in a real sense, for proteins in whole cells. The intricate isotope labelling schemes that were developed to obtain complete assignment and hence the structure of the proteins, were extended in innovative ways to in-cell structural biology. The greatest challenges that were to be overcome were:

1) **Signal Ambiguity:** Reducing signals from the "cellular background" by ensuring that the protein of interest is the most abundant species in the sample, or by targeted labelling.

2) **Low Sensitivity:** Solving the inherent low sensitivity due to the dilution brought about by the cellular background. Biomolecules exist in dynamic states in their native milieu, which calls for faster experimental times to avoid degradation and rapid changes leading to uninterpretable spectra. As in-cell NMR gained popularity, there was an increasing need to decrease experimental times. Subsequent developments in NMR pulse programs, in particular, the SOFAST-HMQC sequence[76] offered an exciting prospect in recording multidimensional experiments, within the lifetimes of cells[77].

3) **Reduced Tumbling:** In solutions, reduced molecular tumbling in a crowded cell leads to acquisition of no signal or poorly resolved spectra owing to faster T_2 relaxation. Hence, only small- to medium sized protein targets (<20 kDa) were often chosen, as highlighted below.

In 2001, the first 2D *in-cell* NMR experiment in bacteria using solution NMR of a metal binding protein, NmerA and the changes following metal introduction were reported[78]. Following this demonstration, the same research group detailed all the parameters that must be taken into account to carry out in-cell solution-state NMR for a protein of interest[79] through overexpression in bacterial cells. As eukaryotic systems tend to be more complex, protein overexpression was not initially employed. Rather, the isotope labelled protein was delivered into unlabelled cells to ensure complete obliteration of background labelling. This approach was for the first time demonstrated on frog eggs (*Xenopus laevis* oocytes), as they were large enough to be subjected to microinjection[59,80].

As most eukaryotic cells are usually much smaller than frog eggs, protein delivery through microinjection would not be feasible. Therefore, other means such as pore forming toxins[81] (eg. SLO), protein fusions containing cell penetrating peptides[82] (CPP) and electroporation were explored[56,83]. In their first demonstration of multi-dimensional NMR in HeLa cells harboring exogenously delivered isotope-

enriched proteins, Inomata *et al.* showed that using a cell penetrating peptide had no deleterious effect on the cells[82]. This observation largely holds true for pore forming toxins such as SLO when used in moderate amounts and if the pores are repaired in time through Ca^{2+} treatment[81]. Using pore-forming toxins offered the added advantage that no protein modification was necessary for their delivery into cells. Electroporation mediated protein delivery came into prominence much later[56,83] which also did not require the protein to be altered in any manner. It was initially used to transfect cells with DNA, and it offers the most robust approach as long as the cells can be 2D cultured. By 2D culturing the cells subjected to electroporation, live cells are specifically selected based on their ability to re-attach firmly to the cell culture surface. The advantage of this technique is that one can somewhat control the amount of protein delivered into the cells, thereby increasing experimental reproducibility. Complementary to electroporation, mammalian expression systems evolved over time allowing for direct protein overexpression as done in bacterial systems. This approach has been demonstrated on the metalloprotein superoxide dismutase (SOD) whose maturation process and metal binding properties have been studied using in-cell NMR[84,85]. Note that such studies on folding/ maturation is not possible when folded isotope labelled proteins are delivered into the cells.

Despite being small, the highly promiscuous protein Ubiquitin (8.5 kDa), could not be detected by Inomata *et al*, without mutating its sites of molecular recognition[82]. This highlights the need for the protein of interest to be relatively inert in the cells for solution-NMR studies. Examples of proteins that adhere to such criteria are SOD (see above) or the protein alpha-synuclein (α-syn) which can exist in a multitude of states- ranging from disordered monomer to a membrane associated state or as an aggregated multimer. Next to structural insights, such in-cell NMR studies have also provided novel and unexpected insights into largely unexplored cellular processes such as post-translational modifications (PTMs). This was first observed in the form of phosphorylation in *Xenopus laevis* oocytes[86]. Later on, the N-terminal acetylation of exogenously delivered proteins thymosin β4[81] and α-syn[56] was detected. N-terminal acetylation of exogenous proteins is particularly interesting as it has been reported to occur purely by co-translational means[87], which is disproved by two separate in-cell NMR studies. Due to its ability to study PTMs, NMR shows potential to uncover labile PTMs which are difficult to study in-vitro.

In-cell solution-state NMR continues to hold on to its robust base, in studying small/ medium sized soluble proteins if they tumble well enough in cells. The molecular size limitation and solubility requirement, which is specific for solution-state NMR have triggered research into the potential use of biomolecular MAS-ssNMR for cellular studies. For in-cell measurements, MAS-ssNMR is difficult

to use for all cell types (eg: 2D cultured mammalian cells), due to the high G-force present during sample spinning. Such restrictions do not apply to membrane proteins, leading to the development of *cellular* or in-situ ssNMR which is described in the following section.

NMR on membrane proteins in their native environment

The primary goal behind conducting in-situ studies on membrane proteins is to ensure that the investigations are carried out in the native lipid environments. Since the amount of membranous parts in the cell is only a fraction of the total amount of the whole cell present, isolating intact membranes proved to be beneficial for NMR sensitivity by avoiding the "dilution" problem. Using intact membranes instead of cells not only improves sensitivity and signal specificity, but also allows one to use very high MAS frequencies without risking cell lysis, which is often closely followed by degradation via intracellular proteases. Obtaining insoluble fractions from the cell is rather trivial in theory, and only entails mechanical or osmotic cell lysis followed by pelleting of the insoluble material by centrifuging at the appropriate G-force. The caveat for obtaining biologically relevant samples was that the membrane(s) must be largely intact, and its composition remain unaltered. Therefore, different approaches were often used for different organisms and membranes.

1) **From small to large proteins:** Early progress in membrane proteins were made in bacterial membranes including the two landmark innovations were made by Renault *et al.*, in 2011 and 2012. They first managed to successfully implement DNP in *Cellular* ssNMR studies of membrane proteins, which enabled them to measure spectra of whole cells enriched with a membrane protein of their interest[38,88]. Further on, they showed that *E. coli* outer membrane proteins OmpA & OmpF which are present in high abundance are not vital for cell viability. Therefore, eliminating these proteins via mutation produced a cleaner spectrum, devoid of much background[88]. In 2015, it was furthermore demonstrated that such concepts can be applied to large proteins. Kaplan *et al.* studied a 1MDa periplasm spanning complex by combining sample preparation approaches from Renault *et al.* with amino acid selective isotope labelling to reduce spectral crowding[89]. In 2016, the approach was extended to eukaryotic systems. This research involved studying the dynamic changes brought about by a ligand-mediated activation followed by autophosphorylation of a large eukaryotic membrane protein (150 kDa), epidermal growth factor receptor (EGFR) that was overexpressed in the native membrane environment of a cancer cell line[90].

2) **Complete obliteration of background labelling:** In 2015, Baker *et al.*, demonstrated targeted labelling of a membrane protein of their interest, by timed expression using non-native RNA polymerases and the subsequent suppression of

native *E. coli* RNA polymerases[91]. In the following year, this approach was applied to the studies of a potassium channel protein KcsA, with high sensitivity & resolution yielded by ^{1}H detection at very high MAS frequencies[13]. Targeted labelling is almost mandatory for ^{1}H detected methods as high sensitivity combined with high proton density in biological samples would lead to the enhancement of sparsely labelled cellular background. Additionally, by fractional[92] or inverse fractional deuteration[13] and amino-acid selective labelling approaches[40], it is possible to focus specific protein regions using ^{1}H detected ssNMR.

In conclusion, the ssNMR methods to probe membrane proteins in native membranes and cells is highly evolved in terms of sample preparation and makes use of currently available cutting-edge NMR approaches. On the other hand, very large membrane protein assemblies can be easily detected in the EM tomograms of whole cells and cell envelopes. Increasingly, EM-based studies are delivering in-cell structures of mostly very rigid parts of the large protein machinery[51] at a progressively higher resolution. In the spirit of cellular structural biology, combination with ^{1}H detected motion filtered experiments, should make it possible to model the flexible parts of protein machinery irrespective of the size. Thus, these complementary approaches which are already emerging in literature[93], hold very bright prospects for our future understanding of membrane proteins.

A timeline of landmark developments in in-situ NMR and its evolution is shown in Figure 7, which is summarized in the above sections.

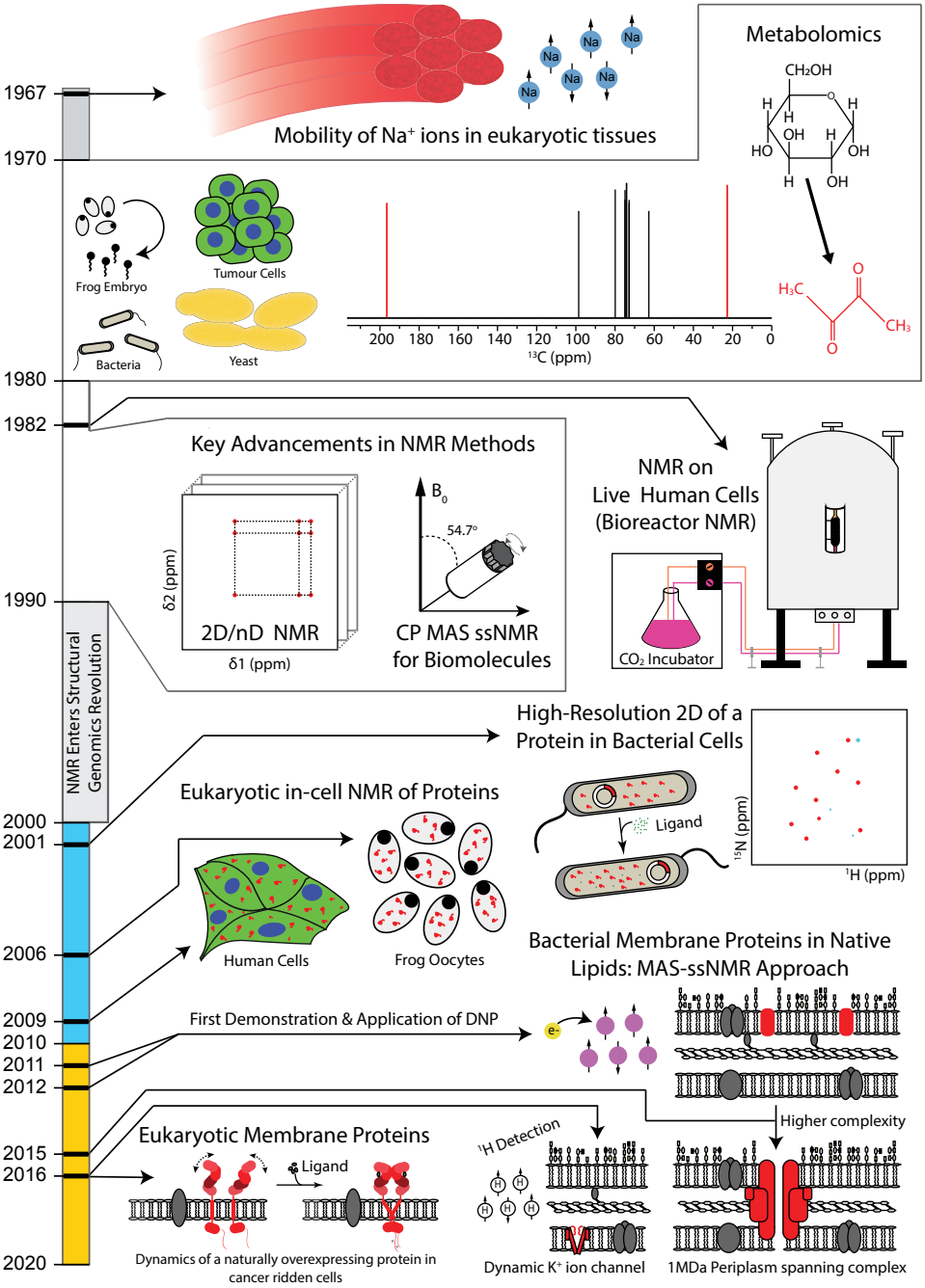


Figure 7: Evolution of NMR spectroscopy for *cellular/* in-cell and in-situ applications, leading up to now.

Scope for ssNMR on select soluble proteins

If one is interested in observing the behavior of a given protein in its native habitat, there are enough approaches available at one's disposal. For a small and soluble well tumbling protein, solution NMR may be used while cryo-EM tomography is cut out for very large and stable assemblies/machinery. Membrane proteins can be tackled using ssNMR approaches or cryo-EM, depending on the information that one wishes to seek and the size of the system where the former deals with smaller systems. This leaves us with one unexplored territory: small protein units or peptides that either assemble into larger assemblies in often an inhomogeneous manner or interact strongly with one or many binding partners. For instance, hub proteins have been described in interactomes, that form central links to other proteins by being involved in key regulatory processes. Reduced or lack of tumbling upon complex formation would hinder detection by solution-NMR. If the proteins are small or even disordered, it renders them undetectable by single particle EM methods. Therefore, small promiscuous proteins have continued to be elusive to in-cell structural studies.

The aim of this thesis to attempt to include such candidates for in-cell structural studies. An example of such a protein is Ubiquitin, which is inherently small (8.5 kDa) and thus making it easily detectable. But as it has been highlighted before, due to its ability to be conjugated to almost any protein in the cell, only an inert population of wild-type Ubiquitin is detectable in solution-NMR[82]. Attempting to study whole cells with MAS ssNMR not only poses the problem of significant G-force on eukaryotic and some prokaryotic cells, but one would also be confronted with low sensitivity due to smaller sample volumes containing NMR active nuclei. Both issues can be addressed by employing low-temperature DNP in ssNMR. As discussed in section 1.5, cross-effect DNP experiments in ssNMR are performed at <110K, where the samples are frozen and hence the effect of G-force is negligible on the sample in the rotor being spun. This approach has been successfully implemented on a eukaryotic and a prokaryotic system, as will be elucidated in the following chapters of this thesis.

Scope of this thesis

This chapter (**Chapter 1**) provides a general introduction which is divided into two parts. The first part deals with the various aspects of biomolecular solution- and solid-state NMR and related DNP approaches. The second part contains a historic overview of cellular structural biology, the role of NMR, and finally outlines the scope of the work presented in the thesis.

Chapter 2 describes a protocol to prepare molecule specific isotope labelled samples for small proteins in 2D cultured human cells for the purpose of in-cell DNP supported multidimensional NMR experiments.

Chapter 3 contains first results from studies on Ubiquitin populations in HeLa cells upon proteasomal inhibition. The studies presented are based on a similar protocol described in chapter 1, thereby highlighting a potential application area for the protocol.

Chapter 4 describes a protocol to achieve targeted molecule specific isotope labelling of soluble proteins in *E. coli* cells for DNP supported multidimensional ssNMR experiments. With a goal to improve DNP sensitivity and completely obliterate signal ambiguity, background deuteration has been implemented.

In **Chapter 5**, the assembly of an artificial metalloenzyme in cell is observed using DNP-ssNMR by applying the approach described in chapter 4. Furthermore, solution-state NMR analysis on the cell lysate provides useful insights into the protein dynamics.

Chapter 6 deals with the step-by-step guide to using an in-house computational tool which can be used to predict spectra upon providing assignments, mainly targeted for ssNMR approaches. Some of its features are useful in obtaining rational isotope labelling strategies on the protein of interest.

The findings in the thesis are summarized, and future directions are discussed in **Chapter 7**.

References

1. Bloch, F. Nuclear Induction. *Phys. Rev.* **70**, 460–474 (1946).
2. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**, 37–38 (1946).
3. Bloembergen, N., Purcell, E. M. & Pound, R. V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **73**, 679–712 (1948).
4. Ishima, R. & Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **7**, 740–743 (2000).
5. Hocking, H. G., Zangger, K. & Madl, T. Solution PRE NMR. in *Protein NMR: Modern Techniques and Biomedical Applications* 133–157 (Springer, Boston, MA, 2015). doi:10.1007/978-1-4899-7621-5_4
6. Knight, M. J. *et al.* Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11095–11100 (2012).
7. Andrew, E. R., Bradbury, A. & Eades, R. G. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* **182**, 1659 (1958).
8. Pines, A., Gibby, M. G. & Waugh, J. S. Proton-Enhanced Nuclear Induction Spectroscopy. A Method for High Resolution NMR of Dilute Spins in Solids. *J. Chem. Phys.* **56**, 1776–1777 (1972).
9. Hartmann, S. R. & Hahn, E. L. Nuclear Double Resonance in the Rotating Frame. *Phys. Rev.* **128**, 2042–2053 (1962).
10. Zhou, D. H., Graesser, D. T., Franks, W. T. & Rienstra, C. M. Sensitivity and resolution in proton solid-state NMR at intermediate deuteration levels: Quantitative linewidth characterization and applications to correlation spectroscopy. *J. Magn. Reson.* **178**, 297–307 (2006).
11. Ishii, Y. & Tycko, R. Sensitivity Enhancement in Solid State 15N NMR by Indirect Detection with High-Speed Magic Angle Spinning. *J. Magn. Reson.* **142**, 199–204 (2000).
12. Ishii, Y., Yesinowski, J. P. & Tycko, R. Sensitivity enhancement in solid-state 13C NMR of synthetic polymers and biopolymers by 1H NMR detection with high-speed magic angle spinning. *Journal of the American Chemical Society* **123**, 2921–2922 (2001).
13. Medeiros-Silva, J. *et al.* 1 H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angew. Chem. Int. Ed. Engl.* **55**, 13606–13610 (2016).
14. Lecoq, L. *et al.* 100 kHz MAS Proton-Detected NMR Spectroscopy of Hepatitis B Virus Capsids. *Front. Mol. Biosci.* **6**, 58 (2019).
15. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chemie Int. Ed.* **52**, 10858–10861 (2013).
16. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **70**, 2–20 (2014).
17. Bax, A. & Clore, G. M. Protein NMR: Boundless opportunities. *J. Magn. Reson.* **306**, 187–191 (2019).
18. Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **21**, 432–440 (2011).
19. Camilloni, C., De Simone, A., Vranken, W. F. & Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **51**, 2224–2231 (2012).
20. Pellecchia, M., Sem, D. S. & Wüthrich, K. Nmr in drug discovery. *Nat. Rev. Drug Discov.* **1**, 211–219 (2002).
21. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192 (2009).

22. Schütz, S. & Sprangers, R. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in Nuclear Magnetic Resonance Spectroscopy* 56–84 (2019). doi:10.1016/j.pnmrs.2019.09.004
23. Fändrich, M., Fletcher, M. A. & Dobson, C. M. Amyloid fibrils from muscle myoglobin. *Nature* **410**, 165–166 (2001).
24. Frost, B., Ollesch, J., Wille, H. & Diamond, M. I. Conformational diversity of wild-type Tau fibrils specified by templated conformation change. *J. Biol. Chem.* **284**, 3546–51 (2009).
25. Strohäker, T. *et al.* Structural heterogeneity of α-synuclein fibrils amplified from patient brain extracts. *Nat. Commun.* **10**, 5535 (2019).
26. Paravastu, A. K., Leapman, R. D., Yau, W.-M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's -amyloid fibrils. *Proc. Natl. Acad. Sci.* **105**, 18349–18354 (2008).
27. Elkins, M. R. *et al.* Structural Polymorphism of Alzheimer's β-Amyloid Fibrils as Controlled by an E22 Switch: A Solid-State NMR Study. *J. Am. Chem. Soc.* **138**, 9840–9852 (2016).
28. Gelenter, M. D. *et al.* The peptide hormone glucagon forms amyloid fibrils with two coexisting β-strand conformations. *Nat. Struct. Mol. Biol.* **26**, 592–598 (2019).
29. Heise, H. *et al.* Molecular-level secondary structure, polymorphism, and dynamics of full-length α-synuclein fibrils studied by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15871–15876 (2005).
30. Andronesi, O. C. *et al.* Characterization of Alzheimer's-like paired helical filaments from the core domain of tau protein using solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **130**, 5922–5928 (2008).
31. Weingarth, M. *et al.* Supramolecular Structure of Membrane-Associated Polypeptides by Combining Solid-State NMR and Molecular Dynamics Simulations. *Biophys. J.* **103**, 29–37 (2012).
32. Ladizhansky, V. Applications of solid-state NMR to membrane proteins. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1577–1586 (2017).
33. Mandala, V. S., Williams, J. K. & Hong, M. Structure and Dynamics of Membrane Proteins from Solid-State NMR. *Annu. Rev. Biophys.* **47**, 201–222 (2018).
34. Cordero-Morales, J. F. *et al.* Molecular determinants of gating at the potassium-channel selectivity filter. *Nat. Struct. Mol. Biol.* **13**, 311–318 (2006).
35. van der Cruijsen, E. A. W. *et al.* Biomolecular DNP-Supported NMR Spectroscopy using Site-Directed Spin Labeling. *Chem. - A Eur. J.* **21**, 12971–12977 (2015).
36. Lange, A. *et al.* Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. *Nature* **440**, 959–962 (2006).
37. Ader, C. *et al.* A structural link between inactivation and block of a K+ channel. *Nat. Struct. Mol. Biol.* **15**, 605–12 (2008).
38. Renault, M., Bos, M. P., Tommassen, J. & Baldus, M. Solid-State NMR on a Large Multidomain Integral Membrane Protein: The Outer Membrane Protein Assembly Factor BamA. *J. Am. Chem. Soc.* **133**, 4175–4177 (2011).
39. Sinnige, T. *et al.* Insight into the conformational stability of membrane-embedded BamA using a combined solution and solid-state NMR approach. *J. Biomol. NMR* **61**, 321–332 (2015).
40. Sinnige, T., Daniëls, M., Baldus, M. & Weingarth, M. Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J. Am. Chem. Soc.* **136**, 4452–5 (2014).
41. Loquet, A. *et al.* Atomic model of the type III secretion system needle. *Nature* **486**, 276–9 (2012).

42. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
43. Thompson, R. F., Walker, M., Siebert, C. A. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15 (2016).
44. Passmore, L. A. & Russo, C. J. Specimen Preparation for High-Resolution Cryo-EM. in *Methods in enzymology* **579**, 51–86 (2016).
45. Russo, C. J. & Passmore, L. A. Ultrastable gold substrates for electron cryomicroscopy. *Science (80-.).* **346**, 1377–1380 (2014).
46. Baker, M. Cryo-electron microscopy shapes up. *Nature* **561**, 565–567 (2018).
47. Lin, Z. *et al.* TTC5 mediates autoregulation of tubulin via mRNA degradation. *Science (80-.).* **367**, 100–104 (2020).
48. Takizawa, Y. *et al.* Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. *Structure* **28**, 44-53.e4 (2020).
49. Kühlbrandt, W. Cryo-EM enters a new era. *eLife* **3**, e03678 (2014).
50. Dawson, N. L. *et al.* CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* **45**, D289–D295 (2017).
51. Bykov, Y. S. *et al.* The structure of the COPI coat determined within the cell. *Elife* **6**, (2017).
52. Baker, L. A. *et al.* Combined 1 H-Detected Solid-State NMR Spectroscopy and Electron Cryotomography to Study Membrane Proteins across Resolutions in Native Environments. *Structure* **26**, 161-170.e3 (2018).
53. Toro-Nahuelpan, M. *et al.* Tailoring cryo-electron microscopy grids by photo-micropatterning for in-cell structural studies. *Nat. Methods* **17**, 50–54 (2020).
54. de Boer, P., Hoogenboom, J. P. & Giepmans, B. N. G. Correlated light and electron microscopy: ultrastructure lights up! *Nat. Methods* **12**, 503–513 (2015).
55. Yang, Y. *et al.* A Reactive, Rigid Gd ^{III} Labeling Tag for In-Cell EPR Distance Measurements in Proteins. *Angew. Chemie Int. Ed.* **56**, 2914–2918 (2017).
56. Theillet, F.-X. X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).
57. Dimura, M. *et al.* Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **40**, 163–185 (2016).
58. Smyth, M. S. & Martin, J. H. J. x Ray crystallography. *Journal of Clinical Pathology - Molecular Pathology* **53**, 8–14 (2000).
59. Serber, Z. *et al.* Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protoc.* **1**, 2701–9 (2006).
60. Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105 (2009).
61. Tanaka, T. *et al.* High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells. *Angew. Chemie Int. Ed.* **58**, 7284–7288 (2019).
62. Cope, F. W. NMR Evidence for Complexing of Na+ in Muscle, Kidney, and Brain, and by Actomyosin. The Relation of Cellular Complexing of Na+ to Water Structure and to Transport Kinetics. *J. Gen. Physiol.* **50**, 1353–1375 (1967).
63. Shulman, R. G. *et al.* Cellular applications of 31P and 13C nuclear magnetic resonance. *Science* **205**, 160–166 (1979).
64. Salhany, J. M., Yamane, T., Shulman, R. G. & Ogawa, S. High resolution 31P nuclear magnetic resonance studies of intact yeast cells. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 4966–70 (1975).
65. Navon, G., Ogawa, S., Shulman, R. G. & Yamane, T. 31P nuclear magnetic resonance studies of Ehrlich ascites tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 87–91 (1977).

66. Navon, G., Navon, R., Shulman, R. G. & Yamane, T. Phosphate metabolites in lymphoid, Friend erythroleukemia, and HeLa cells observed by high-resolution 31P nuclear magnetic resonance. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 891–5 (1978).
67. Colman, A. & Gadian, D. G. 31P Nuclear-Magnetic-Resonance Studies on the Developing Embryos of Xenopus laevis. *Eur. J. Biochem.* **61**, 387–396 (1976).
68. Alshamleh, I. *et al.* Real-Time NMR Spectroscopy for Studying Metabolism. *Angew. Chemie - Int. Ed.* **59**, 2304–2308 (2020).
69. Gonzalez-Mendez, R., Wemmer, D., Hahn, G., Wade-Jardetzky, N. & Jardetzky, O. Continuous-flow NMR culture system for mammalian cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **720**, 274–280 (1982).
70. Foxall, D. L. & Cohen, J. S. NMR studies of perfused cells. *J. Magn. Reson.* **52**, 346–349 (1983).
71. Nagayama, K. Two-dimensional NMR spectroscopy: an application to the study of flexibility of protein molecules. *Adv. Biophys.* **14**, 139–204 (1981).
72. Oschkinat, H. *et al.* Three-dimensional NMR spectroscopy of a protein in solution. *Nature* **332**, 374–376 (1988).
73. Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. M. Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **29**, 1689–1696 (1990).
74. Jacob, G. S., Schaefer, J. & Wilson, G. E. Solid-state 13C and 15N nuclear magnetic resonance studies of alanine metabolism in Aerococcus viridans (Gaffkya homari). *J. Biol. Chem.* **260**, 2777–81 (1985).
75. Gronenborn, A. M. & Clore, G. M. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci.* **5**, 174–177 (1996).
76. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).
77. Barbieri, L., Luchinat, E. & Banci, L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat. Protoc.* **11**, 1101–1111 (2016).
78. Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy inside living cells. *Journal of the American Chemical Society* **123**, 2446–2447 (2001).
79. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 8895–8901 (2001).
80. Selenko, P., Serber, Z., Gadea, B., Ruderman, J. & Wagner, G. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc. Natl. Acad. Sci.* **103**, 11904–11909 (2006).
81. Ogino, S. *et al.* 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**, 10834–10835 (2009).
82. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106–109 (2009).
83. Binolfi, A. *et al.* Intracellular repair of oxidation-damaged α-synuclein fails to target C-terminal modification sites. *Nat. Commun.* **7**, 10251 (2016).
84. Banci, L. *et al.* Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat. Chem. Biol.* **9**, 297–9 (2013).
85. Luchinat, E. & Banci, L. In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Acc. Chem. Res.* **51**, 1550–1557 (2018).

86. Selenko, P. *et al.* In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **15**, 321–9 (2008).
87. Starheim, K. K., Gevaert, K. & Arnesen, T. Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem. Sci.* **37**, 152–161 (2012).
88. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
89. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12**, 5–9 (2015).
90. Kaplan, M. *et al.* EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **167**, 1241-1251.e11 (2016).
91. Baker, L. A., Daniëls, M., van der Cruijsen, E. A. W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62**, 199–208 (2015).
92. Nand, D., Cukkemane, A., Becker, S. & Baldus, M. Fractional deuteration applied to biomolecular solid-state NMR spectroscopy. *J. Biomol. NMR* **52**, 91–101 (2012).
93. Kaiser, C. J. O. *et al.* The structure and oxidation of the eye lens chaperone αA-crystallin. *Nat. Struct. Mol. Biol.* **26**, 1141–1150 (2019).

Chapter 2:

DNP-supported solid-state NMR spectroscopy of proteins inside mammalian cells

This chapter is adapted from a published manuscript:
"DNP-supported solid-state NMR spectroscopy of proteins inside mammalian cells"
Published in *Angew. Chemie - Int. Ed.* **58**, 12969–12973 (2019)
Siddarth Narasimhan[#], Stephan Scherpe[#], Alessandra Lucini Paioni, Johan van der Zwan, Gert Folkers, Huib Ovaa and Marc Baldus
[#]- These authors contributed equally

Major modifications made to the original text include rearrangement of supporting information, addition of titles for the introduction and results sections and reformatting the bibliography.

Abstract

Elucidating at atomic level how proteins interact and are chemically modified in cells represents a leading frontier in structural biology. We have developed a tailored solid-state NMR spectroscopic approach that allows studying protein structure inside human cells at atomic level under high-sensitivity dynamic nuclear polarization (DNP) conditions. We demonstrate the method using ubiquitin (Ub), which is critically involved in cellular functioning. Our results pave the way for structural studies of larger proteins or protein complexes inside human cells, which have remained elusive to in-cell solution-state NMR spectroscopy due to molecular size limitations

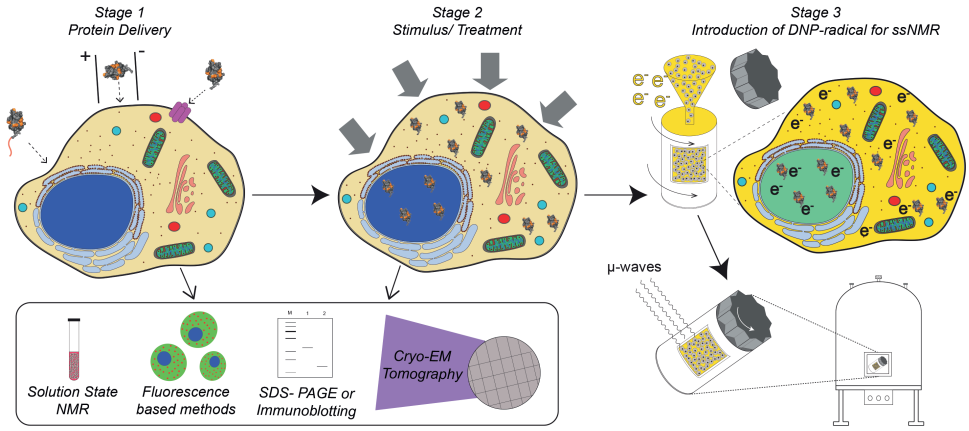
Introduction

Increasing evidence suggests that the highly complex and dynamic environment of the cell interior and its physiochemical setting imposes critical control on cellular functions, which is hardly reproducible under in-vitro conditions. In-cell solution-state NMR spectroscopy can track such structural and dynamic interactions at the atomic level provided that proteins or other molecular units are small and tumble rapidly[1–4]. On the other hand, solid-state NMR spectroscopy (ssNMR) has been used to probe proteins and large protein complexes in bacterial cells[5–12] and at the cell membrane periphery of human cells[13].

However, extending such studies to investigating proteins and molecular complexes inside human cells poses additional challenges. Firstly, molecule-specific isotope labelling must be achieved to spectroscopically detect the protein of interest in a complex cellular background. Furthermore, cellular ssNMR studies should be possible at endogenous protein concentrations to ensure proper cell functioning. As a result, high-sensitivity ssNMR methods are needed that allow such proteins to be studied in an intact cellular environment. Dynamic nuclear polarization[14] (DNP), in which polarization is transferred from free electrons to atomic nuclei, greatly enhances ssNMR sensitivity. Previous work has shown that DNP is readily compatible with ssNMR on bacterial[9] and human[13,15] cells, cell compartments[9–11,13,16] as well as cell lysates[17,18]. However, the strong reducing environment inside the cells can be deleterious to DNP radicals[19], thus far precluding protein studies inside human cells.

Herein, we describe a dedicated ssNMR approach, detailed in Scheme 1, to overcome the aforementioned challenges by separating the biochemical and cell preparation steps from the NMR procedures, allowing us to directly examine molecular interactions inside cells at high-sensitivity DNP conditions. In the

first stage of Scheme 1, an isotope-labelled protein, prepared, for example, using recombinant expression in Escherichia coli or solid-phase peptide synthesis (SPPS, ref.[20]), is delivered into human cells of interest. In the current context, isotope-labelled protein was delivered into human cells by electroporation[21], which, after a recovery period, can be followed by stage 2 in which cells are subjected to a stimulus or other functional treatments. Finally, and preceding the ssNMR measurements, DNP agents are introduced into the cells which are then filled into DNP-ssNMR magic angle spinning (MAS) rotors. At every stage, biochemical and biophysical methods can be used to monitor cellular processes, thus paving the way for correlative studies including ssNMR.



Scheme 1: In-cell DNP-supported solid-state NMR protocol. A general approach to studying molecular interactions inside human cells using DNP-supported ssNMR consisting of three steps, that is, isotope labelling and protein delivery, followed by applying a stimulus to the cells, and finally preparing the cells for DNP-ssNMR measurements using DNP agents (shown as electrons). Possible reference experiments, including solution-state NMR and microscopy studies utilized in the current work, are indicated.

Materials and Methods

Expression and purification of Ub

Glycerol stocks of *E. coli* BL21 Rosetta (R2) cells containing a plasmid DNA with wild type Ub sequence, a T7 promoter and, ampicillin and chloramphenicol resistance were cultured overnight in LB medium containing 35 mg/L of chloramphenicol and 50 mg/L of ampicillin. A 100 μL overnight preculture was added to 50 mL unlabelled M9 medium and grown until the OD_{600} reached 0.6, upon which it was scaled up to half of the required final volume of labelled medium (usually 1L or 0.5L) and grown until the OD_{600} reached 0.3. The cells were then

spun down and resuspended in the required volume of fully labelled M9 medium (supplemented with 2g/l of ^{13}C-D-glucose and/ or 0.5 g/L of $^{15}NH_4Cl$). When the OD_{600} reached 0.6- 0.7, the expression of Ub was induced by the addition of 0.5 mM IPTG (final concentration), and the expression was carried out at 37°C for 6 hours. For the production of unlabelled Ub, an overnight culture was added to the required final volume of LB medium and the expression was carried out in the same way as with the M9 medium. After the expression period, the cells were spun down, resuspended in 10 mM Tris buffer pH 7.4 containing lysozyme. After a freeze-thaw cycle, the cells were lysed by sonication. The lysate was supplemented with ammonium acetate buffer pH 5.1 at a final concentration of 20 mM and the pH was adjusted to 4.5-5.0 using acetic acid. A clear lysate was obtained by spinning down the insoluble components at 40,000 x g for 30 mins and was subsequently loaded on to a 5 mL HiTrap SP HP cation exchange column (GE Life Sciences), equilibrated with a 20 mM ammonium acetate buffer pH 5.1. A salt gradient of 0- 1M NaCl was used for elution. The fractions containing Ub were pooled and Ub was isolated by size exclusion chromatography using a Superdex 75 gel filtration column (GE Life Sciences) equilibrated in 20 mM ammonium acetate pH 5.1, 150 mM NaCl buffer. The fractions containing pure Ub were pooled and buffer-exchanged with water prior to use. For storage, the protein was lyophilized and stored in dry form at -20°C.

Preparation of Ub using solid phase peptide synthesis

Solid phase peptide synthesis (SPPS) was performed in syringes using a Syro II synthesizer (Multisyntech GmbH, Witten, Germany). Peptides were synthesized using standard 9-fluorenylmethoxycarbonyl (Fmoc) based solid phase synthesis as described before[20] with preloaded Fmoc amino acid trityl resin (0.2 mmol/g, Rapp Polymere GmbH) on a 25 µmol scale. Swelling of the resin was achieved by addition of 1.25 mL N-methyl-2-pyrrolidone (NMP) for 5 min (x2). Fmoc was removed by treatment with 20% piperidine in NMP thrice for 3, 5 & 5 min, followed by five washing steps of NMP. Amino acids were coupled by applying a fourfold excess in the presence of 4 equivalents PyBOP and 8 equivalents DiPEA in NMP for 45 min, followed by three washing steps with NMP. After completion of all coupling cycles, the resin was washed with Et_2O and dried under vacuum. The ubiquitin peptide sequences were synthesized on resin following the procedures as described before[20].

Coupling of tetramethylrhodamine (TMR) on synthesized Ub

Coupling of TMR at the N-terminus of Ub (1-76) was carried out as previously described[22]. Resin bound Ub (25 µmol) was incubated overnight with 1 mL DMF containing TAMRA (100 µmol), PyBOP (100 µmol), DiPEA (200 µmol) under shaking at room temperature. The resin was washed with NMP, DCM and Et_2O and dried under

high vacuum. The peptide was cleaved from the resin and deprotected by treatment with 2 mL trifluoroacetic acid/water/phenol/triisopropylsilane (90/5/2.5/2 vol%) for 3 h followed by precipitation into cold Et_2O/pentane (3/1 vol%). The precipitated peptide was washed twice using Et_2O, dried under vacuum and lyophilized from water/acetonitrile (70/30 vol%) and lyophilized. The crude product was dissolved in a minimal amount of DMSO and diluted into water, resulting in a final DMSO concentration of <10%. The peptide was purified by preparative RP-HPLC using 3 mobile phases: A: water, B: CH_3CN and C: 1% TFA in water. Gradient: 0 – 5 min: 90% A, 5% B, 5% C; 5 – 7 min: 90 – 75% A, 5 – 20% B, 5% C; 7 – 18 min: 75 – 50% A, 20 – 45% B, 5% C, on a Waters 2535 HPLC using a Waters C18-Xbridge 5 µm OBD (30 x 150 mm) column at a flowrate of 37.5 mL/min, detected by a Waters 2489 UV/Vis detector. HPLC chromatograms were measured using a Waters 1525EF binary HPLC pump and a Waters 2487 Dual λ Absorbance Detector using 2 mobile phases: A: 0.05% TFA in water and B: CH_3CN. Gradient: 0 – 1 min: 95% A, 5% B; 1 – 11 min: 95 – 10% A, 5 – 90% B on a C18-Xbridge 5 µm (4.6 x 150 mm) column at a flowrate of 1.4 mL/min, visualized by a Waters 2487 Dual λ Absorbance Detector. Pure fractions were pooled and lyophilized resulting in 18.7 mg (8%). MS ES+ (amu) calc: 8959, found: 8959.

LC-MS measurements were performed on a Waters LCT-premier mass spectrometer system equipped with a Waters 2795 Separation Module (Alliance HT), Waters 2996 Photodiode Array Detector (190-750 nm) Phenomenex Kinetex C18 (2.1x50, 2.6 µm, flow rate= 0.8 mL/min, runtime= 6 min, column T= 40°C. Gradient: 0 – 0.5 min: 95% A, 5% B; 0.5 – 4 min: 95 – 5% A, 5 – 95% B; 4 – 5.5 min: 5% A, 95% B) and LCTTM ESI-Mass Spectrometer. Samples were run using 2 mobile phases: A = 1% CH_3CN, 0.1% formic acid in water and B = 1% water and 0.1% formic acid in CH_3CN. Data processing was performed using Waters Mass Lynx Mass Spectrometry Software 4.1 (deconvolution with Maxent1 function). High resolution mass spectra were recorded on a Waters XEVO-G2 XS Q-TOF mass spectrometer equipped with an electrospray ion source in positive mode (source voltage 30 V, desolvation gas flow 900 L/hr, temperature 250°C) with resolution R = 26000 (mass range m/z = 50-2000 m/z, capillary Voltage 1.2 kV) and 200 pg/µL Leu-Enk (m/z = 556.2771) as a "lock mass".

Cell culture & electroporation mediated Ub delivery

HeLa cells were grown as 2D cultures, in a 37°C CO_2 incubator, in Dubelco's Modified Eagle's Medium (DMEM Glutamax- Produced by Gibco- ThermoFisher Scientific) supplemented with 10% FBS (Gibco- ThermoFisher Scientific) devoid of any antibiotics. Due to the absence of antibiotics, the cells were only grown in T75 or

T175 flasks to reduce the risk of microbial contamination. The electroporation protocol was adapted from the work of Selenko Lab[21]. Healthy cells at 70-80% confluency were harvested and resuspended in electroporation buffer composed of 100 mM $NaHPO_4$ pH 7.0, 50 mM HEPES pH 7.0, 5 mM KCl, 15 mM $MgCl_2$, 2 mM ATP, 2 mM reduced glutathione (GSH) and 300-1200 μM Ub (unless mentioned, 1200 μM was used). 4 million cells were resuspended in every 100 μL of electroporation buffer and 100 μL of the suspension was added per electroporation cuvette (Sigma-Aldrich 0.2 cm gap). The cells were pulsed twice with a gap of 10 seconds between each pulse, using a Lonza Nucleofector II device (program B-28). Subsequently, cells were collected in prebuffered and prewarmed (37°C) DMEM medium supplemented with 10% FBS. The cells were allowed to recover for at least 5 h in the CO_2 incubator at 37°C. After recovery and optional epoxomicin treatment, the (attached and healthy) cells were harvested by trypsinization. Please note that there is a small batch to batch variation in the amount of delivered Ub, as seen in the western blots and the SDS-PAGE.

Proteasome inhibition by Epoxomicin

The cells were prepared by delivering the required variant of Ub (unlabelled, isotope labelled or TAMRA- labelled) through electroporation. After 5 h of recovery, the medium was replaced with DMEM medium (containing 10% FBS) with 2 μM epoxomicin (Enzo Life Sciences) and 1% penicillin & streptomycin (Gibco-ThermoFisher Scientific). The treatments were carried out for 6 h.

Western blots

After optional electroporation recovery & epoxomicin treatment, cells were harvested by trypsinization, resuspended in 1X laemmli buffer and boiled for 10 mins at 95°C. Samples were loaded into Bis-Tris gradient gels (Invitrogen, NuPAGE 10%) and electrophoresis was performed using MES SDS running buffer (NuPAGE) at a steady voltage of 150V. The proteins were electroblotted onto a 0.45 μm nitrocellulose membrane (Bio-Rad) with Tris-glycine-SDS buffer containing 20% ethanol using a wet transfer apparatus (Bio-Rad) at a steady current of 140 mA for 150 min. Subsequently, the blot was washed thrice (5 min each time) using phosphate buffered saline (PBS) containing 0.1% Tween 20 (PBST buffer). Blocking was done for 1 h at room temperature using PBST with 5% milk. Primary antibody incubation was done overnight in PBST with 5% milk at 4°C, against Ub (mouse anti-Ub: sc-8017, Santa Cruz Biotechnology) or β-Actin (mouse anti-Actin: A5441, Sigma). The blots were washed thrice using PBST buffer and secondary antibody (Goat Anti-Mouse 800CW, Licor IRDye, 926-32210) incubation was done for 1 h in PBST with 5% milk at room temperature. The blots were visualized using an Odyssey imaging system. The blots were processed using FIJI Software[23].

Confocal microscopy

Recovery after electroporation was done on sterile 0.01% Poly-L-Lysine coated 12 mm coverslips in 12 well plates (Cell Star). After recovery and optional epoxomicin treatment, the coverslips were washed with PBS and the cells were fixed using 4% paraformaldehyde in PBS for 15 mins. Cells were permeabilized using 0.1% Triton-X in PBS for 10 mins and 100 nM phalloidin-alexa 488 (Acti-stain 488 Fluorescent Phalloidin, Cytoskeleton Inc.) staining was done for 30 mins followed by 10 μg/mL DAPI (Sigma) staining for 5 mins. After every step (fixing, permeabilization and staining), the coverslips were washed twice using PBS. The coverslips were finally mounted on a glass slide with SlowFade Gold Antifade Mounting Medium (ThermoFisher Scientific) and the edges of the coverslip was sealed using transparent nail polish. For imaging the diffusion of DNP agents by confocal microscopy we used PyPol-MTSSL[24,25] conjugated to TAMRA. Samples were imaged on a Leica SP8 microscope equipped with appropriate solid-state lasers. For all confocal imaging, HCX PL 63x 1.32 oil objectives and HyD detectors were used. Digital zoom ranging from 1.5x-3x was employed as applicable.

Solution-state NMR experiments on HeLa cells & lysates

48 million cells were electroporated with ^{15}N labelled Ub at a final concentration of 1200 μM in the electroporation buffer. After a recovery period of five hours, cells were harvested by trypsinization, washed with PBS and resuspended in Leibowitz Medium with 10% FBS and D_2O. After NMR measurements on intact cells, a freeze-thaw step was performed on the cell pellet followed by lysis in lysis buffer (50 mM Tris pH 7.0, 5 mM $MgCl_2$, 250 mM sucrose and 0.1% NP-40) using a dounce homogenizer. Clear lysate was obtained by spinning down the lysate at 12,000 x g for 30 min. Prior to NMR experiments, 10% D_2O (final concentration) was added to the lysate to enable field locking. ^{15}N-^{1}H SOFAST-HMQC spectrum[26] was acquired at 298 K with a triple channel (^{1}H, ^{13}C, ^{15}N) cryogenically cooled-probe, at a static magnetic field of 21.1 T, corresponding to a proton frequency of 900 MHz with 1024 scans with a delay of 0.25 s. Acquisition times were 56 ms (2048 data points- 20 ppm spectral width) and 17 ms (96 data points- 31 ppm spectral width) for the ^{1}H and ^{15}N dimensions, respectively. The spectrum was processed using a cosine squared window function in both dimensions.

Sample preparation for DNP solid-state NMR experiments

For ssNMR sample preparation, 8 million cells were electroporated with (^{13}C, ^{15}N) Ub at a final concentration of 1200 μM in the electroporation buffer. After a recovery period of five hours, the cells were harvested by trypsinization, washed with

PBS and resuspended in 50 μL of deuterated DNP juice (60% deuterated d_7- and ^{12}C enriched (99.95%) glycerol in D_2O or just d_7- glycerol) comprising of 15-30 mM AMUPol[25] in Hank's Buffered Salt Solution (HBSS: 140 mM NaCl, 5 mM KCl, 1 mM $CaCl_2$, 0.4 mM $MgSO_4$, 0.5 mM $MgCl_2$, 0.3 mM Na_2HPO_4, 0.4 mM KH_2PO_4, 6 mM glucose and 4 mM $NaHCO_3$). The cells were then spun down into DNP rotors by mild centrifugation (lower than 1000 x g). The supernatant containing excess DNP juice was removed, and the rotor was snap frozen in liquid nitrogen prior to magic angle spinning (MAS) at 100 K.

Solid-state NMR experiments

We conducted solid-state NMR experiments under low temperature (100 K) DNP conditions using 3.2 mm triple-resonance (^1H, ^{13}C, and ^{15}N) magic-angle-spinning (MAS) probe head in a static magnetic field of 9.4 T, corresponding to proton/electron resonance frequencies of 400 MHz/263 GHz and 3.2 mm double-resonance (^1H, ^{13}C) magic-angle-spinning (MAS) probe head in a static magnetic field of 18.8 T, corresponding to proton/electron resonance frequencies of 800 MHz/527 GHz (Bruker BioSpin). All spectra were recorded at a MAS rate of 8 kHz using SPINAL-64 proton decoupling[27] at 84 kHz. A recycle delay of 2s was used for all experiments.

^1H-^{13}C cross polarization experiments for measuring DNP enhancements

50-70 μs contact times were used for H-C CP transfer. 100/200 Hz line-broadening were applied prior to Fourier transformation.

2D ^{13}C, ^{13}C spindiffusion (PDSD) experiments

For all the experiments, PDSD mixing time of 30ms was used. Contact times of 110 μs and 70 μs were used for H-C cross polarization on in-vitro and in-cell Ub samples respectively. Acquisition times were 15 ms (600 data points used- 405.4 ppm spectral width) and 5 ms (282 data points- 280 ppm spectral width) for the direct and indirect dimensions, respectively. Cumulative number of scans of 32 and 256 were used for the in-vitro and in-cell Ub samples respectively. The 2D spectra were processed using a 0.28 π shifted sine squared window function on both dimensions.

2D 2Q-1Q ^{13}C, ^{13}C Correlation Experiment

The 2D 2Q-1Q ^{13}C, ^{13}C correlation spectrum, was obtained using 10 blocks of SPC5[28] mixing (2.5 ms) for generation and subsequent reconversion of double quantum coherences. A contact time of 70 μs was used for H-C cross polarization and continuous wave (CW) ^1H decoupling was applied during the SPC5 sequence with a field strength of 85 kHz. Acquisition times were 10 ms (400 data points used-

307 ppm spectral width) and 1.1 ms (60 data points- 280 ppm spectral width) for the 1Q direct and 2Q indirect dimensions, respectively. 512 scans were acquired. With forward linear predictions (4 points using 16 coefficients) on DQ dimension, the 2D spectra were processed using a 0.25π shifted sine squared window function in both dimensions.

2D ^{15}N-^{13}C correlation (NCA) experiment

^{15}N-^{13}C-correlation experiment was performed using a ^{1}H-^{15}N CP step of 1 ms and a SPECIFIC-CP[29] ^{15}N-^{13}C step of 2.5 ms. The center frequencies were 50 ppm and 120 ppm for ^{15}N and ^{13}C dimensions, respectively. Acquisition times were 10 ms (800 data points used- 497 ppm spectral width) and 3.4 ms (25 data points- 90 ppm spectral width) for the direct and indirect dimensions, respectively. 384 scans were acquired. the 2D spectra were processed using a 0.33 π shifted sine squared window function in both dimensions.

3D 2Q-1Q-1Q ^{13}C, ^{13}C correlation experiment

The 3D 2Q-1Q-1Q ^{13}C, ^{13}C correlation spectrum, was obtained using 10 blocks of SPC5[28] mixing (2.5 ms, MAS: 8kHz) for generation and subsequent reconversion of double quantum coherences. A contact time of 70 μs was used for H-C cross polarization and continuous wave (CW) H decoupling was applied during the SPC5 sequence with a field strength of 85 kHz. Following reconversion of double quantum coherences, PARIS[30] recoupling with an amplitude of 4 kHz was applied on the ^{1}H channel to achieve a spin diffusion mixing time of 30 ms. Acquisition times were 10 ms (400 data points used- 307 ppm spectral width) on the F_3 1Q direct dimension, 1.5 ms (64 data points- 212 ppm spectral width) for the F_2 1Q indirect dimension and 0.8 ms (48 data points- 290 ppm spectral width) for the F_1 2Q indirect dimension. 64 scans were acquired. With forward linear predictions (8 points using 8 coefficients on F2 dimension and 4 points using 8 coefficients on the F1 dimension), the 2D spectra were processed using a 0.28 π shifted sine squared window function on all dimensions.

Multidimensional spectra analysis

For spectral analysis, NMRFAM-SPARKY software package was used[31]. ssNMR assignments were obtained from BMRB ID: 7111 of microcrystalline Ub. For Ub loops for which ssNMR assignments were not available, we supplemented assignments from solution NMR (BMRB ID: 15410). As it has been shown before, solution NMR assignments do not differ greatly from microcrystalline solid state NMR assignments, especially on carbons[32]. The correlations expected per experiment were prepared using the web-based chemical shift prediction tool FANDAS 2.0[33].

Results and Discussion

2.1 Exogenously delivered wild-type Ubiquitin is functional in HeLa cells

To demonstrate the feasibility of the entire approach, we concentrated on ubiquitin (Ub), a post-translational modifier that regulates a large variety of cellular functions, particularly protein degradation[34,35]. The process known as ubiquitination entails the covalent attachment of Ub to the N-terminus or an internal lysine of a substrate, mediated by a concerted cascade of specialized classes of enzymes (E1, E2, and E3). Ub can itself be ubiquitinated in the same manner, leading to the formation of a variety of Ub chains that determine the fate of the substrate. While magnetic resonance studies of Ub have provided valuable insight into Ub chain formation in vitro (see, for example, Ref.[34,36]), wild-type Ub has remained elusive to in-cell solution-state NMR[2] and previous studies have required the mutation of Ub residues known to be involved in numerous protein-protein interactions[2]. Using SPPS, we first introduced an N-terminal tetramethylrhodamine (TMR)-tag to the synthesized Ub (TMR–Ub). Confocal microscopy images of cells after recovery from electroporation (Figure 1A) revealed that the cells retained their normal morphology and that TMR–Ub was well integrated into the nuclear and cytoplasmic compartments. Importantly, electroporation enabled us to control the amount of protein delivered into cells, as evident from the presence of free monomeric and conjugated Ub at increasing concentrations on the western blot (Supporting Information, Figure S1A,C). Further analysis revealed that the concentration of Ub delivered into the cell is comparable to endogenous Ub concentrations[37] as seen from our western blots (Supporting Information, Figure S1 B,D). Furthermore, an SDS-PAGE analysis using TMR–Ub allowed us to selectively track the exogenously introduced Ub population, which is, in addition to free monomeric Ub, mostly attached to 14–17 kDa proteins (Figure 1B), in line with earlier work[38] predicting Ub interactions with the nucleosomal proteins H2A and H2B. Indeed, such interactions would also be in line with the location of Ub in the nucleus as revealed by our confocal microscopy studies (Figure 1A). To further test the functionality of exogenous Ub, we performed a proteasome inhibition assay. The addition of the proteasome inhibitor epoxomicin led to the depletion of the nuclear Ub pool and an increase in the level of cytoplasmic Ub (Figure 1A, left), in line with earlier studies[38]. An increase in the levels of Ub conjugation to substrates was also apparent from western blotting against wild-type Ub (Supporting Information, Figure S1C,D) and SDS-PAGE (Figure 1B) probing exogeneous TMR–Ub. Taken together, these results suggested that exogenous Ub can be introduced into HeLa cells, and that it is functional as evident from conjugate formation in response to proteasome inhibition.

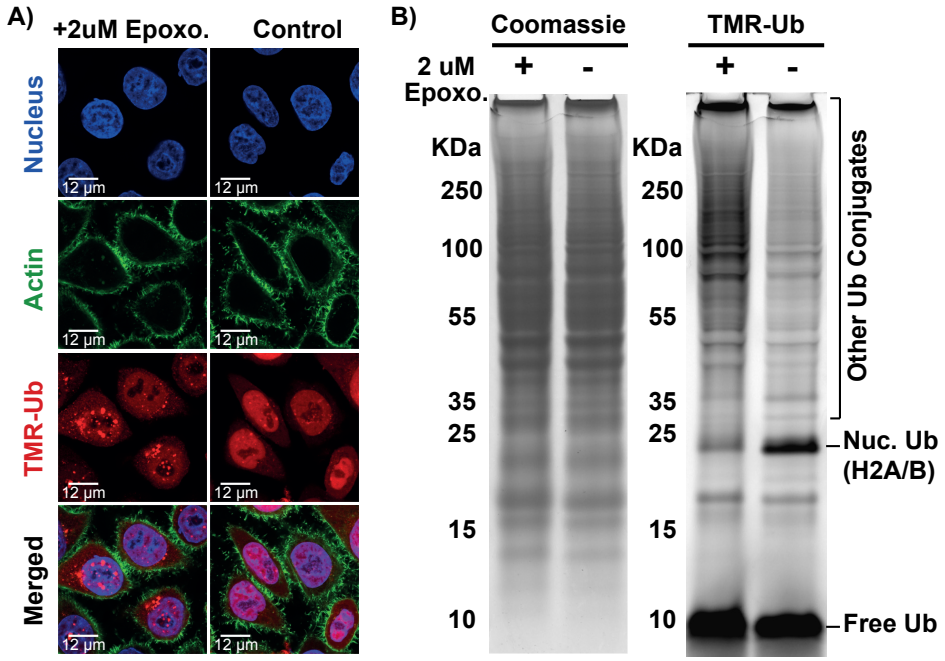


Figure 1: A) Single Z-slices of confocal microscopy images of electroporated TMR–Ub before (right) and after (left) proteasome inhibition. Note that in the latter case, the nuclear Ub pool is depleted and puncta-like formations are seen in the cytoplasm. B) SDS-PAGE (TMR scan) analysis of exogeneous TMR–Ub showing the formation of higher molecular weight conjugates and the depletion of nuclear Ub (25 KDa) upon proteasome inhibition.

2.2 Stability of in-cell Ub is revealed through solution NMR analysis

To check degradation of delivered Ub after electroporation and recovery, we conducted solution-state NMR experiments of ^{15}N-labelled wild-type Ub electroporated into HeLa cells (Figure 2). In line with earlier studies[2,21], ^{15}N-^{1}H SOFAST-HMQC[26] experiments on cells suspended in Leibowitz L-15 medium revealed no traces of degradation. The spectrum (Figure 2A) contained five backbone amide correlations, of which R74, G75, and G76 in the C-Terminal tail were unambiguously identified. We confirmed that there was no protein leakage by recording a ^{15}N-edited ^{1}H 1D experiment on the medium (Figure 2B). In line with earlier work[2], we attribute the absence of signals from the structured regions of Ub to the myriad of molecular recognition events and to the high viscosity of the cell interior. This notion was further confirmed by additional NMR experiments after cell-lysis in which the NMR signals of folded Ub were readily recovered without any sign of protein unfolding or degradation (Supporting Information, Figure S2). Most likely, the major population available to in-cell solution NMR studies is the free, unconjugated Ub, which constitutes only a fraction of the total Ub pool (Figure 1B

and Supporting Information, Figure S1), further strengthening the need for ssNMR approaches.

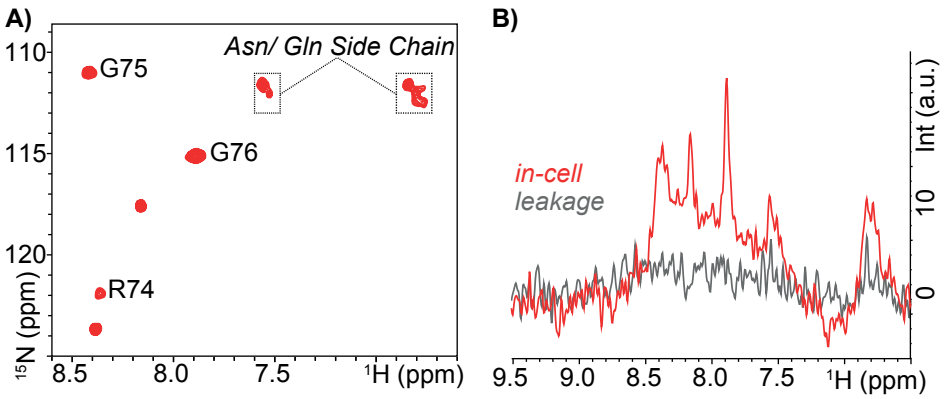


Figure 2: A) Solution-state NMR two-dimensional (^{1}H-^{15}N) SOFAST-HMQC spectra of ^{15}N-labelled Ub electroporated into HeLa cells performed 5 h after electroporation. Only backbone amide correlations emanating from the flexible C-terminal tail are detected, most likely from the free monoubiquitin pool. B) 1D slices were extracted from the spectra of in-cell ^{15}N-labelled Ub obtained before (red) and after (gray) depletion of the cells and showed negligible leakage of the protein during the NMR experiment.

2.3 DNP agents diffuse into the cells and provide high signal enhancements

For our ssNMR experiments described in the next section, we estimated that 3.2 mm MAS rotors can hold about 5–8 million cells, corresponding to approximately 10–30 µg of labelled Ub in our ssNMR preparations (Supporting Information, Figure S3). Such quantities underline the need for efficient DNP for in-cell ssNMR studies. Previous calculations in our laboratory[39] have suggested that, to establish efficient DNP inside mammalian cells, the DNP agent must be localized inside the detached cells, for example as a result of rapid diffusion to minimize the effect of the reducing environment inside the cells. Using a DNP biradical variant of AMUPol[24,25] conjugated to TMR (Supporting Information, Figure S4) that was resuspended in a DNP buffer (see Materials and Methods), we tracked the location of DNP agents by confocal microscopy. Indeed, the water-soluble DNP agents were present in both nuclear and cytoplasmic cell compartments (Figure 3A) after a period of 10–15 minutes. Hence, we applied a similar time period in the final stage of our approach (Scheme 1) in which DNP-ssNMR MAS rotors were filled with cells using mild centrifugation. Subsequently, the DNP rotor was rapidly frozen in liquid nitrogen and transferred into the pre-cooled DNP-ssNMR probe.

Using this procedure, we measured up to 130- and 35- times signal enhancement at 400 and 800 MHz DNP conditions (Figure 3B and Supporting Information, Figure S5), respectively, compared to the case without DNP. These results were significantly higher than previous studies on in situ systems, including human-cell vesicles[13] or fully labelled mammalian cells[15]. Moreover, ^{13}C T_1 relaxation times of protein signals ranged around 5 seconds further supporting the notion that the DNP agents were in close proximity to the target protein Ub.

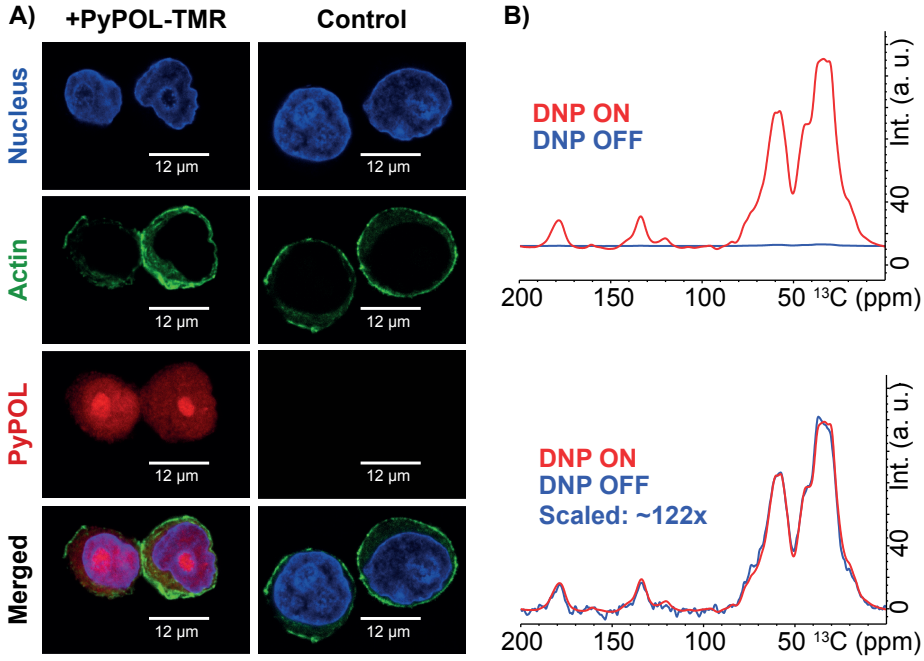


Figure 3: A) Z-Slices of confocal microscopy images showing that the radical (PyPOL-TMR, a variant of AMUPol) is well distributed in both nuclear and cytoplasmic compartments of the cell. B) DNP signal enhancement of approximately 122 times as seen on ^{1}H-^{13}C CP experiments under 400 MHz DNP conditions.

2.4 2D & 3D in-cell DNP-ssNMR spectra of Ub

The findings discussed in the previous section allowed us to conduct a series of 2D ssNMR experiments to detect Ub in our DNP-ssNMR preparations including double quantum-single quantum (2Q-1Q) $^{13}C,^{13}C$ and N to C_α (NCA) 2D experiments (Supporting Information, Figures S6,7). In Figure 4, the aliphatic region of a 2D ^{13}C, ^{13}C proton-driven spin diffusion (PDSD) experiment (see Figure S8 in the Supporting Information for the complete spectrum) is shown. For comparison, Figure 4 also includes results of a 2D DNP-ssNMR experiment on Ub before delivery into cells (black) in DNP buffer and in vitro NMR assignments of Ub at ambient temperature (black crosses, see also Materials and Methods). Overall, the in vitro PDSD experiment was in good agreement with the in-cell ssNMR spectrum of Ub and included well-resolved correlations, such as for Ile30 and Pro37.

To further improve spectral resolution, we resorted to 3D ssNMR (2Q-1Q-1Q) $^{13}C,^{13}C$ spectroscopy[40] (Figure 5 and Supporting Information, Figures S9–S12). Analysis of various 2D F_2,F_3 planes (Figure 5 and Supporting Information, Figures S9–S12) and F_1,F_3 planes (Supporting Information, Figures S9–S12) readily allowed us to unambiguously identify the chemical shift correlations of Ub, including Lys63, which is a functional hotspot of polyubiquitination that is linked to regulatory functions other than proteasomal degradation. When plotted on the 3D structure of Ub, these residues for which correlations were identified unambiguously in the 2D and 3D experiments were located throughout the protein sequence, including the N-terminus as well as the five beta sheets and the central alpha helix (Figure 6).

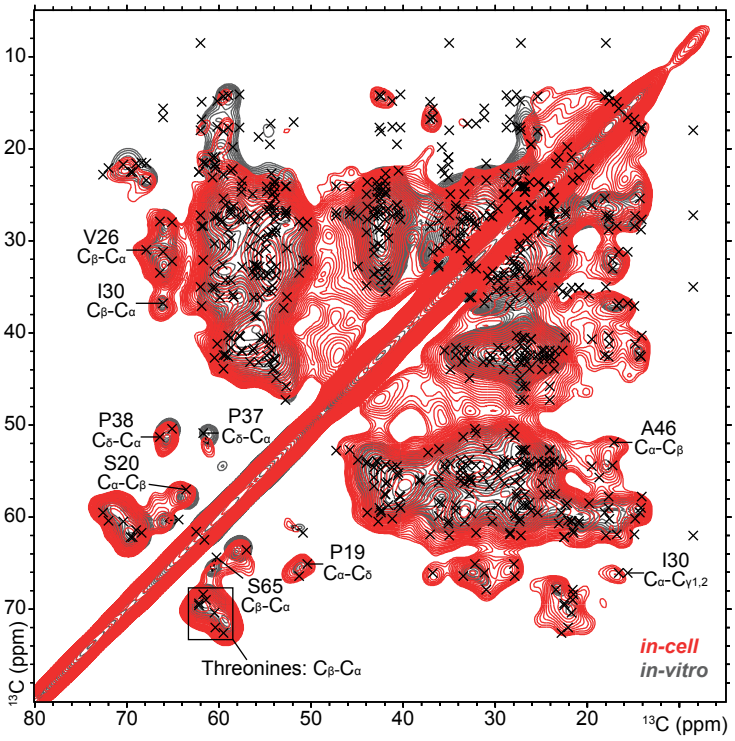


Figure 4: Aliphatic region of the 2D ^{13}C, ^{13}C-correlated spin diffusion (PDSD) experiment showing clear similarities in the spectrum of Ub in-vitro (gray) and in cell (red). This confirms that Ub remains folded after delivery into cells. This is further confirmed by our 3D DNP-ssNMR spectra. Indicated peaks were identified unambiguously.

With the combined analysis of 2D and 3D spectra, we could unambiguously trace back correlations (Supporting Information, Table ST1) for 25 residues, spread throughout the protein, which matched with the in-vitro assignments. This strongly suggests that uniformly ^{13}C- ,^{15}N- labelled Ub remains folded after delivery into cells. Such a conclusion would also be consistent with our microscopy results and SDS-PAGE analysis in Figure 1, which confirmed the proper location and biological activity of our exogenously introduced Ub.

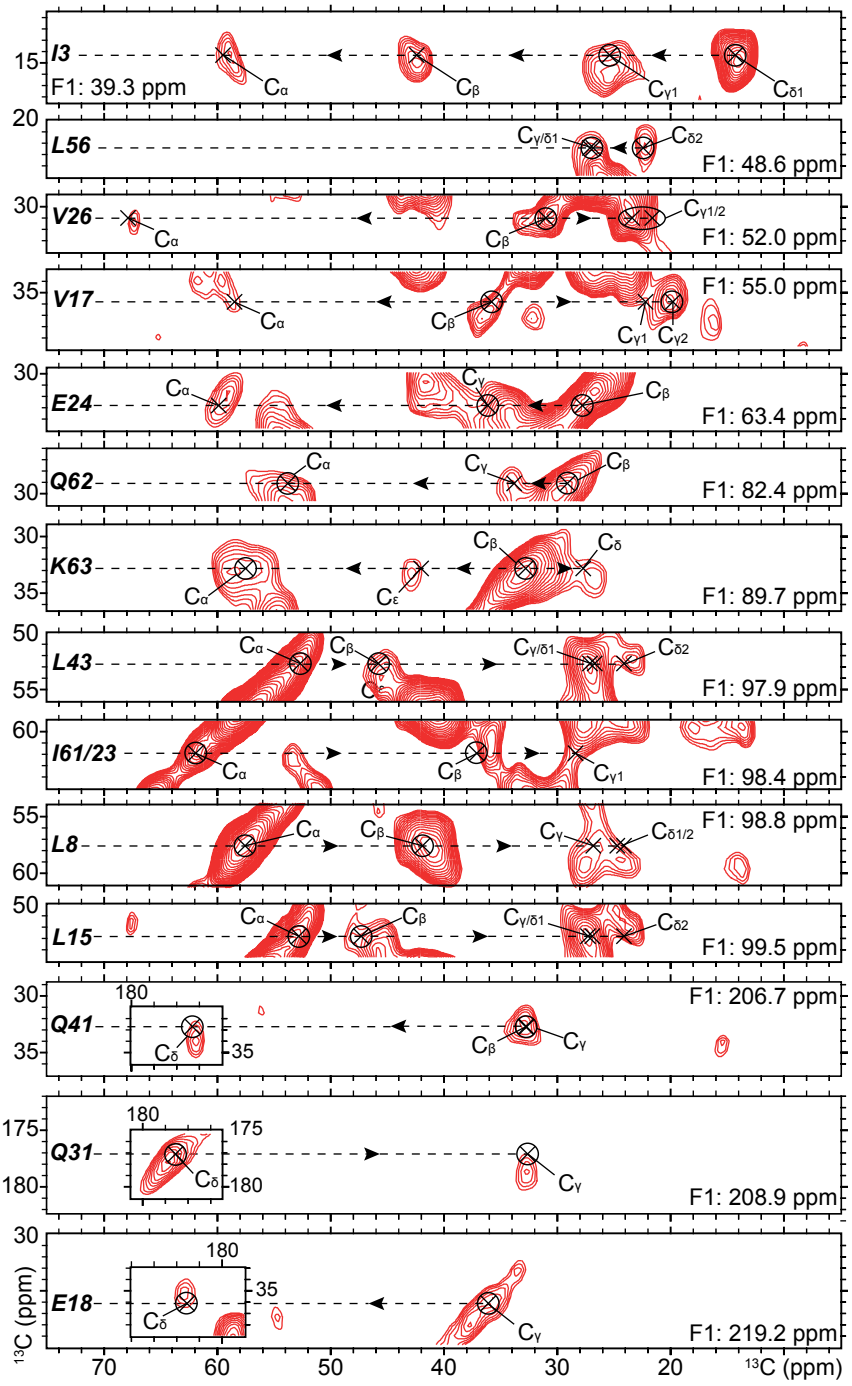


Figure 5: 2D (F_2, F_3) slices from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to the unambiguously identifiable peaks. The F_1 (2Q) frequency, indicated in the slices corresponds to the chemical shifts of encircled peaks.

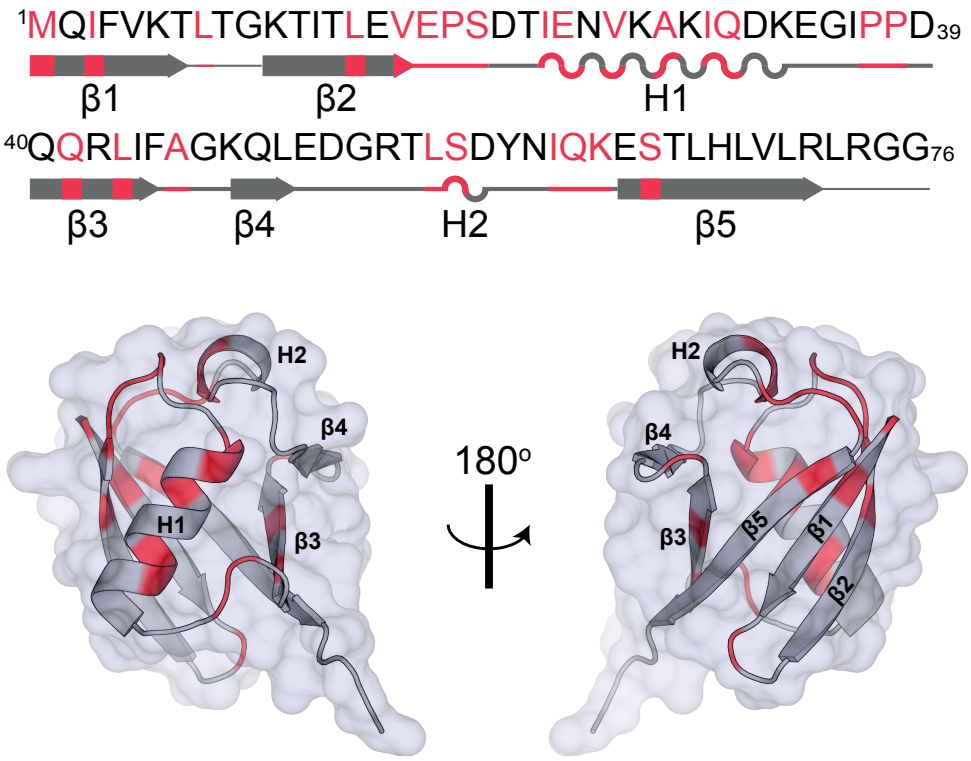


Figure 6: Residues of Ub for which correlations were identified unambiguously in the 2D and 3D experiments, plotted on the Ub crystal structure (PDB ID: 1UBQ). For further information, see the Supporting Information.

Conclusions and Perspectives for the Future

In summary, the presented in situ DNP-ssNMR scheme allowed us to detect Ub at endogenous levels in human cells by NMR. To further increase spectral resolution, SPPS as well as tailored isotope-labelling, including amino-acid-specific forward or reverse labelling[41], will be useful. For example, further studies that elucidate the effect of proteasomal inhibition by epoxomicin and other stress factors on Ub structures in cells are ongoing in our laboratory. Stress-induced polyubiquitination is followed by a plethora of molecular interactions within Ub chains[35,38]. Uncovering the interactions between Ub molecules and specific interacting proteins in vivo is a long-term prospect for our work. By potentially co-introducing mixed-labelled[42] proteins (or other biomolecules), molecular interactions can be targeted and studied at atomic level in the cellular environment. Furthermore, such experiments could be readily extended to other Ub-like proteins, such as SUMO or NEDD8[35].

In the current context, we have used electroporation, which has successfully been used for the delivery of proteins (see, for example, Ref.[21]) or nucleic acids into cells. In cases where this technique is not suited, for example due to protein size or solubilization issues, the present approach could be combined with other delivery methods (Scheme 1), such as cell-penetrating peptides (CPP[2]), pore-forming toxins[43], or directed protein expression[4], to achieve molecule-specific labelling in the cells of interest. Likewise, our method complements advanced microscopy, such as confocal microscopy, used in the present study, or electron tomography[8], studies to obtain unprecedented insights from the atomic to the nanometer level on molecular interactions and modifications that take place inside human cells. Such studies could significantly expand our understanding of cellular processes ranging from protein synthesis, folding, and clearance to cellular signalling.

Supporting Information

Table ST1: Peak list corresponding to correlations (in ppm) of residues indicated in Figure 6. When multiple peaks were identified (Figure 4-5, S6-12), only the correlations corresponding to the most resolved peak are shown.

Residue	Correlation (^{13}C)			Correlating Nuclei
	F_1	F_2	F_3	
M1	54.4	170.6	-	C_α, C_o
I3	39.7	14.3	42.4	C_δ+$C_{\gamma 1}$, $C_{\delta 1}$, C_β
L8	99.48	57.57	178.78	C_α+C_β, C_α, C_o
L15	100.1	52.8	47.3	C_α+C_β, C_α, C_β
V17	55.6	19.8	174.4	$C_{\gamma 2}$+C_β, $C_{\gamma 2}$, C_o
E18	219.2	183.1	36.1	C_δ+C_γ, C_δ, C_γ
P19	78.4	50.4	175.3	C_δ+C_γ, C_δ, C_o
S20	57.0	63.6	-	C_α, C_β
I23	97.0	62.0	35.0	C_α+C_β, C_α, C_β
E24	63.93	27.8	59.9	C_β+C_γ, C_β, C_α
V26	98.9	67.9	23.4	C_α+C_β, C_α, $C_{\gamma 2}$
A28	74.4	55.7	-	C_α +C_β, C_α
I30	66.1	36.8	-	C_α, C_β
Q31	209.7	32.6	177.1	C_γ+C_δ, C_γ, C_δ
P37	50.9	61.7	-	C_δ, C_α
P38	79.2	51.3	178.3	C_δ+C_γ, C_δ, C_o
Q41	207.3	174.6	32.7	C_δ+C_γ, C_δ, C_γ
L43	98.5	45.8	175.5	C_β+C_α, C_β, C_o
A46	51.9	17.1	-	C_α, C_β
L56	49.2	26.8	27.1	C_γ+$C_{\delta 2}$, C_γ, $C_{\delta 1}$
S57	124.1	61.6	178.4	C_α+C_β, C_α, C_o
I61	99.0	61.9	175.1	C_α+C_β, C_α, C_o
Q62	82.9	29.1	33.79	C_β+C_α, C_β, C_γ
K63	90.3	32.8	42.0	C_β+C_α, C_β, C_ε
S65	60.3	64.4	-	C_α, C_β

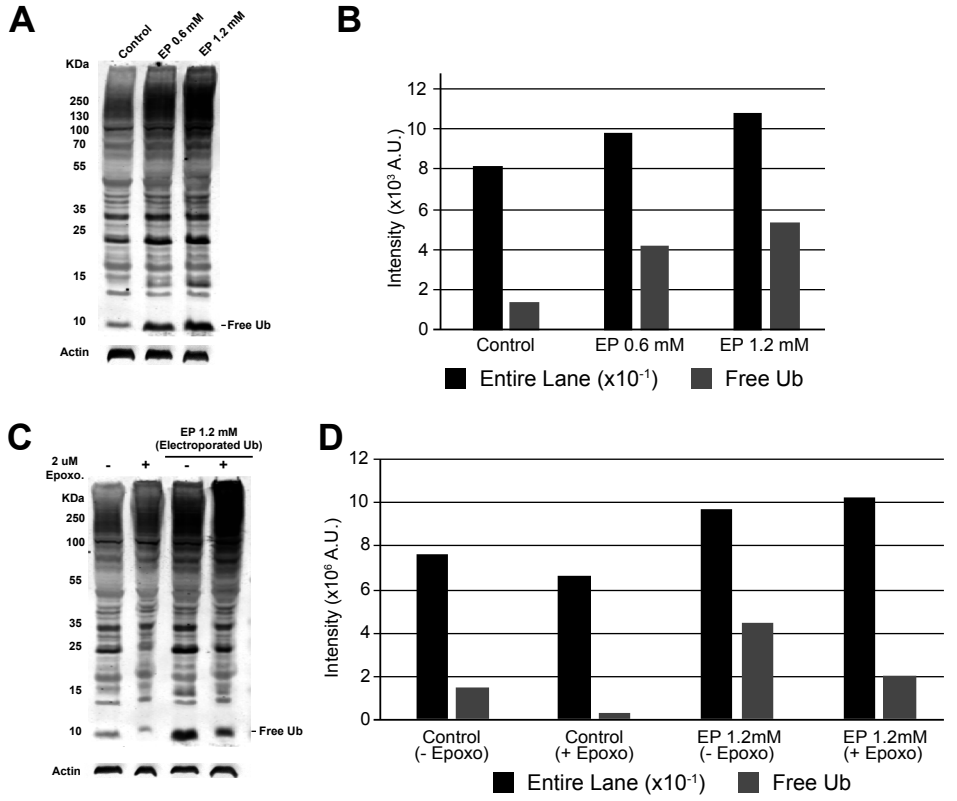


Figure S1: A) Western blot with an anti-Ub antibody showing an increase of Ub amounts in cells upon electroporation when increasing concentrations of Ub are used for electroporation (shown as EP 0.6 mM and 1.2 mM). Both the free pool of Ub and higher molecular weight conjugates are considered in the analysis. It is also noteworthy that we do not deviate largely from physiological concentrations as seen by comparing band intensities in unmanipulated cells (shown as control). B) FIJI analysis of lane intensities and monomeric intensities corresponding to Ub in the western blot in panel B. C) Western blot (against Ub, with actin control) of cells electroporated with Ub showing the depletion of the free Ub and formation of higher order Ub conjugates upon proteasome inhibition. D) FIJI analysis of lane intensities and monomeric intensities corresponding to Ub in the western blot in panel C. The values corresponding to the bar plot are available in Tables ST2 and ST3 for panel B and D respectively.

Table ST2: Raw Data- analysis of western blot shown in Figure S1(B).

Sample	Intensity (A. U.)	
	Entire lane	Free Ub
Control	81889.607	1372.092
0.6 mM	98358.829	4214.749
1.2 mM	108469.436	5353.991

Table ST3: Raw Data- analysis of western blot shown in Figure S1(D). + or – is indicative of epoxomicin treatment.

Sample	Intensity (A. U.)	
	Entire Lane	Free Ub
Control (-)	76308.172	1473.335
Control (+)	66728.595	345.556
1.2 mM (-)	97459.779	4489.406
1.2 mM (+)	102309.838	2028.456

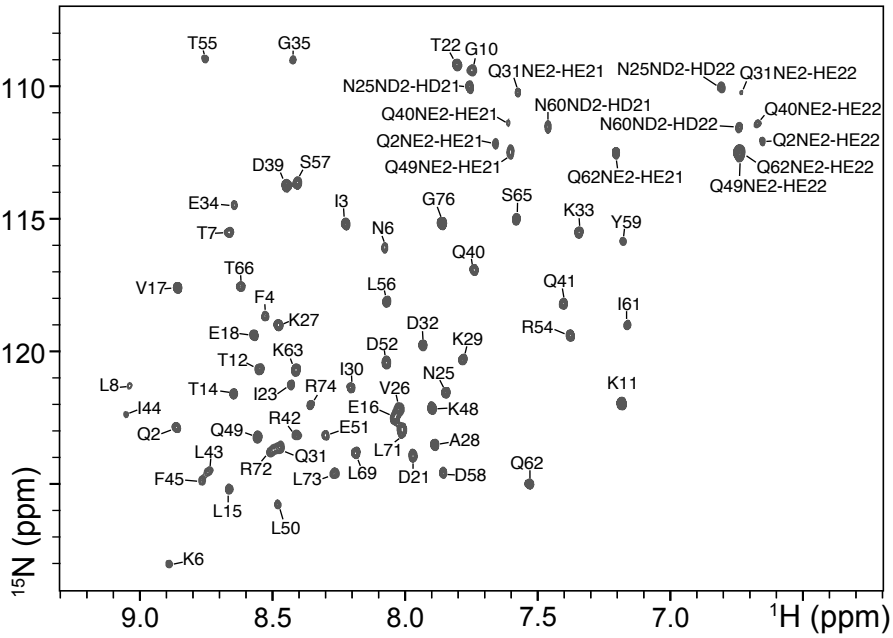


Figure S2: Solution-state NMR two-dimensional (^{1}H-^{15}N) SOFAST-HMQC spectra of ^{15}N- Ub electroporated into HeLa cells performed at a ^{1}H Larmor frequency of 900 MHz after 5 h of recovery and lysis the two-dimensional (^{1}H-^{15}N) SOFAST- HMQC spectrum recorded in identical conditions shown in Figure 2, still shows no signs of protein degradation or unfolding.

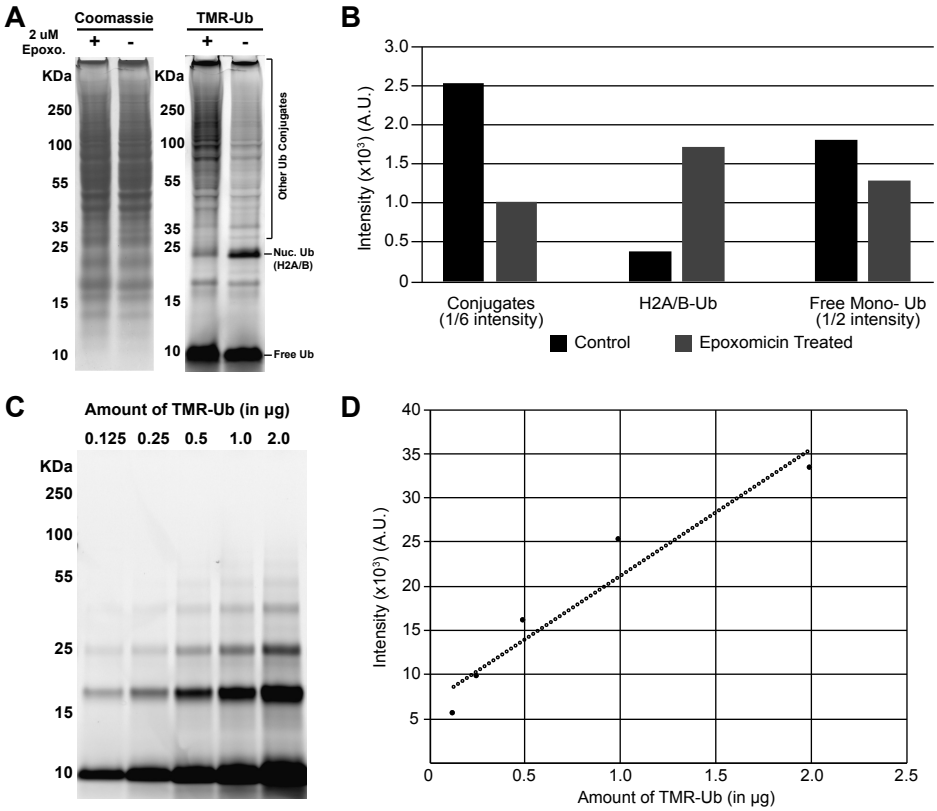


Figure S3: Quantification & analysis of Exogenous Ub A) SDS PAGE analysis (Coomassie, left and TMR scan, right) of TMR-Ub introduced into HeLa cells with & without Epoxomicin treatment B) FIJI analysis of intensities of described bands in panel A. C) Increasing amounts of TMR-Ub (Standard -D). The values corresponding to the plots are available in Tables ST4 and ST5 for panels D and B respectively. Standard curve for the plotted intensities in panel C is used for quantification of labelled material in the rotor, described in the main text. See also notes below table ST5 for calculation.

Table ST4: Raw Data- analysis of concentration series of TMR-Ub on an SDS-PAGE gel shown in Figure S3(D). Polymeric bands of TMR-Ub were seen on the gel. Combinded intensities corresponding to the mono-, di- and trimeric bands were plotted against the amount of TMR-Ub loaded in the lane. A linear trendline was plotted, yielding a function***: $y = -6729.1x^2 + 27004x + 2971.1$*** where y corresponds to the intensity and x corresponds to the amount of TMR-Ub.

TMR-Ub (ng)	Intensity (A. U.)			
	Mono	**Di**	**Tri**	**Total**
125	4429.983	947.648	161.778	5539.409
250	7425.640	1758.205	253.435	9437.280
500	12554.589	2946.276	587.406	16088.271
1000	16646.397	4612.64	1118.255	22377.292
2000	21871.033	6754.296	1573.033	30198.362

Table ST5: Raw Data- analysis and quantification of TMR-Ub in HeLa cells on an SDS-PAGE gel shown in Figure S3(B). Due to the presence of background noise in the lanes, sum of intensities from four distinct areas in each lane were considered for the analysis, namely Higher Molecular Weight Conjugates (<30 KDa), Nuc- Ub (indicated on the gel), a band between Free-Ub and Nuc-Ub (referred to as 18 kDa band) and Free-Ub. To calculate the amount of TMR-Ub in each lane, corresponding to 125,000 cells, the final intensity was plugged into the trendline on Figure S3(D) (See caption of Table ST4). Quantity of Ub in the rotor was extrapolated by assuming 5,000,000 cells per rotor, yielding a lower limit of ~10,000 ng (when total intensity in the Control lane was used) and an upper limit of 30,000 ng (when total intensity in the Epoxo Treated lane was used).

Sample	Intensity (A. U.)				
	H. Mw. Conj.	**Nuc-Ub**	**18 KDa**	**Free-Ub**	**Total**
Epoxo Treated	15211.392	397.092	243.263	3622.054	19473.801
Control	6156.058	1726.426	174.435	2583.719	10640.638

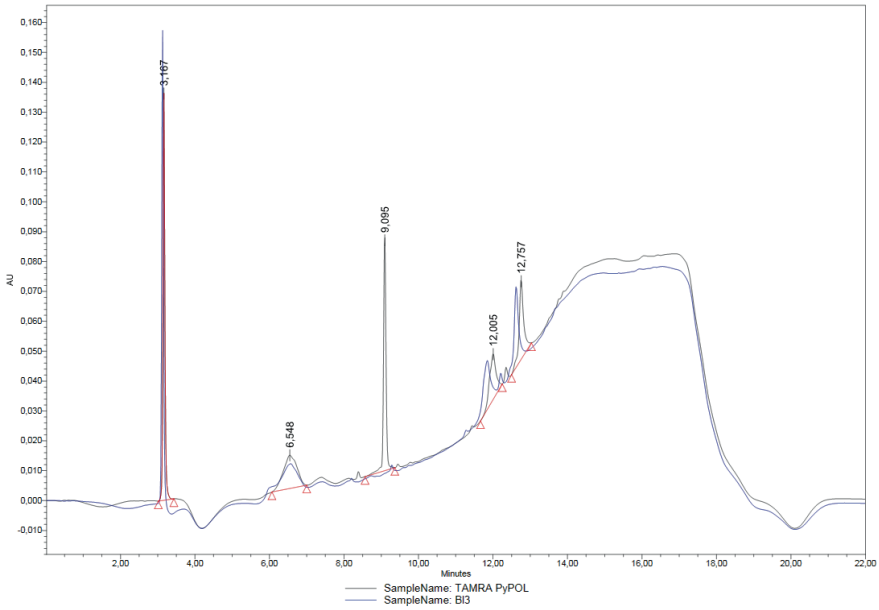


Figure S4: Analytical chromatogram of TAMRA-PyPOL (retention time of 9.1 min) overlaid with a background measurement.

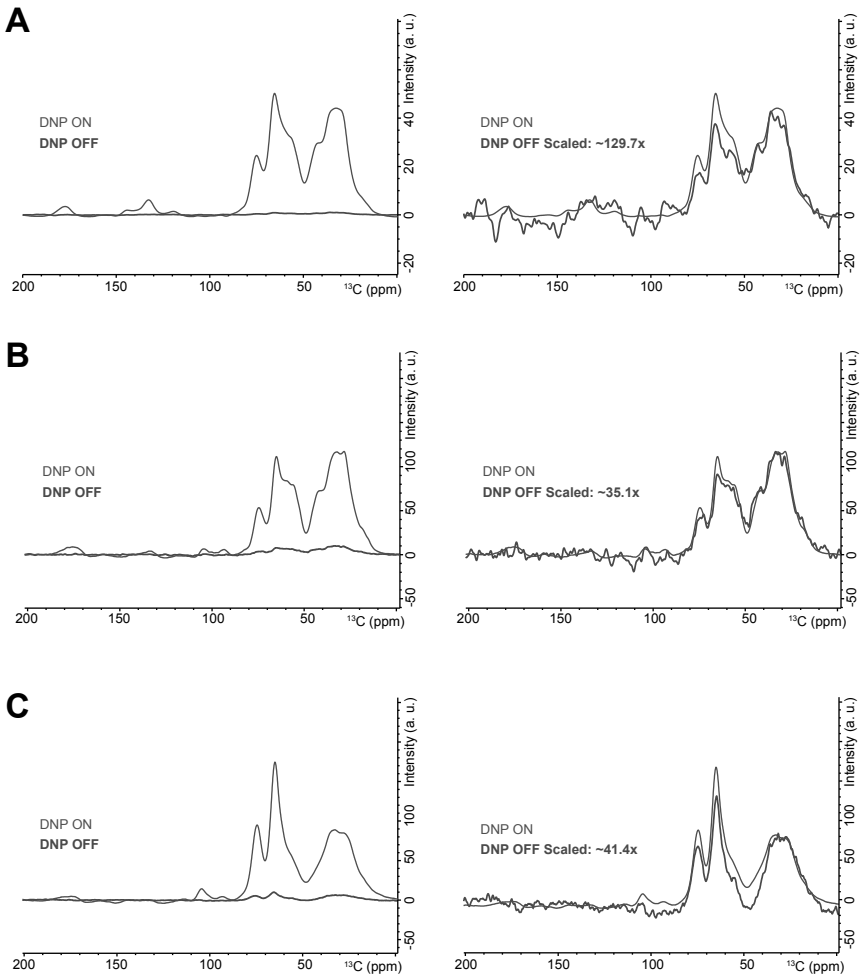


Figure S5: DNP enhancements observed in 1D ^{1}H-^{13}C cross-polarization (CP) spectra of HeLa cells containing ^{13}C, ^{15}N- Ub at 400 MHz (A) and 800 MHz (B) DNP conditions. For comparison, C) shows enhancements obtained on unmanipulated HeLa cells at 800 MHz which were not isotope labelled.

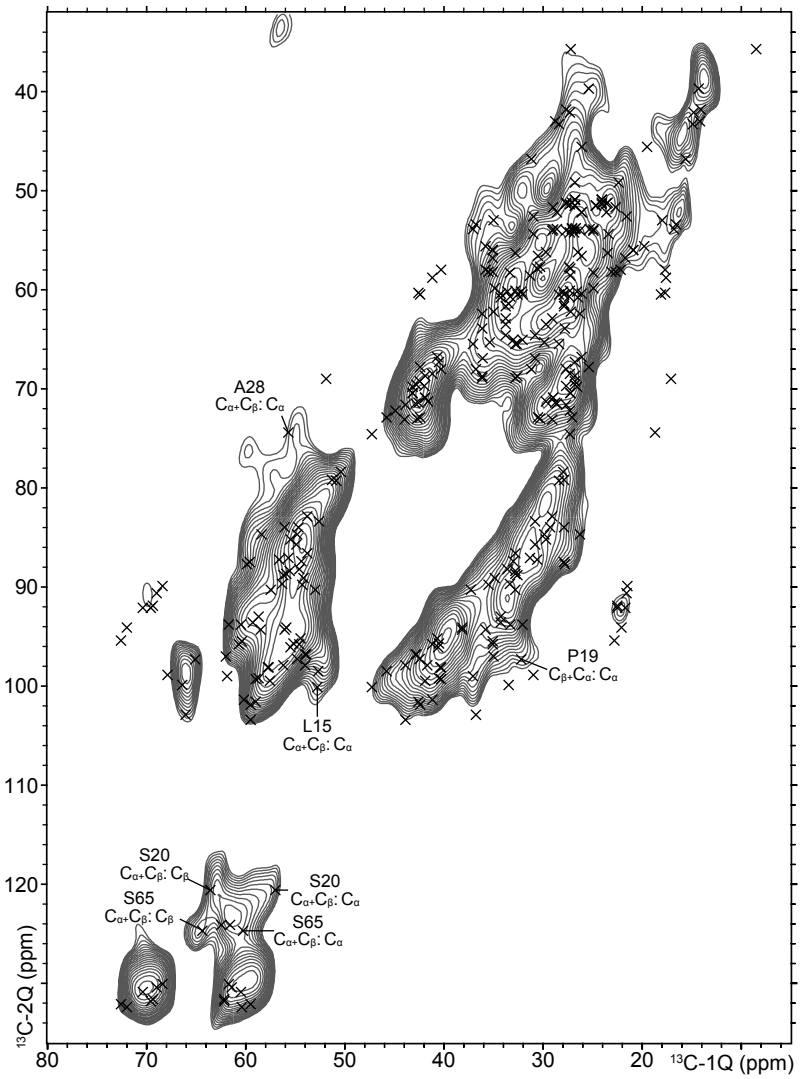


Figure S6: 2D- 2Q-1Q ^{13}C, ^{13}C correlation experiment of Ub in cell (red) overlaid with the assignments obtained in-vitro. Indicated correlations were readily identifiable because they are resolved in 3D.

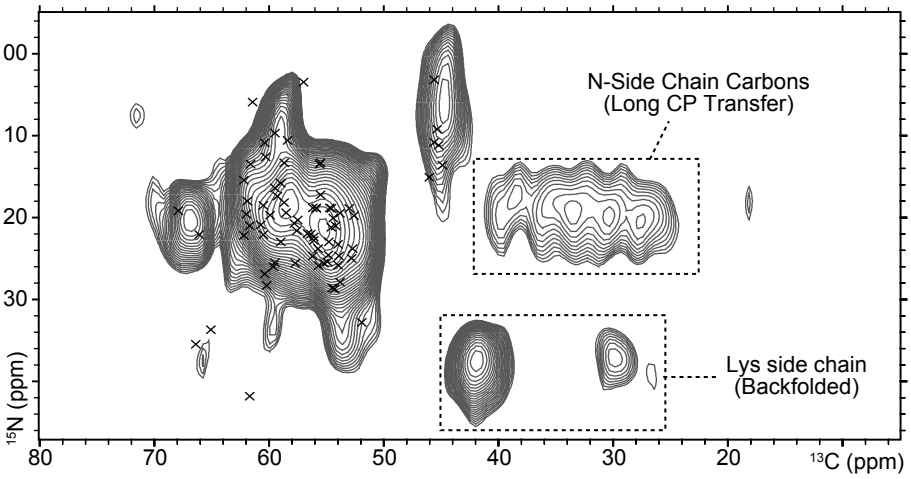


Figure S7: 2D NCA experiment of Ub in cell (red) overlaid with in-vitro assignments

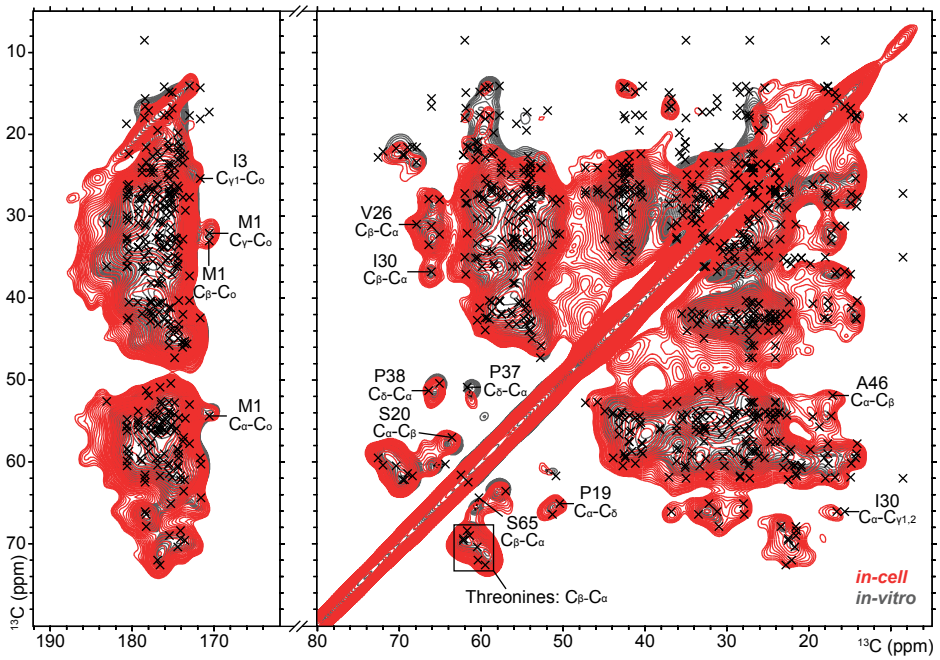


Figure S8: 2D- ^{13}C,^{13}C Proton Driven Spin Diffusion (PDSD) experiment of Ub in-cell (red) vs. in-vitro (grey) whose aliphatic region is shown in Figure 4. Black crosses relate to in-vitro NMR assignments of free Ub.

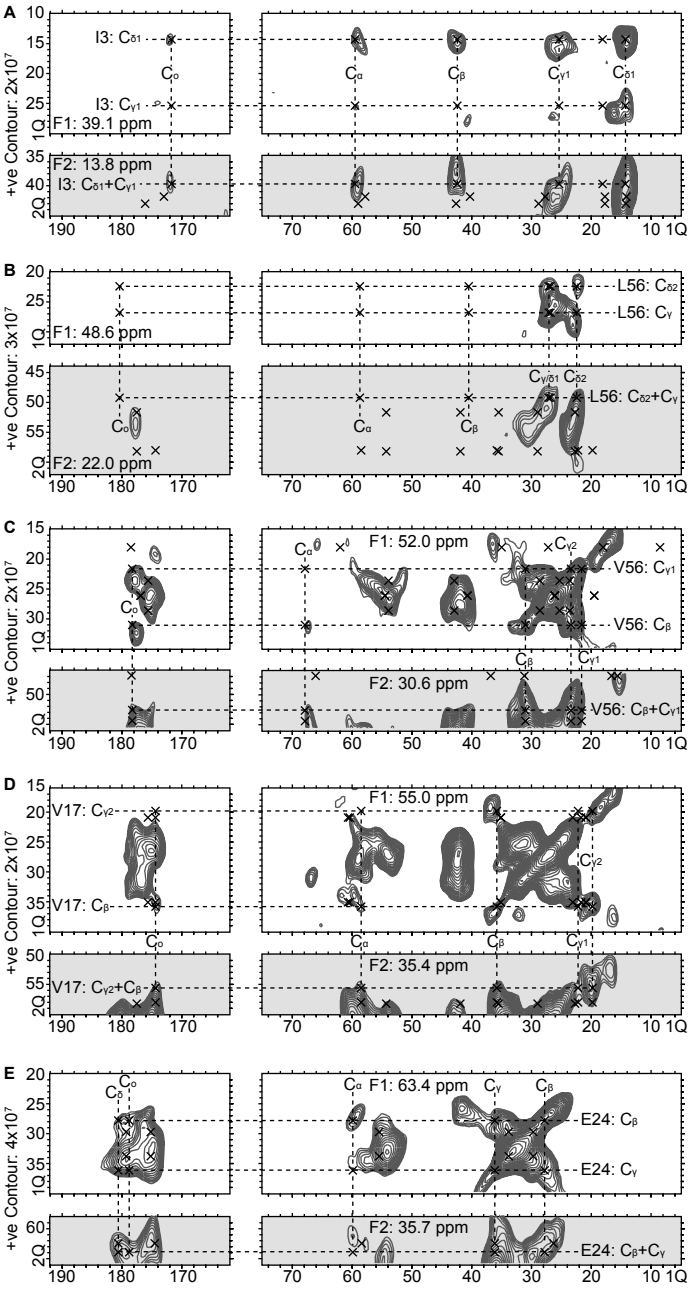


Figure S9: 2D slices (F_2, F_3- white background and F_1-F_3 grey background) from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to I3 (A), L56 (B), V56 (C), V17 (D) and E24 (E). Contour levels used here (indicated on the left) also correspond to the respective panels in Figure 5. Again, crosses relate to in-vitro NMR assignments of free Ub.

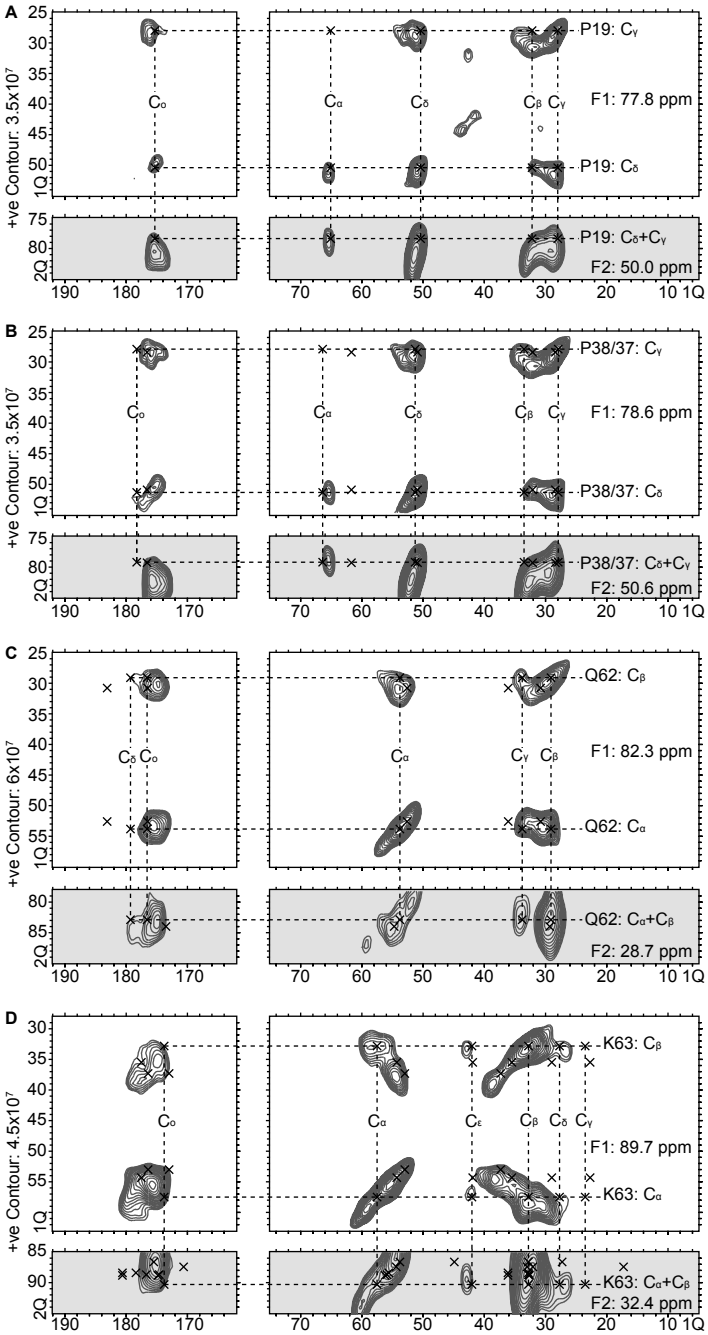


Figure S10: 2D slices (F_2, F_3- white background and F_1-F_3 grey background) from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to P19 (A), P38/37 (B), Q62 (C) and K63 (D). Contour levels used here (indicated on the left) also correspond to the respective panels in Figure 5. Again, crosses relate to in-vitro NMR assignments of free Ub.

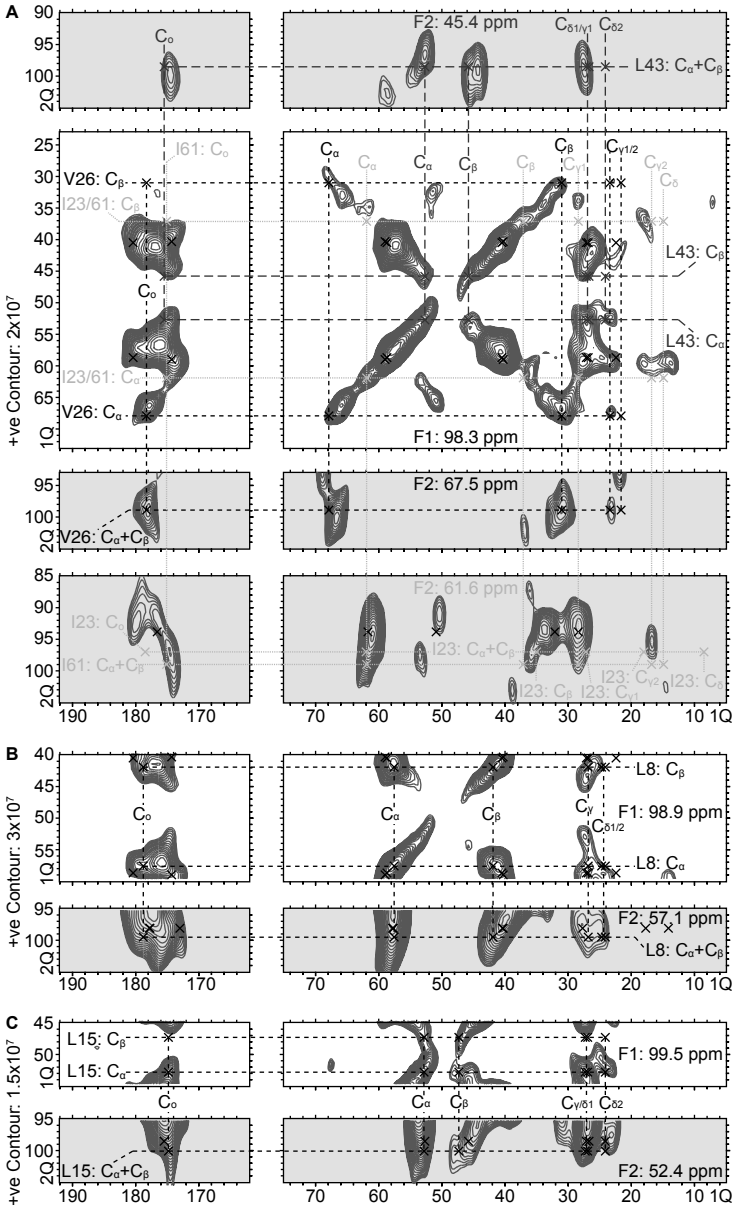


Figure S11: 2D slices (F_2, F_3- white background and F_1-F_3 grey background) from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to L43, V26 & I23/61 (A), L8 (B) and L15 (C). Contour levels used here (indicated on the left) also correspond to the respective panels in Figure 5. Again, crosses relate to in-vitro NMR assignments of free Ub.

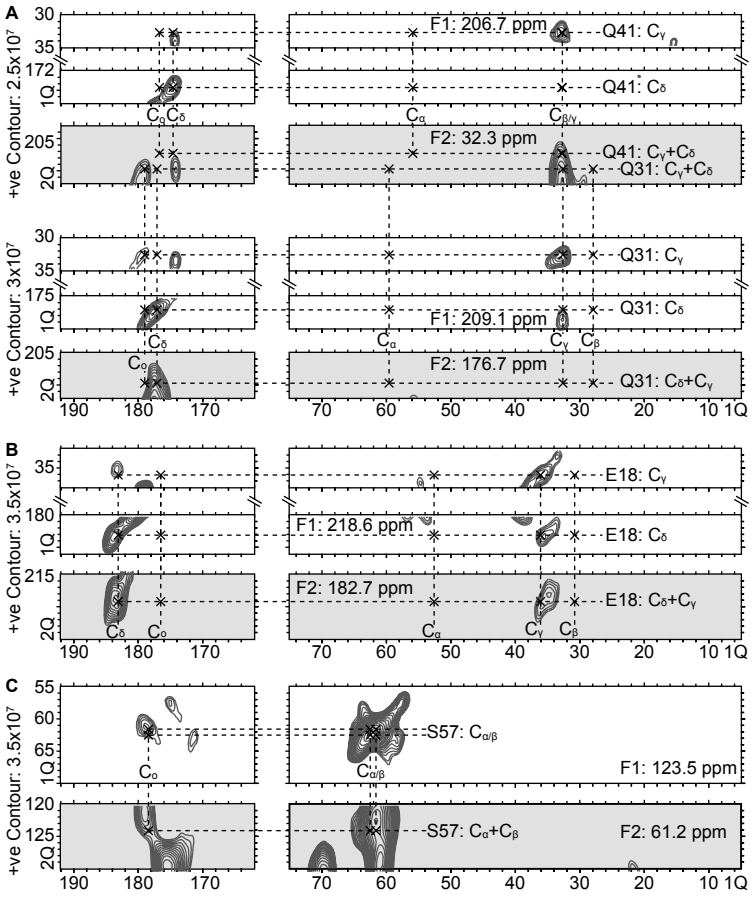


Figure S12: 2D slices (F_2, F_3- white background and F_1-F_3 grey background) from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to Q41 & Q31 (A), E18 (B) and S57 (C). Contour levels used here (indicated on the left) also correspond to the respective panels in Figure 5. Again, crosses relate to in-vitro NMR assignments of free Ub.

References

1. Li, C. *et al.* Magnetic Resonance Spectroscopy as a Tool for Assessing Macromolecular Structure and Function in Living Cells. *Annu. Rev. Anal. Chem.* **10**, 157–182 (2017).
2. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106–109 (2009).
3. Hänsel, R., Luh, L. M., Corbeski, I., Trantirek, L. & Dötsch, V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angewandte Chemie (International ed. in English)* **53**, 10300–10314 (2014).
4. Luchinat, E. & Banci, L. In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. *Acc. Chem. Res.* **51**, 1550–1557 (2018).
5. Reckel, S., Lopez, J. J., Löhr, F., Glaubitz, C. & Dötsch, V. In-Cell Solid-State NMR as a Tool to Study Proteins in Large Complexes. *ChemBioChem* **13**, 534–537 (2012).
6. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4863–8 (2012).
7. Ward, M. E. *et al.* In Situ Structural Studies of Anabaena Sensory Rhodopsin in the E. coli Membrane. *Biophys. J.* **108**, 1683–1696 (2015).
8. Baker, L. A. *et al.* Combined 1 H-Detected Solid-State NMR Spectroscopy and Electron Cryotomography to Study Membrane Proteins across Resolutions in Native Environments. *Structure* **26**, 161-170.e3 (2018).
9. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
10. Jacso, T. *et al.* Characterization of membrane proteins in isolated native cellular membranes by dynamic nuclear polarization solid-state NMR spectroscopy without purification and reconstitution. *Angew. Chemie - Int. Ed.* **51**, 432–435 (2012).
11. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12**, 5–9 (2015).
12. Yamamoto, K., Caporini, M. A., Im, S.-C., Waskell, L. & Ramamoorthy, A. Cellular solid-state NMR investigation of a membrane protein using dynamic nuclear polarization. *Biochim. Biophys. Acta - Biomembr.* **1848**, 342–349 (2015).
13. Kaplan, M. *et al.* EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **167**, 1241-1251.e11 (2016).
14. Ni, Q. Z. *et al.* High frequency dynamic nuclear polarization. *Acc. Chem. Res.* **46**, 1933–41 (2013).
15. Albert, B. J. *et al.* Dynamic Nuclear Polarization Nuclear Magnetic Resonance in Human Cells Using Fluorescent Polarizing Agents. *Biochemistry* **57**, 4741–4746 (2018).
16. Medeiros-Silva, J. *et al.* High-resolution NMR studies of antibiotics in cellular membranes. *Nat. Commun.* **9**, 3963 (2018).
17. Frederick, K. K. *et al.* Sensitivity-Enhanced NMR Reveals Alterations in Protein Structure by Cellular Milieus. *Cell* **163**, 620–628 (2015).
18. Viennet, T. *et al.* Selective Protein Hyperpolarization in Cell Lysates Using Targeted Dynamic Nuclear Polarization. *Angew. Chemie Int. Ed.* **55**, 10746–10750 (2016).
19. Krstić, I. *et al.* Long-range distance measurements on nucleic acids in cells by pulsed EPR spectroscopy. *Angew. Chemie - Int. Ed.* **50**, 5070–5074 (2011).
20. El Oualid, F. *et al.* Chemical Synthesis of Ubiquitin, Ubiquitin-Based Probes, and Diubiquitin. *Angew. Chemie Int. Ed.* **49**, 10149–10153 (2010).
21. Theillet, F.-X. X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian

cells. *Nature* **530**, 45–50 (2016).
22. Geurink, P. P., El Oualid, F., Jonker, A., Hameed, D. S. & Ovaa, H. A General Chemical Ligation Approach Towards Isopeptide-Linked Ubiquitin and Ubiquitin-Like Assay Reagents. *ChemBioChem* **13**, 293–297 (2012).
23. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
24. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chemie Int. Ed.* **52**, 10858–10861 (2013).
25. van der Cruijsen, E. A. W. *et al.* Biomolecular DNP-Supported NMR Spectroscopy using Site-Directed Spin Labeling. *Chem. - A Eur. J.* **21**, 12971–12977 (2015).
26. Schanda, P., Kupče, Ē. & Brutscher, B. SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
27. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *J. Magn. Reson.* **142**, 97–101 (2000).
28. Hohwy, M., Rienstra, C. M., Jaroniec, C. P. & Griffin, R. G. Fivefold symmetric homonuclear dipolar recoupling in rotating solids: Application to double quantum spectroscopy. *J. Chem. Phys.* **110**, 7983 (1999).
29. Baldus, M., Petkova, A. T., Herzfeld, J. & Griffin, R. G. Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **95**, 1197–1207 (1998).
30. Weingarth, M., Demco, D. E., Bodenhausen, G. & Tekely, P. Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem. Phys. Lett.* **469**, 342–348 (2008).
31. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–7 (2015).
32. Seidel, K., Etzkorn, M., Schneider, R., Ader, C. & Baldus, M. Comparative analysis of NMR chemical shift predictions for proteins in the solid phase. *Solid State Nucl. Magn. Reson.* **35**, 235–42 (2009).
33. Narasimhan, S. *et al.* Rapid prediction of multi-dimensional NMR data sets using FANDAS. in *Methods in Molecular Biology* **1688**, 111–132 (2018).
34. Pickart, C. M. & Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr. Opin. Chem. Biol.* **8**, 610–616 (2004).
35. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).
36. Kniss, A. *et al.* Chain Assembly and Disassembly Processes Differently Affect the Conformational Space of Ubiquitin Chains. *Structure* **26**, 249-258.e4 (2018).
37. Kaiser, S. E. *et al.* Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods* **8**, 691–696 (2011).
38. Dantuma, N. P., Groothuis, T. A. M., Salomons, F. A. & Neefjes, J. A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *J. Cell Biol.* **173**, 19–26 (2006).
39. Mance, D., Weingarth, M. & Baldus, M. Solid-state NMR on complex biomolecules: Methods and applications. in *Modern Magnetic Resonance* 487–503 (Springer International Publishing, 2018). doi:10.1007/978-3-319-28388-3_33
40. Heise, H. *et al.* Molecular-level secondary structure, polymorphism, and dynamics of full-length α-synuclein fibrils studied by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15871–15876 (2005).
41. Renault, M., Cukkemane, A. & Baldus, M. Solid-state NMR spectroscopy on complex

biomolecules. *Angew. Chem. Int. Ed. Engl.* **49**, 8346–57 (2010).
42. Weingarth, M. & Baldus, M. Solid-state NMR-based approaches for supramolecular structure elucidation. *Acc. Chem. Res.* **46**, 2037–2046 (2013).
43. Ogino, S. *et al.* 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**, 10834–10835 (2009).

Acknowledgements

We are indebted to Professor Paul Tordo and Dr. Olivier Ouari (Aix-Marseille Université, Marseille) for providing AMUpol and PyPol-MTSSL. We thank Mark Daniëls for technical support as well as helpful discussions, along with Suzan Leemans for the latter. We thank Hugo van Ingen and Rolf Boelens for providing access to the 900 MHz solution-state NMR instrument. We are grateful to Dr. Philipp Selenko and his group (Weizmann Institute) for sharing their expertise with protein delivery into eukaryotic cells using electroporation. This work was funded in part by the Netherlands Organisation for Scientific Research (NWO) (Graduate Program; project 022.005.029, Netherlands' Magnetic Resonance Research School, NMARRS, a VICI grant to H.O. as well as grants 700.26.121 and 700.10.443 to M.B.) and iNEXT (project number 653706), a Horizon 2020 program of the European Union.

Chapter 3:

The effect of proteasomal inhibition on Ubiquitin residues as seen by solid-state NMR spectroscopy

This chapter forms part of a manuscript in preparation with the same title

Siddarth Narasimhan, Alessandra Lucini Paioni, Stephan Scherpe, Gert Folkers, Huib Ovaa and Marc Baldus

Abstract

Inhibiting the proteasome has adverse effects on the cellular ubiquitin (Ub) pools, leading to histone deubiquitination and a subsequent increase in the amount of ubiquitinated proteins in the cytoplasm. Many studies on proteasomal inhibition have focused on either the characterization of the resulting polyubiquitin chains and/or the subsequent cellular response. Using a DNP-ssNMR approach, we studied the changes occurring on ubiquitin on a residue-specific level. Next to observing an increase in polyubiquitin linkages, we find spectroscopic evidence for a large-scale conformational heterogeneity in the flexible parts of the protein along with structural changes and post-translational modifications on the K48 residue of Ub.

Introduction

Ub is a protein that can covalently be linked to a variety of protein substrates in the cells, mediated by the E1, E2 and E3 classes of ubiquitin activating, conjugating and ligating enzymes, respectively[1,2]. Despite its small size (8.5 kDa), Ub is involved in a complex set of cellular functions. Ubiquitination is a process by which a protein becomes covalently attached by a peptide bond to the C-terminal glycine (G76) residue of Ub via any of its internal Lysines or the N-terminal residue. In the enzymatic cascade leading up to a protein being ubiquitinated, the E1 enzyme first activates Ub by sequestering free Ub and adenylating the C-terminus. Next, it transfers the Ub to an active site of the E2 class of enzymes, which act as a carrier to transfer Ub to the E3 class of enzymes that finally mediate the attachment of Ub to the substrates. E3 class of enzymes are substrate specific and hence are the most diverse of all the three enzyme types (Figure 1). As Ub is involved in key regulatory roles in the cell (discussed below) a feedback mechanism exists to detach them from their substrates when needed. This is carried out by an additional diverse class of enzymes called deubiquitinating enzymes (DUBs) or Ub specific proteases (USPs) that detach conjugated Ub molecules by cleaving the isopeptide linkage[3,4]. Together, the four enzymes regulate the different populations of Ub (free and conjugated pools in both, the cytoplasm and the nucleus).

When a single Ub molecule is bound to the substrate, the substrate is referred to as being mono-ubiquitinated (Figure 1). Subsequently, the substrate conjugated Ub can be further ubiquitinated through any of the 7 internal Lysines (K6, K11, K27, K29, K33, K48 & K63) or the N-terminal residue (M1) giving rise to diubiquitin and subsequently longer polyubiquitin chains[5,6]. The concomitant addition of Ub molecules does not necessarily have to occur on the same position as the previous ubiquitination, which gives rise to the so-called branched chains[7]. Often, various

Ub molecules can covalently attach to multiple sites on the substrates resulting in multi-monoubiquitination[5]. The aforementioned types of Ubiquitin modifications are summarized in Figure 1. Note that the different types of ubiquitination or the different linkage types and lengths of polyubiquitin chains result in different overall structures[8,9]. As a result, they can regulate various cellular functions. For example, nuclear proteins are mostly monoubiquitinated, while K63- linked polyubiquitin chains trigger DNA damage response[10,11] and K11- or K48- linked polyubiquitin chains target the substrates for degradation by the proteasome[5,6].

The proteasome is a large protein machinery consisting of a catalytic core (20s subunit) and two caps (19s subunits). The proteasomal cap is known to recognize K11- and K48- linked polyubiquitin chains, whereupon the poly-Ub conjugated protein would be directed into the catalytic core of the 20s subunit for degradation. The majority of eukaryotic proteins are thought to be degraded by this intricate system known as the Ubiquitin Proteasome System (UPS)[12]. Disrupting the UPS has adverse effects on the cell due to aggregation of proteins, eventually leading to cell death[13].

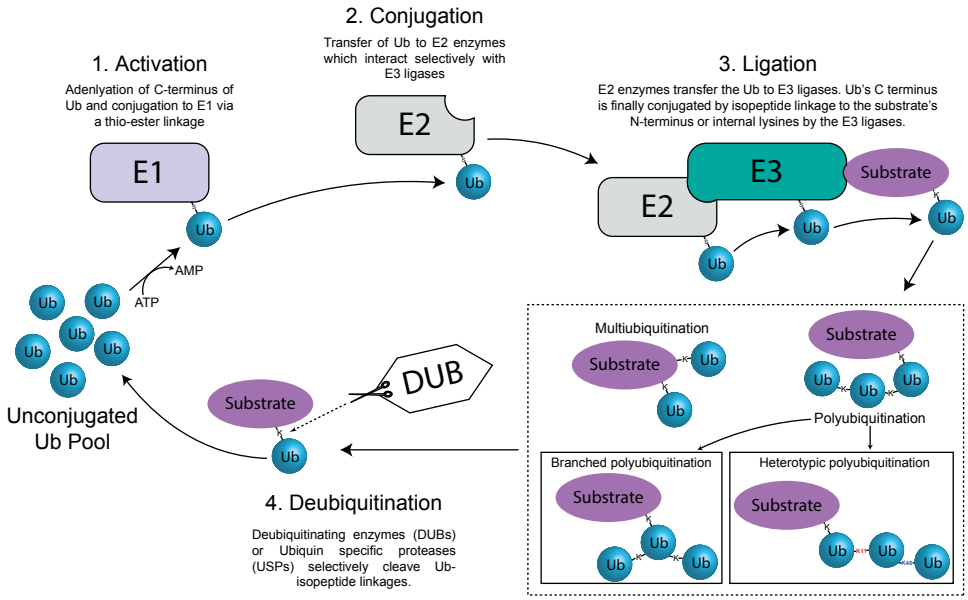


Figure 1: Protein ubiquitination by E1, E2 & E3 enzymes and finally the cleavage of the conjugated Ub by DUB enzymes. Examples of poly/multi- ubiquitination complexes are shown within dashed boxes.

Inhibition of the proteasome leads to accumulation of (poly)ubiquitinated proteins in the cells along with a drastic reduction in the monoubiquitinated nuclear

Ub population (See Chapter 2 and Ref[14]). Previously, immunoblotting and several mass spectrometry (MS) approaches have been employed to characterize the chain linkage types in these (poly)ubiquitin chains. Combination of both approaches and, in particular, MS has provided a wealth of novel insights into the chain formation and their associated cellular functions even when the detected chains are present in very low abundance[15,16]. Additionally, MS studies have revealed interactomes associated with specific chains[17] under normal cellular homoeostasis without proteasome inhibition. Together, MS approaches have helped to construct an overall view of the effect proteasome inhibition upon Ub-chain accumulation and have elucidated a multitude of interactions that the chains are capable of engaging in. However, these studies have been carried out in vitro where cell lysis and further sample treatments may lead to a loss of labile protein modifications. Moreover, such studies have been unable to track the structural changes of Ub residues that may accompany chain formation. As highlighted in Chapter 1, one of the goals of in-cell NMR is to enable detection of even the labile PTMs that may occur in cells at very low abundance. Here, we attempted to obtain direct insight into the structural and chemical changes during Ub chain formation due to the effect of proteasome inhibition inside human cells.

Materials and Methods

Expression and purification of Ub

Growth and purification methods were identical to the ones described in chapter 2. To obtain Lysine labelled Ub (K-Ub), M9 medium which are not isotope labelled, supplemented with 100 mg/L of $^{13}C_6$- and $^{15}N_2$ L-lysine hydrochloride and 200 mg/L of all the other 19 amino acids was used for expression.

Preparation of K48 labelled Ub using Solid Phase Peptide Synthesis

Solid phase peptide synthesis (SPPS) was performed in syringes using a Syro II synthesizer (Multisyntech GmbH, Witten, Germany). Peptides were synthesized using standard 9-fluorenylmethoxycarbonyl (Fmoc) based solid phase synthesis as described before[18] with preloaded Fmoc amino acid trityl resin (0.2 mmol/g, Rapp Polymere GmbH) on a 25 µmol scale. Swelling of the resin was achieved by addition of 1.25 mL N-methyl-2-pyrrolidone (NMP) for 5 min (x2). Fmoc was removed by treatment with 20% piperidine in NMP thrice for 3, 5 & 5 min, followed by five washing steps of NMP. Amino acids were coupled by applying a fourfold excess in the presence of 4 equivalents PyBOP and 8 equivalents DiPEA in NMP for 45 min, followed by three washing steps with NMP. After completion of all coupling cycles, the resin was washed with diethyl ether and dried under vacuum. The ubiquitin

peptide sequences were synthesized on resin following the procedures as described before[18].

For K48 incorporation, $^{13}C_6{}^{15}N_2$ Lysine was protected with as previously reported[19]. The peptide was synthesized on a 10 μmol scale using Fmoc-$^{13}C_6{}^{15}N_2$-Lys for position K48. The PEG spacer and, afterwards, the biotin were coupled to the N-terminus and the purification was achieved as described above. MS ES+ (amu) calc: 8925.8. See following section for LC-MS protocol.

LC-MS measurements

LC-MS measurements were performed on a Waters LCT-premier mass spectrometer system equipped with a Waters 2795 Separation Module (Alliance HT), Waters 2996 Photodiode Array Detector (190-750 nm) Phenomenex Kinetex C18 (2.1x50, 2.6 μm, flow rate= 0.8 mL/min, runtime= 6 min, column T= 40°C. Gradient: 0 – 0.5 min: 95% A, 5% B; 0.5 – 4 min: 95 – 5% A, 5 – 95% B; 4 – 5.5 min: 5% A, 95% B) and LCT™ ESI-Mass Spectrometer. Samples were run using 2 mobile phases: A = 1% CH_3CN, 0.1% formic acid in water and B = 1% water and 0.1% formic acid in CH_3CN. Data processing was performed using Waters Mass Lynx Mass Spectrometry Software 4.1 (deconvolution with Maxent1 function). High resolution mass spectra were recorded on a Waters XEVO-G2 XS Q-TOF mass spectrometer equipped with an electrospray ion source in positive mode (source voltage 30 V, desolvation gas flow 900 L/hr, temperature 250°C) with resolution R = 26000 (mass range m/z = 50-2000 m/z, capillary Voltage 1.2 kV) and 200 pg/μL Leu-Enk (m/z = 556.2771) as a "lock mass".

Preparation of cells for DNP Solid-State NMR experiments

For ssNMR sample preparation, HeLa cells were grown as 2D cultures in DMEM Glutamax medium supplemented with 10% FBS without the addition of antibiotics in a 37°C incubator. 8 million cells were electroporated with (^{13}C-, ^{15}N-labelled ^{13}C-, ^{15}N- K- labelled / ^{13}C-, ^{15}N- K48- labelled-) Ub at a final concentration of 1200 μM in the electroporation buffer. See Chapter 2 for detailed electroporation protocol. After 5 h of recovery, cells were optionally treated with 2 μM Epoxomicin for 6 h.

Cells were harvested by washing twice with phosphate buffered saline (PBS) followed by trypsin treatment. The cell pellet was collected and washed twice with PBS. The cells were resuspended in DNP buffer with similar composition as used in Chapter 2, with the only difference being the use of ^{2}H- enriched glycerol instead of the ^{2}H- enriched, ^{13}C- depleted glycerol. In some cases (cells with uniformly ^{13}C-, ^{15}N-labelled Ub) where no glycerol was used, D_2O was substituted for glycerol to account for the same buffer volume as in the other samples.

DNP Solid-State NMR experiments

We conducted solid-state NMR experiments under low temperature (100 K) DNP conditions using a 3.2 mm triple-resonance (^{1}H, ^{13}C, and ^{15}N) magic-angle-spinning (MAS) probe head in a static magnetic field of 9.4 T, corresponding to proton/electron resonance frequencies of 400 MHz/263 GHz (Bruker BioSpin). All spectra were recorded at an MAS rate of 8 kHz using a 84 kHz SPINAL-64 proton decoupling[20].

Solid-state NMR experiments at ambient temperatures

We conducted solid-state NMR experiments at an ambient temperature (280 K) using a 3.2 mm triple-resonance (^{1}H, ^{13}C, and ^{15}N) magic-angle-spinning (MAS) probe head in a static magnetic field of 9.4 T, corresponding to a proton resonance frequency of 400 MHz (Bruker BioSpin). All spectra were recorded at an MAS rate of 13 kHz using a 84 kHz SPINAL-64 proton decoupling[20].

1D- ^{1}H-^{13}C and ^{1}H-^{15}N Cross-Polarization experiments

The 1D H-C cross-polarization (CP) spectra in DNP samples were recorded using a 2 s recycle delay and a 10 ms acquisition time. Cross-polarization was achieved using a contact time of 50 μs. 200 Hz line-broadening was applied prior to Fourier transformation.

The 1D H-C/ H-N cross-polarization (CP) spectra in conventional ssNMR of Homoarginine and Homocitrulline were recorded using a 2 s recycle delay and a 10 ms acquisition time. Cross-polarization was achieved using a contact time of 1000 μs. 50 Hz and 80 Hz line-broadening were applied prior to Fourier transformation for the H-C and H-N CP experiments respectively.

Proton-driven spin diffusion (PDSD) experiments

For the 2D proton-driven spin diffusion (PDSD) experiments of cells containing uniformly ^{13}C-, ^{15}N- Ub, a contact time of 100 μs was used for H-C cross polarization and a mixing time of 30ms was used. Acquisition times were 15 ms (900 data points used- 405.3 ppm spectral width) and 7.2 ms (408 data points- 280 ppm spectral width) for the direct and indirect dimensions, respectively. 200 scans were used for the control sample which was not treated with epoxomicin. For the epoxomicin treated sample, two separate spectra were recorded with 512 scans and 320 scans, which were then added to match the signal to noise of the control sample. A recycle delay of 2s was used. The 2D spectra were processed using a 0.33 π shifted sine squared window function on both dimensions. For spectral analysis, ssNMR assignments were obtained in the similar manner as described in Chapter 2.

^{13}C DQ-SQ correlation experiment

The 2D ^{13}C DQ-SQ correlation spectrum, on cells containing uniformly ^{13}C-, ^{15}N- labelled Ub, resuspended in DNP buffer containing ^{2}H-enriched glycerol, was obtained using 10 blocks of SPC5[21] mixing (2.5 ms) for generation and subsequent reconversion of double quantum coherences. A contact time of 50 μs was used for H-C cross polarization and continuous wave (CW) ^{1}H decoupling was applied during the SPC5 sequence with a field strength of 85 kHz. Acquisition times were 10 ms (640 data points- 307 ppm spectral width) and 1.6 ms (88 data points- 280 ppm spectral width) for the SQ direct and DQ indirect dimensions, respectively. 800 scans were acquired, with a recycle delay of 2 s. We used a forward linear prediction (8 coefficients) in the DQ dimension and the 2D spectra were processed using a 0.33 π shifted sine squared window function in both dimensions.

^{15}N-^{13}C correlation experiments

^{15}N-^{13}C-correlation experiments for cells with uniformly ^{13}C-, ^{15}N- Ub were performed using a ^{1}H-^{15}N CP step of 900-1100 μs and a SPECIFIC-CP[22] ^{15}N-^{13}C step of 2.5 ms. On resonance frequencies were set to 118.77 ppm and 47.98 ppm for ^{15}N and ^{13}C dimensions, respectively. A recycle delay of 2 s was used, and 1k scans were acquired for control cells which were not treated with Epoxomicin and 4k scans were acquired for cells treated with Epoxomicin. Acquisition times were set to 10 ms (500 data points used- 497 ppm spectral width) and 4.1 ms (30 data points- 90 ppm spectral width) for the direct and indirect dimensions, respectively. The 2D spectra were processed using a 0.33 π shifted sine squared window function in both dimensions.

^{15}N-^{13}C-correlation experiments for cells with K- Ub (^{13}C-, ^{15}N- K labelled) and K48-Ub (^{13}C-, ^{15}N- K labelled) were performed using a ^{1}H-^{15}N CP step of 900 μs and a SPECIFIC-CP[22] ^{15}N-^{13}C step of 3.5 ms. The on resonance frequencies were set to 75 ppm and 50 ppm for ^{15}N and ^{13}C dimensions, respectively. A recycle delay of 2 s was used, and 8k scans were acquired. Acquisition times for the K-Ub samples were set to 10 ms (1024 data points- 497 ppm spectral width) and 2.2 ms (29 data points- 160 ppm spectral width) for the direct and indirect dimensions, respectively. Acquisition times for K48-Ub samples were 10 ms (1024 data points- 497 ppm spectral width) and 1 ms (15 data points- 180 ppm spectral width) for the direct and indirect dimensions, respectively. Prior to Fourier transformation, 150 Hz line-broadening was used in both dimensions.

Homocitrulline synthesis

L-lysine monohydrochloride (1.1 g, 6 mmol) and sodium hydrogen carbonate (0.5 g, 6 mmol) were dissolved in water (6 mL). Copper sulphate pentahydrate

(1.5 g, 6 mmol) was added and the mixture was heated to reflux for 1 hour. Urea (2.4 g, 40 mmol) was added and the mixture was heated to reflux for 4 hours. The precipitated copper-homocitrulline was filtered and washed with water (3x 50 mL). The precipitate, water (20 mL) and sodium sulfide (2.4 g, 30 mmol) were heated to reflux for 1 hour. The precipitated copper sulide was filtered off and the solvent was evaporated. The product was obtained after recrystallization from methanol and dried under high vacuum. Calculated mass for $[C_7H_{15}N_3O_3+H^+]$: 190.1192.

Homoarginine synthesis

Nα-(tert-Butoxycarbonyl)-L-lysine (1 g, 4 mmol) was dissolved in *n*-butanol (10 mL). After addition of acetic acid (250 µL, 4 mmol) the mixture was heated to 90°C. Cyanamide (170 mg, 4 mmol) in *n*-butanol (1 mL) was added dropwise. The reaction was heated to reflux for 3 hours, let cool to room temperature, and the solvent was evaporated under high vacuum. The product was purified by flash column chromatography (95% $CHCl_3$, 5% MeOH) and dried under high vacuum. Calculated mass for $[C_{12}H_{24}N_4O_4 + H^+]$: 289.1826.

Chemical shift prediction of crystal structures of Ub- K48 analysis

To compare experimental chemical shifts to previous Ub structures, the Ub structures mentioned in Ref[23] were used (PDB IDs: 1NBF, 1AAR, 1CMX, 1P3Q, 1S1Q, 1TBE, 1UZX, 1XD3, 1YD8, 2AYO, 2C7M, 2C7N, 2D3G, 2FCQ, 2FID, 2FIF, 2G45, 1WR6, 1WRD, 1OTR, 1WR1, 1YIW, 1UBI, 1UBQ & 1F9J). The chemical shifts relating to NMR signatures of Ub in the aforementioned structures were predicted using the SHIFTX2[24] software package. The Cα chemical shifts of these residues were used to calculate the standard deviation, which was plotted on the crystal structure of free Ub (PDB ID: 1UBQ).

Results and Discussion

3.1 DNP enhancements as a function of the DNP buffer conditions

The experiments and the data presented in this chapter predate the DNP-ssNMR experiments presented in the previous chapter. At the time that the data presented here were collected, the optimal composition of the DNP buffer (with ^{2}H- enriched and ^{13}C depleted glycerol) was yet to be discovered. Below we discuss different strategies to overcome the adverse effects of using suboptimal DNP buffer conditions. As a result, the DNP buffer composition used for the work here is slightly suboptimal. In the future, we intend to perform these experiments with samples with the optimal DNP buffer and obtain more resolved spectral information. Thus, the data presented here is deemed as being preliminary.

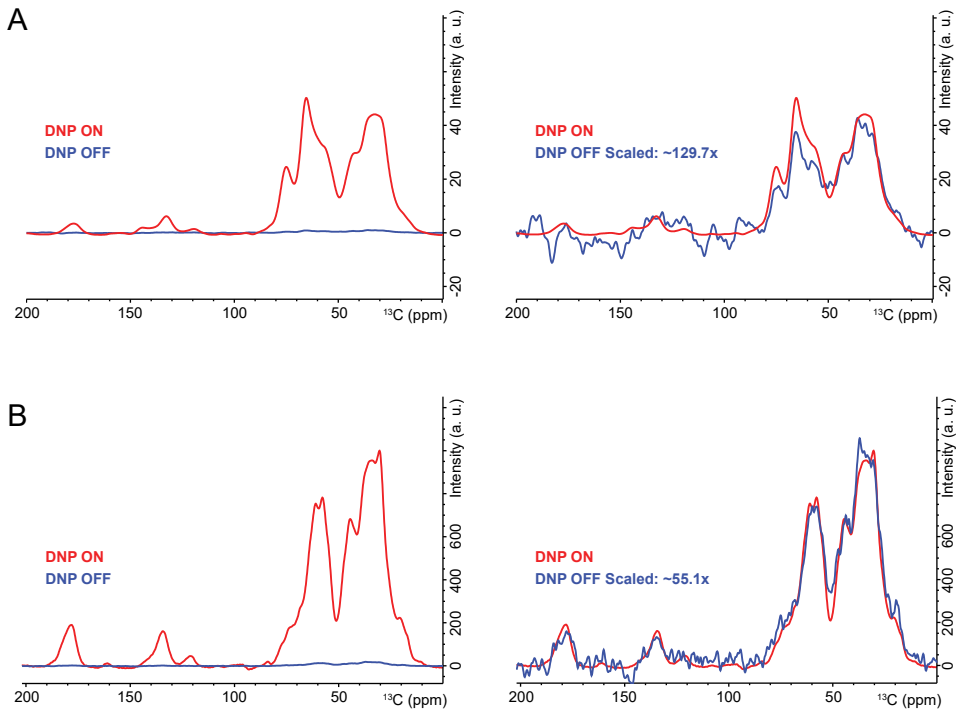


Figure 2: DNP enhancements in the presence (A) and absence (B) of ^{2}H- glycerol in the DNP buffer. Glycerol peaks occur at 65 & 75 ppm.

As mentioned in the previous chapter, cells prepared for use in DNP-ssNMR need to be frozen immediately after resuspension in the DNP juice and packing into the rotor. As a result, radicals that do not permeate into the cells cannot be

completely washed off. Prior to using ^{2}H- enriched and ^{13}C depleted glycerol, we used ^{2}H- enriched glycerol. Despite the low abundance of ^{13}C in this glycerol preparation, the NMR signal of glycerol peaks was still higher than the protein signals in the cells (Figure 2A, Supporting Information, Table ST1). due to high concentration of glycerol and their proximity to the bulk of the active radicals.

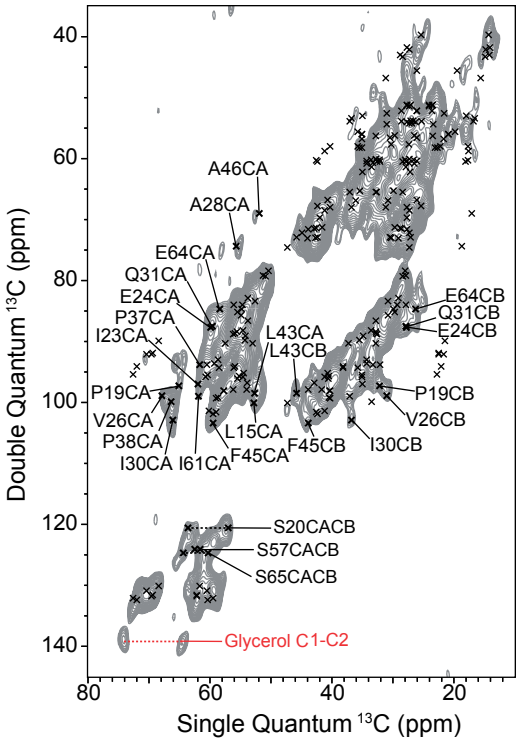


Figure 3: ^{13}C DQ-SQ correlation spectrum of Ub in HeLa cells. The DNP buffer contains ^{2}H- enriched glycerol. Overlaid as black crosses are the predictions for the spin-diffusion spectra for Ub, where some resonances in resolved regions of the spectra are highlighted. Highlighted in red is the correlation for Glycerol C1 and C2 resonances which appears purely through high abundance and larger DNP enhancements.

Therefore, the ^{13}C-^{13}C correlated spin-diffusion experiments would be dominated with the glycerol peaks. Likewise, glycerol signals could not be completely suppressed in 2D ^{13}C double quantum- single quantum (DQ-SQ) correlation experiments, where natural abundant ^{13}C signal intensities are reduced by a factor 10^{4} (Figure 3).On the other hand, ^{15}N correlated experiments indeed filtered out the glycerol signals, but we obtained poorer resolution due to the broader ^{15}N linewidths in comparison to ^{13}C (See Chapter 2).

To suppress the dominant glycerol resonances from the DNP buffer, we

prepared our DNP sample of HeLa cells using uniformly ^{13}C-, ^{15}N- labelled Ub in a DNP buffer devoid of any glycerol (also see Materials and Methods). The 1D ^{1}H-^{13}C CP spectrum clearly showed the absence of glycerol resonances (Figure 2B). Additionally, we observed a drop in the DNP enhancements by almost a factor one half compared to the case where glycerol was present in the buffer (Figure 2 & Supporting Information, Table ST1). Despite the low enhancement, we recorded 2D experiments (vide infra). However, we could not perform 3D experiments that would yield higher spectral resolution.

3.2 Effect of proteasome inhibition on distinguishable residues of Ub

To study the overall effect of proteasome inhibition on residues that exhibit resolved resonances in the ^{13}C-^{13}C spin diffusion spectrum, we studied ^{13}C and ^{15}N labelled Ub in HeLa cells. Despite the reduction in sensitivity (see section 3.1), we could record 2D ^{13}C-^{13}C (Figure 4) and ^{15}N-^{13}C (Figure 5) correlation experiments on HeLa cells with and without a 6h Epoxomicin treatment. In the Epoxomicin treated cells, we observed a large-scale disappearance of resonances in several regions of the spectrum which mostly includes the side chains along with some backbone correlations (Figure 4, 5). Due to the lack of spectral resolution, we could only pinpoint a few resonances with certainty (Supporting Information, Table ST2). The most likely explanation for signals disappearing from the spectrum is the heterogeneity of Ub populations brought about by the crowding and/or molecular interactions that Ub molecules engage in. Likewise, there is a lack of heterogeneity prior to epoxomicin treatment, as Ub is most prominently present in the free (unconjugated) and the histone-conjugated states[14] (See Figure 1 in Chapter 2).

The notion that increased structural heterogeneity is an important factor for the increased signal loss is further strengthened by the fact that many resonances corresponding to the residues in the protein loop regions disappear upon epoxomicin treatment. These include all proline residues, which are also absent in both the spin diffusion (Figure 4) and ^{15}N-^{13}C correlation experiment (Figure 5). The residues corresponding to these peaks are indicated on the structure of Ub (Figure 6). Intriguingly, the absence of resonances from the alpha-helix such as I23, V26 and I30 could not be explained by structural heterogeneity, as previous reports do not identify these residues as being critically involved in Ub recognition[25]. On the other hand, disappearing residues in the beta-sheets (I3, L15, I13 and I44) are hydrophobic, which would be in line with increased interactions with different binding partners upon inhibitor treatment. Indeed, I44 is a central binding hotspot in the hydrophobic patch of Ub consisting of L8, I44 and V70, which is important for recognition by the E1 activating enzymes[26].

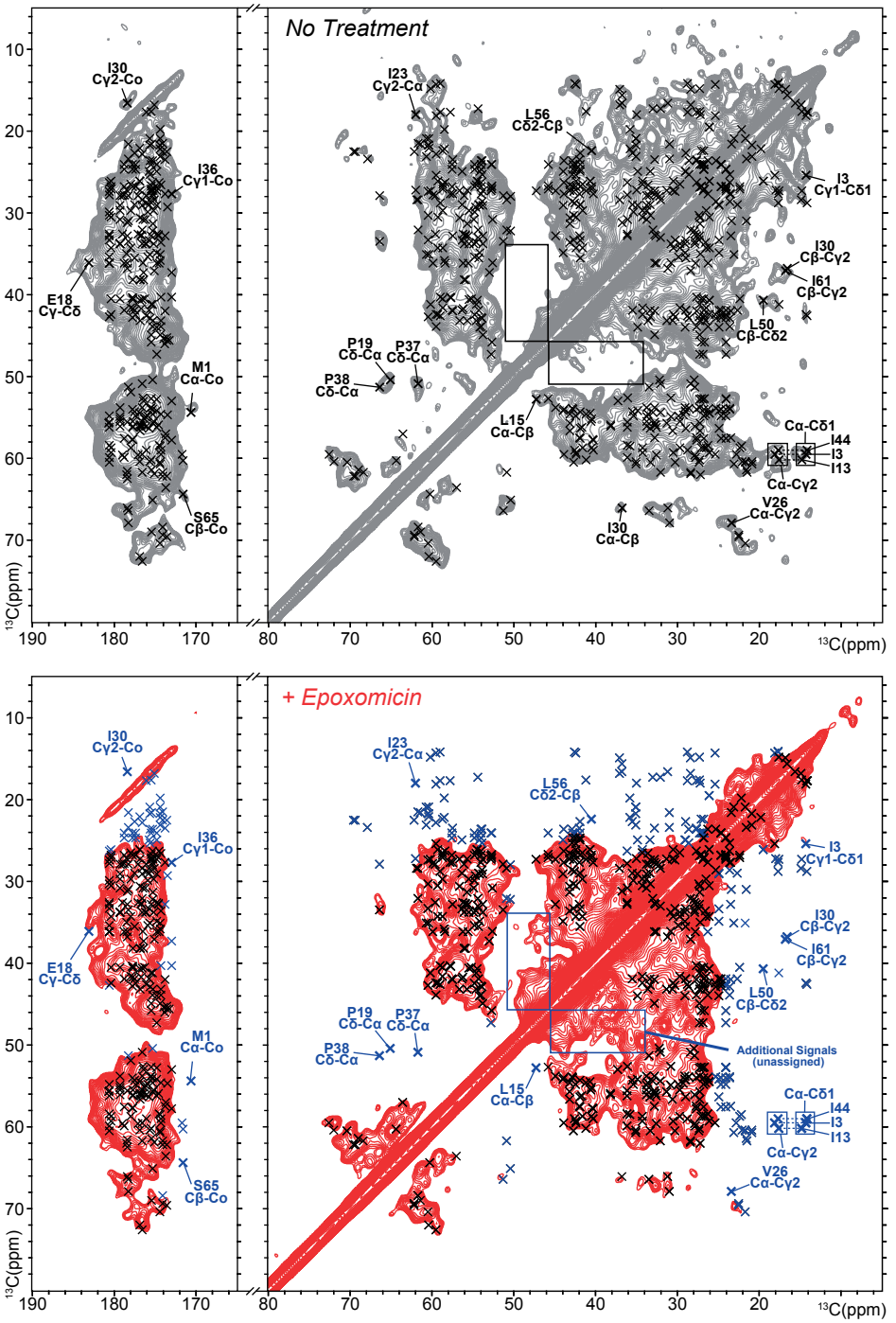


Figure 4: 2D ^{13}C-^{13}C spin-diffusion spectrum recorded in HeLa cells containing the uniformly ^{13}C- & ^{15}N-labelled Ubiquitin before (grey) and (red) after epoxomicin treatment. The peaks that disappeared upon epoxomicin treatment are shown in blue.

Moreover, M1 (Figure 4A&B, left panels- carbonyl region) & G47 (Figure 5) also showed disappearance of resonances after proteasome inhibition. Absence of M1 could be attributed to the characteristic Ub chain formation through the N terminus, and/ or conformational heterogeneity. A likely explanation for absence of G47, is its proximity to K48, which is another site of chain propagation linked to the UPS[10]. Finally, a few signals for which no assignments are currently available, appeared in the spin diffusion (Figure 4, solid box) and in the N-C spectrum (Figure 5, solid box).

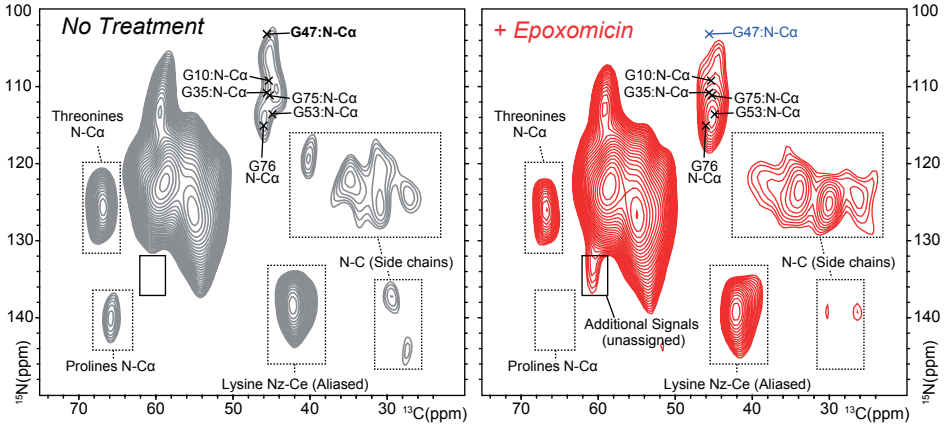


Figure 5: Comparison of the 2D aliphatic ^{15}N-^{13}C correlation spectra recorded in HeLa cells containing the U-^{13}C & ^{15}N labelled Ub before (grey) and (red) after epoxomicin treatment. The G47 and Proline signals that disappeared upon epoxomicin treatment are shown in blue.

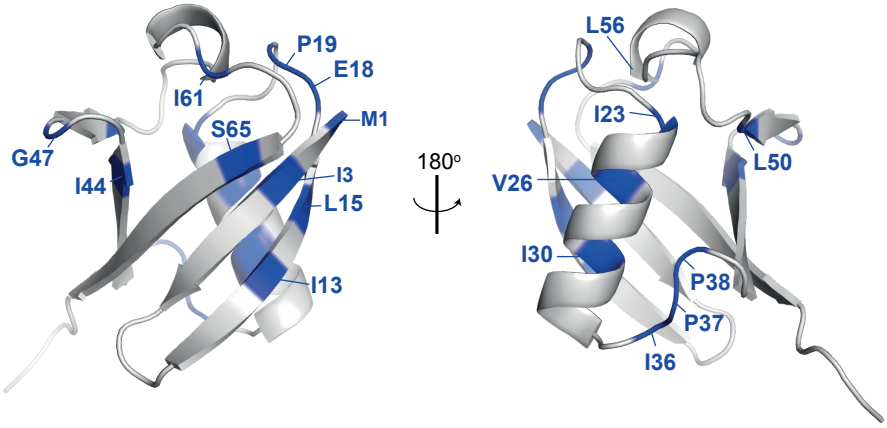


Figure 6: Residues corresponding to the resonances that disappeared upon epoxomicin treatment in the 2D ^{13}C-13C spin diffusion (Figure 4) and ^{15}N-^{13}C correlation (Figure 5) spectra plotted on the Ub structure (PDB ID: 1UBQ) in blue

Despite the changes described above, a definitive indication of chain formation could not be obtained, as none of the Lysine and C-terminal Glycine resonances could be distinguished in the 2D ^{13}C-^{13}C spin diffusion and the N-C spectra.

3.3 Effect of proteasome inhibition specifically on Lysine residues of Ub

To further zoom into the sites of chain propagation upon proteasome inhibition, we conducted ^{15}N-^{13}C correlation experiments starting with a Ub variant in which all seven lysine residues were ^{13}C-, ^{15}N- labelled (in the following referred to as K-Ub) (Supporting Information, Figure S1). We expected that upon chain formation, the large-scale chemical shift changes should occur on the side-chain nitrogen of the Lysine ($^{15}N_\zeta$) due to increased formation of isopeptide linkages with another Ub's C-terminal Glycine. Note that such an event would shift the $^{15}N_\zeta$ chemical shifts from the shielded region (~30 ppm) to the protein backbone (between 100-130 ppm) region.

Due to the inherent lower sensitivity of ^{15}N-^{13}C correlation experiments compared to ^{13}C,^{13}C 2D ssNMR, we used DNP buffer containing ^{2}H- glycerol to obtain maximum enhancements (see also Materials and Methods). While we did not observe major spectral changes readily in the 2D data set (Figure 7A), analysis of the 1D slices of the 2D spectra indicated an increase in nitrogen de-shielded $^{15}N_\zeta$-$^{13}C_\varepsilon$ correlations upon epoxomicin treatment, providing a first indication for an increase in the isopeptide linkage formation (Figure 7B). Intriguingly, from the signal intensities, it is evident that the amount of isopeptide linkages is rather low, as 8096 scans were required in the 2D experiments to clearly detect them. Note that longer proteasomal inhibition times could increase the amount of isopeptide linkages on Ub, as maximal effect is reported to occur on most Lysines at 24 hours long treatment with inhibitors[27]. We also note that the resolution in the ^{15}N dimension was not sufficient to pin-point the exact Lysine residues that were preferably engaged in chain formation upon proteasome inhibition.

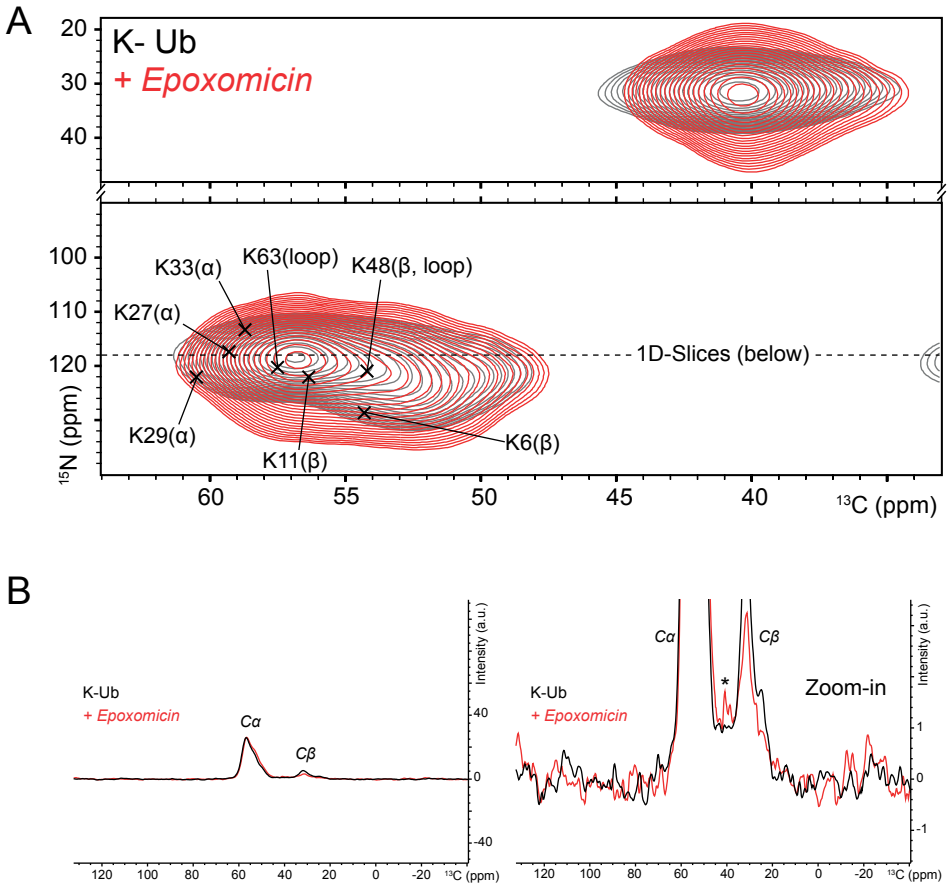


Figure 7: The effect of proteasomal inhibition on K- Ub. A) 2D- ^{15}N-^{13}C correlation experiments of K-Ub in HeLa cells before (black) and after (red) inhibitor treatment. The in-vitro backbone chemical shift (^{15}N-$^{13}C_\alpha$) assignments for Ub lysines (BMRB ID: 7111) as well as the corresponding secondary structure for each residue (α-helix, random coil, and β-strand) are included as reference. B) 1D slices extracted from the 2D spectra shown in (A) at a ^{15}N frequency representative of amide backbone resonances (δ_N ~121 ppm- indicated in panel A). While the spectrum is largely dominated by C_α resonances (δ_C ~45-65 ppm) (A), the intensity of correlations in the region δ_C ~38-48 ppm (indicated by an asterisk) is increased when the proteasome is inhibited by epoxomicin treatment (B). Such a signal could correspond to a deshielded (isopeptide linked) N_ζ-C_ε correlation.

3.4 Effect of proteasome inhibition specifically on K48 of Ub

To gain unambiguous, residue-specific insight, we subsequently concentrated on K48, which is critically involved in activating the ubiquitin proteasome system (UPS)[6,10]. We produced ^{13}C-, ^{15}N- K48 labelled Ub (K48-Ub) using solid phase peptide synthesis (SPPS[18]) as confirmed by solution state NMR (Supporting Information, Figure S2) and DNP-ssNMR (Supporting Information, Figure S3A). Upon proteasome inhibition, we observed a minor reduction of the free side-chain intensity and the occurrence of an additional correlation at δ_N- 124.5 ppm and δ_C- 44.9 ppm (Figure 8A, also see Supporting Information, Figure S3). The corresponding deshielded $^{15}N_\zeta$-$^{13}C_\varepsilon$ correlation peak (Figure 8A) is in good agreement with previous in vitro NMR studies on lysine-aspartic acid isopeptide linkage[28], suggesting the formation of an isopeptide linkage (Figure 8B) in our cellular preparations. Finally, we also observed a correlation at δ_N - 98.6 ppm and δ_C - 48 ppm (Figure 8A, also see Supporting Information, Figure S3). SsNMR experiments on synthesized homocitrulline and homoarginine (Supporting Information, Figure S4 & Table ST3) suggest that such a correlation possibly reflects carbamylated or guanidinylated lysines.

A recent study shows evidence for urea/ cyannate-induced protein carbamylation as a signal for proteasomal degradation in cultured fibroblasts[29]. Note that urea/ cyannate can spontaneously react with Lysine side chains to form homocitrulline[30]. However, there are no reports suggesting upregulation of carbamylation/ guanidinylation of Ub Lysines in proteasomal inhibited cells. At this stage, we are thus unable to confirm if the observed signal corresponds to carbamylated/ guanidinylated K48, due to the lack of strong biochemical evidence. Such evidence could in the future be obtained from an immunocytochemical analysis of expoxomicin treated cells using antibodies against carbamylated/ guanidinylated proteins. Guanidinylated Lysines are almost identical to Arginines which makes it difficult to discern the two epitopes. Additionally, reliable antibodies specific to carbamylated Lysines are currently unavailable[31,32].

Next, we examined the effect of proteasome inhibition upon the backbone fold of K48-Ub. Before treatment, we detected a backbone ^{15}N-^{13}C correlation that showed strong resemblance to the in vitro K48-Ub spectrum (Supporting Information, Figure S3) and exhibited a narrower line width compared to the K-Ub case. In the proteasome inhibited cells, we observed a significant line broadening of $^{13}C_\alpha$ resonance of the backbone ^{15}N-$^{13}C_\alpha$ correlation, consistent with the occurrence of multiple backbone conformations. Comparison to ssNMR assignments of all seven lysine residues (included in Figure 8A) suggested that these could include α-helix like folds of K48, which were not reported to be observed in vitro or before proteasome inhibition (Figure 8A & Supporting Information, Figure S3A&B). These additional

correlations may arise either due to conjugate formation in a fraction of K48-Ub and/or to other molecular binding events that would entail polyubiquitination. To establish this conclusively, 3D experiments need to be performed.

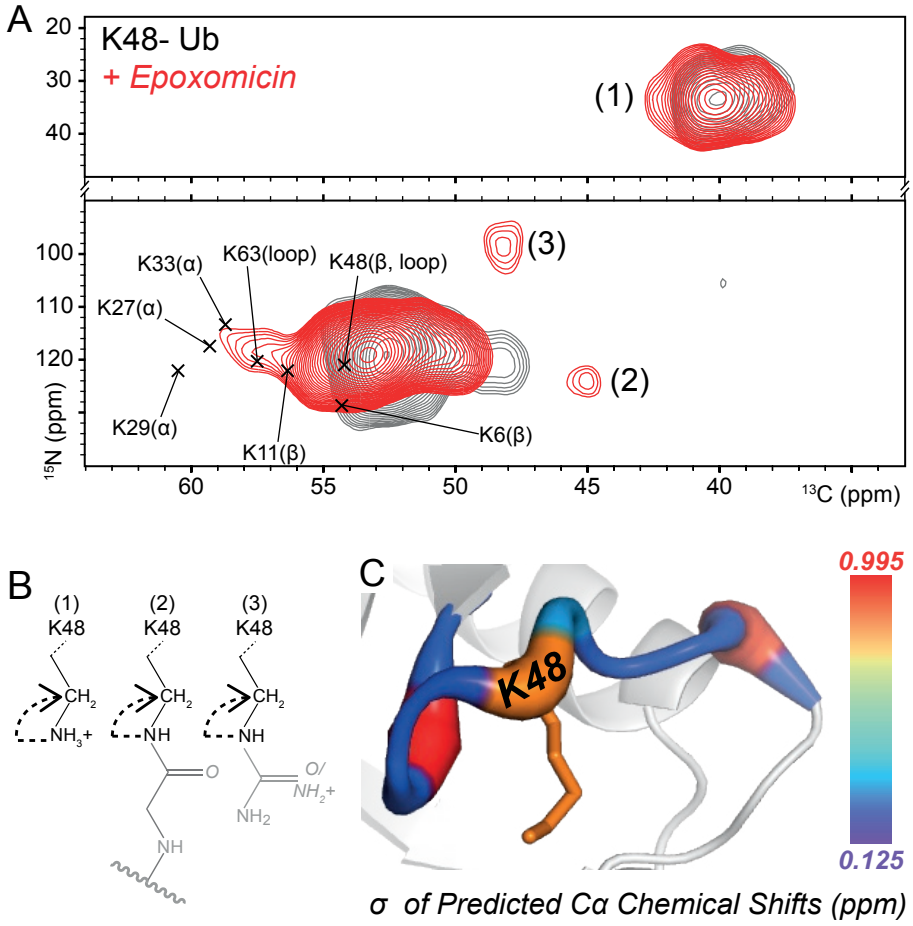


Figure 8: The effect of proteasomal inhibition on K48 of Ub in HeLa cells. A) ^{15}N-^{13}C before (black) and after (red) inhibitor treatment. The in vitro backbone chemical shift (^{15}N-^{13}Cα) assignments for Ub lysines (BMRB ID: 7111) as well as the corresponding secondary structure for each residue (α-helix, random coil, and β-strand) are included as reference. B) K48 side-chain structures related to the three different N-C correlations denoted (1-3) in figures (A). C) Predicted variation of ^{13}Cα chemical shifts for K48 (see Materials and Methods), in comparison to its neighbouring residues based on several crystal structures of Ub complexes, di- as well as monomeric Ub structures.

To this end, we performed a computational analysis of crystal and NMR structures of a set of Ub, poly- Ub and its complexes[23] (See Materials and Methods). This yielded a wide distribution of predicted ^{13}Cα chemical shifts for K48 in comparison to the neighboring residues in the unstructured region where K48 is present (Figure

8C). Note that K48 does not exist in a turn or alpha helix like conformations in the in vitro structures we analyzed. We speculate that such conformations may not be visited in vitro. In summary, our findings seem to provide an in-cell support for earlier in vitro NMR studies, predicting a correlation between residue-specific Ub dynamics and the ability to engage in local protein-protein interactions in cells[23].

Conclusions

Thus far, direct insight into the structural and chemical changes that occur on Ub due to the effect of proteasome inhibition inside human cells has been limited. We have presented solid-state NMR DNP results suggesting that Ub populations undergo large-scale conformational changes upon proteasome inhibition while concomitantly forming poly-Ub chains and undergoing other PTMs. While additional experiments are necessary to pinpoint the details of these alterations, our findings provided us with a direction to focus on specific Ub residues for future studies. Due to the lack of resolution in the 2D spectra, we could in the future acquire 3D NMR spectra (see Chapter 2) before and after epoxomicin treatment on HeLa cells harboring Ub. This strategy would not only provide additional evidence for the disappearance of resonances upon proteasome inhibition but may also yield additional insights into the residues of Ub that are not resolved in the 2D data. Finally, in-vitro reference studies on single Ub populations detected in cell preparations can deliver insights into why conformational heter0geneity is brought about in the cellular setting upon epoxomicin treatment. Moreover, performing 3D experiments on K-Ub and K48-Ub containing cells would also reveal if poly- Ub chain formation and the tentatively identified carbamaylation/ guanidinylation has a direct link to the conformational changes seen on K48. Additional key biochemical experiments (see section 3.4) would further be required to determine if carbamaylation/ guanidinylation do occur in Ub in cells treated with epoxomicin.

Supporting Information

Table ST1: Enhancements seen on various samples corresponding to this chapter.

Sample Description	Enhancements	
	Cell (~0-60 ppm)	Glycerol (65/ 75 ppm)
HeLa cells- $^{13}C\&^{15}N$ Ub	129.7	165.5
HeLa cells- $^{13}C\&^{15}N$ K- Ub	145.8	209.2
HeLa cells- ^{13}C & ^{15}N K- Ub (+ Epoxomicin)	138.9	209.2
HeLa cells- ^{13}C & ^{15}N K48- Ub	124.1	160.2
HeLa cells- ^{13}C & ^{15}N K48- Ub (+ Epoxomicin)	118.1	243.1
HeLa cells- ^{13}C & ^{15}N Ub (No Glycerol)	55.1	-
HeLa cells- ^{13}C & ^{15}N Ub (+ Epoxo., No Glycerol)	50.5	-

Table ST2: Resonances that could be analyzed, and peaks indicated in the 2D PDSD Experiments, shown in Figure 4.

Assignment	$^{13}C\ \delta_1$ ppm	$^{13}C\ \delta_2$ ppm
S65CB-S65C	64.4	171.6
I3CA-I3CD1	59.5	14.3
I3CG1-I3CD1	25.4	14.3
I13CA-I13CG2	60.2	17.6
I13CA-I13CD1	60.2	14.8
I13CG2-I13CB	17.6	41.2
L15CA-L15CB	52.8	47.3
P19CD-P19CA	50.4	65.1
I23CG2-I23CA	18.0	62.0
V26CA-V26CB	67.9	31.0
I30CA-I30CB	66.1	36.8
P37CD-P37CA	50.9	61.7
P38CA-P38CD	66.4	51.3
P38CD-P38CA	51.3	66.4
I44CA-I44CG2	59.0	17.7
I44CA-I44CD1	59.0	14.2
T55CA-T55CB	59.5	72.6
T55CB-T55CA	72.6	59.5
S65CA-S65CB	60.3	64.4
S65CB-S65CA	64.4	60.3
T66CA-T66CB	61.7	68.4
T66CB-T66CA	68.4	61.7
M1CA-M1C	54.4	170.6
E18CG-E18CD	36.1	183.1
I30CG2-I30C	16.6	178.4
I36CG1-I36C	27.7	173.0
I3CA-I3CG2	59.5	18.1

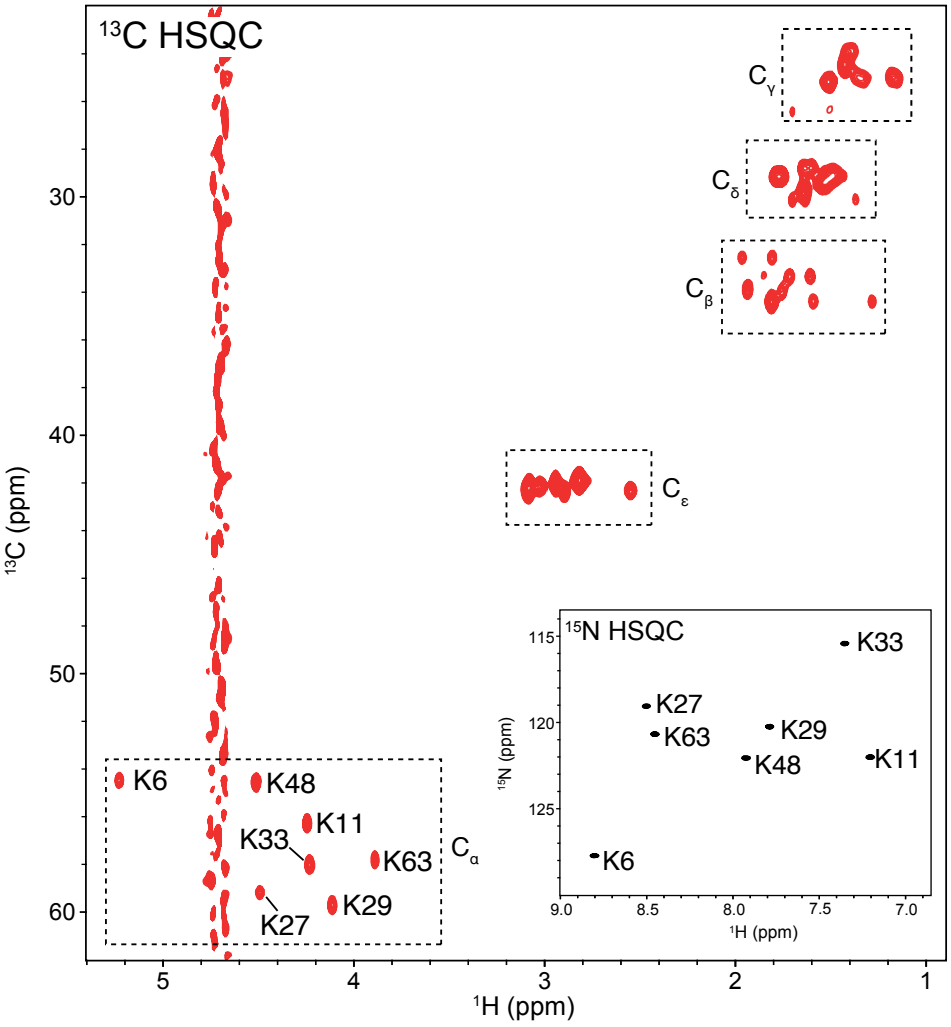


Figure S1: 2D solution-state NMR spectrum ^{13}C- HSQC (red) and ^{15}N- HSQC of K-Ub (black). Assignments of lysine residues are indicated.

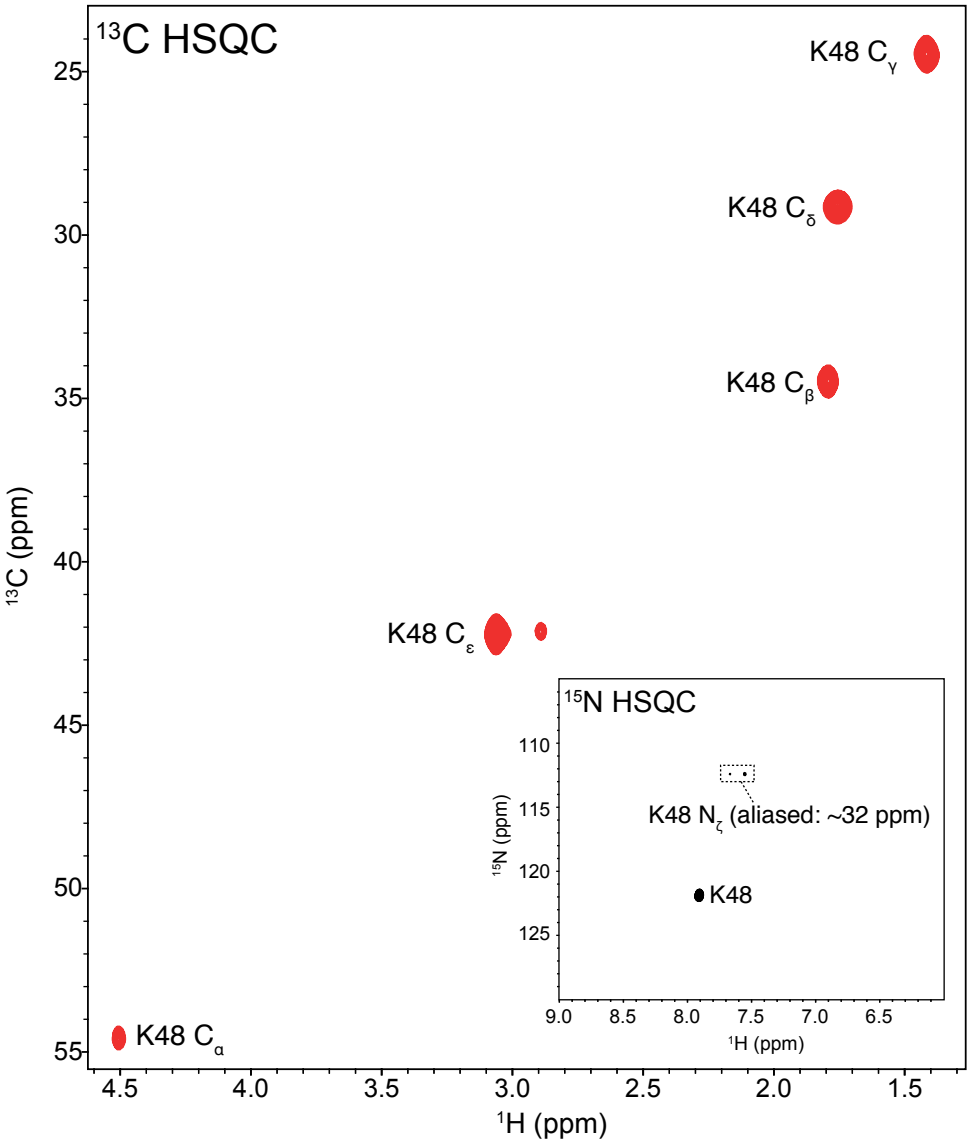


Figure S2: 2D solution-state NMR ^{13}C- HSQC (red) and ^{15}N- HSQC (black) of K48-Ub obtained by solid phase peptide synthesis (See Materials and Methods). Only ^{15}N-^{1}H and ^{13}C-^{1}H correlations from residue K48 were detected, confirming our intended labelling scheme.

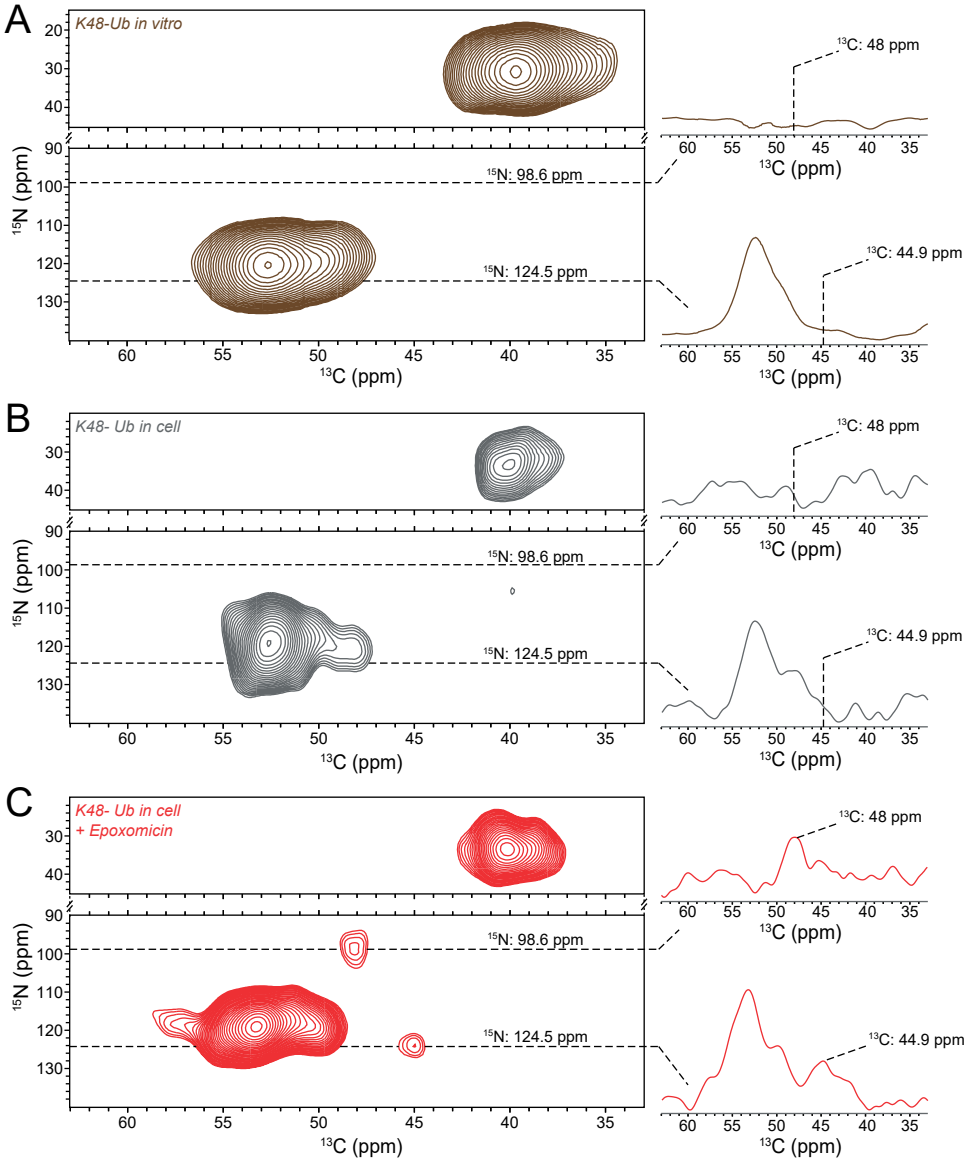


Figure S3: 2D (^{15}N-^{13}C) DNP-ssNMR spectra of K48-Ub in vitro (A), in HeLa cells (B) and in proteasome-inhibited HeLa cells (C) showing directly bonded N-C correlations. 1D slices were extracted at ^{15}N frequencies to indicate the presence of isopeptide-linked K48 (124.5 ppm) carbamylated/ guanidinilated K48 (98.6 ppm) Nζ- Cε correlations.

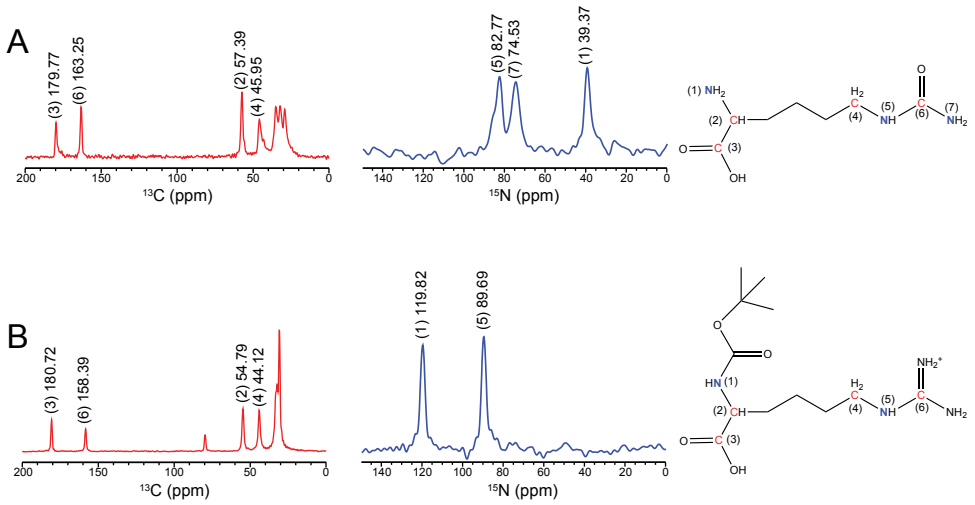


Figure S4: Solid-state NMR 1D ^{13}C (red) and ^{15}N (blue) spectra of homocitrulluine (A) and homoarginine (B) with annotations of the peaks of interest. In both cases, N_ζ resonances observed in the in-cell conditions were significantly deshielded with respect to in vitro conditions (Figure 8) and (Table ST3). The $^{15}N_\zeta$ chemical shifts seen for compounds do not fully match with the peak seen in cells (98 ppm- Figure 8). However, they are well within one standard deviation from average chemical shifts of the analogous N_ε of Arginine, which is seen to exhibit even higher chemical shifts in solution[33]. See Materials and Methods for the protocol used for synthesis of homocitrulline and homoarginine.

Table ST3: NMR assignment table of PTMs identified in the in-cell NMR Spectra.

Description of the Modification	Observed in-cell		Reference		Comments
	δ_1 ppm	δ_2 ppm	δ_1 ppm	δ_2 ppm	
Isopeptide linked K48	124.5 (N_ζ)	44.9 (C_ε)	122.01 (N_ζ)	43.45 (C_ε)	Reference:[28]
Carbamylated/ Guanidinilated K48	98.6 (N_ζ)	48 (C_ε)	82.77 (N_ζ)	45.95 (C_ε)	Homocitrulline (Fig S4A)
			89.69 (N_ζ)	44.12 (C_ε)	Homoarginine (Fig S4B)

References

1. Pickart, C. M. Mechanisms Underlying Ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
2. Scheffner, M., Nuber, U. & Huibregtse, J. M. Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83 (1995).
3. D'Andrea, A. & Pellman, D. Deubiquitinating Enzymes: A New Class of Biological Regulators. *Crit. Rev. Biochem. Mol. Biol.* **33**, 337–352 (1998).
4. Komander, D., Clague, M. J. & Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **10**, 550–563 (2009).
5. Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–29 (2012).
6. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).
7. Yau, R. & Rape, M. The increasing complexity of the ubiquitin code. *Nat. Cell Biol.* **18**, 579–586 (2016).
8. Li, W. & Ye, Y. Polyubiquitin chains: Functions, structures, and mechanisms. *Cellular and Molecular Life Sciences* vol. 65 2397–2406 (2008).
9. Fushman, D. & Wilkinson, K. D. Structure and recognition of polyubiquitin chains of different lengths and linkage. *F1000 Biology Reports* vol. 3 (2011).
10. Xu, P. *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133–45 (2009).
11. Liu, P. *et al.* K63-linked polyubiquitin chains bind to DNA to facilitate DNA damage repair. *Sci. Signal.* **11**, (2018).
12. Ciechanover, A. The ubiquitin-proteasome proteolytic pathway. *Cell* **79**, 13–21 (1994).
13. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–5 (2001).
14. Dantuma, N. P., Groothuis, T. A. M., Salomons, F. A. & Neefjes, J. A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *J. Cell Biol.* **173**, 19–26 (2006).
15. Meierhofer, D., Wang, X., Huang, L. & Kaiser, P. Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *J. Proteome Res.* **7**, 4566–76 (2008).
16. Kaiser, S. E. *et al.* Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods* **8**, 691–696 (2011).
17. Zhang, X. *et al.* An Interaction Landscape of Ubiquitin Signaling. *Mol. Cell* **65**, 941-955.e8 (2017).
18. El Oualid, F. *et al.* Chemical Synthesis of Ubiquitin, Ubiquitin-Based Probes, and Diubiquitin. *Angew. Chemie Int. Ed.* **49**, 10149–10153 (2010).
19. Berkers, C. R. *et al.* Peptide Splicing in the Proteasome Creates a Novel Type of Antigen with an Isopeptide Linkage. *J. Immunol.* **195**, 4075–4084 (2015).
20. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *J. Magn. Reson.* **142**, 97–101 (2000).
21. Hohwy, M., Rienstra, C. M., Jaroniec, C. P. & Griffin, R. G. Fivefold symmetric homonuclear dipolar recoupling in rotating solids: Application to double quantum spectroscopy. *J. Chem. Phys.* **110**, 7983 (1999).
22. Baldus, M., Petkova, A. T., Herzfeld, J. & Griffin, R. G. Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **95**, 1197–1207 (1998).
23. Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived

ubiquitin ensemble in solution. *Science (80-.).* **320**, 1471–1475 (2008).

24. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
25. Winget, J. M. & Mayor, T. The Diversity of Ubiquitin Recognition: Hot Spots and Varied Specificity. *Molecular Cell* vol. 38 627–635 (2010).
26. Lv, Z., Williams, K. M., Yuan, L., Atkison, J. H. & Olsen, S. K. Crystal structure of a human ubiquitin E1– ubiquitin complex reveals conserved functional elements essential for activity. *J. Biol. Chem.* **293**, 18337–18352 (2018).
27. Dammer, E. B. *et al.* Polyubiquitin Linkage Profiles in Three Models of Proteolytic Stress Suggest the Etiology of Alzheimer Disease. *J. Biol. Chem.* **286**, 10457–10465 (2011).
28. Hagan, R. M. *et al.* NMR spectroscopic and theoretical analysis of a spontaneously formed Lys-Asp isopeptide bond. *Angew. Chem. Int. Ed. Engl.* **49**, 8421–5 (2010).
29. Desmons, A. *et al.* Proteasome-dependent degradation of intracellular carbamylated proteins. *Aging (Albany. NY).* **11**, 3624–3638 (2019).
30. Sun, S., Zhou, J.-Y., Yang, W. & Zhang, H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal. Biochem.* **446**, 76–81 (2014).
31. Verheul, M. K. *et al.* Pitfalls in the detection of citrullination and carbamylation. *Autoimmun. Rev.* **17**, 136–141 (2018).
32. Zheng, Z. *et al.* Disordered Antigens and Epitope Overlap Between Anti–Citrullinated Protein Antibodies and Rheumatoid Factor in Rheumatoid Arthritis. *Arthritis Rheumatol.* art.41074 (2019) doi:10.1002/art.41074.
33. Pantoja-Uceda, D. *et al.* The NMR Structure and Dynamics of the Two-Domain Tick Carboxypeptidase Inhibitor Reveal Flexibility in Its Free Form and Stiffness upon Binding to Human Carboxypeptidase B † ‡. *Biochemistry* **47**, 7066–7078 (2008).

Chapter 4:

DNP-supported solid-state NMR spectroscopy of proteins inside *E. coli* cells

This chapter forms part of

"A solid-state NMR protocol to characterize proteins in a native bacterial environment"

(Manuscript in preparation for an invited contribution to Nature Protocols)

Siddarth Narasimhan, Cecilia de Agrela Pinto, Alessandra Lucini Paioni, Gert Folkers and Marc Baldus

Abstract

Previous work has shown that in-cell solution state NMR can make use of flexible protein expression systems available for *E. coli* to study soluble proteins after overexpression, provided that the protein of interest tumbles rapidly. Here we attempt to extend such studies to ssNMR spectroscopy with the final goal to study biomolecules in bacterial cells that remain invisible to in-cell solution state NMR. Similar to previous work on mammalian cells, our ssNMR approach involves the use of Dynamic Nuclear Polarization (DNP) which greatly improves spectroscopic sensitivity. Under such conditions, we have examined the effect of different isotope labelling strategies on spectroscopic signals stemming from the cellular background. Our studies lead to an optimal protocol to conduct DNP-ssNMR experiments on isotope-labelled proteins in *E. coli* cells with maximum DNP enhancements and minimum background signal contributions.

Introduction

Previous work has shown that molecular crowding is significantly more pronounced in bacterial cells such as *E.coli* cells compared to human cells[1]. As a result, the study of even moderately sized globular bacterial proteins can be precluded for in-cell solution-state NMR[2]. Hence, a solid-state NMR approach that allows to study soluble proteins in bacterial (*E. coli*) cells of an arbitrary type is highly desirable. Due to the inherent lack of sensitivity in cellular ssNMR preparations, it is necessary to employ sensitivity enhancing techniques such as DNP[3,4]. Compared to recent DNP-ssNMR approaches to study proteins inside mammalian cells[4], three major challenges must be overcome. Firstly, and unlike in mammalian cells, proteins cannot be delivered by electroporation into bacterial cells. As a result, targeted labelling of the protein of interest must be achieved by overexpression. Secondly, it is vital to ensure that unwanted NMR signals stemming from isotope labelled cellular background, in particular stemming from the non-protein components, are minimized through dedicated labelling strategies. Finally, sufficiently large DNP enhancements must be obtained to facilitate the acquisition of 2D and 3D NMR experiments.

As discussed in the following, we have addressed these challenges using Ub as a test protein. We deliberately chose Ub because of its small size and compact 3D fold. Additionally, unlike in mammalian cells, we expect *a priori* for Ub to be inert and hence not to be involved in the formation of protein complexes in bacteria. We tested if Ub can be labelled in a targeted manner in *E. coli* compatible with DNP-ssNMR approaches. DNP is applied in a similar manner to NMR studies involving mammalian cells[4] discussed in the previous chapters. In particular, previous

reports[5] suggest that DNP agents can be introduced into bacterial cells prior to NMR measurements to prevent radical reduction.

The classical labelling strategy used in bacterial in-cell solution state NMR involves overexpressing the protein of interest to an extent that the NMR signatures of the same are detectable, while the background signals (especially stemming from components like lipids or labelled amino acids) can be neglected in the spectra[6]. The latter approximation is however not valid in DNP-ssNMR as the signals from the entire cell background, particularly from the labelled lipids, are spectroscopically enhanced. Thus, we turned to another approach that uses the antibiotic rifampicin to suppress native *E.coli* RNA-polymerase dependent protein expression, which has been previously applied to cellular ssNMR of membrane proteins[7,8]. Despite successful usage[7,9,10], the rifampicin approach has been deemed to be non-critical/ redundant when protein overexpression rates are sufficiently high for in-cell solution NMR studies[11]. However, we resorted to using the rifampicin approach in developing our method, as in principle it also allows for future studies of proteins at low expression levels.

Materials and Methods

Targeted labelling of Ub in *E. coli* cells using ^{13}C- glucose & ^{15}N- ammonium chloride

Fresh transformants of Rosetta R2 or Lemo cells harboring a pET derived plasmid carrying a Ub sequence were picked from a plate and incubated in an overnight LB culture with 50 mg/L ampicillin and 35 mg/L chloramphenicol. 50 µL of the overnight culture was incubated in 50 mL of LB and grown up to OD600 of 0.05. The cells were then switched to an equal volume (50 mL) of unlabelled M9 medium and grown up to OD600 of 1.0. Expression was induced by adding 0.5 mM IPTG (final concentration) and incubated at 30°C for 30 minutes. The cells were then moved to a 18°C incubator and incubated for 5 mins to cool them down. Subsequently, 100 mg/L rifampicin (from 1000x DMSO stock) was added and the flasks were incubated for another 15 mins. The cells were then switched to precooled (18°C) 50 mL labelled M9 medium (supplemented with 100 mg/L rifampicin, 2g/L of ^{13}C-D-glucose and 0.5 g/L of $^{15}NH_4Cl$). Expression was carried out for a total period of 22 h. In the case of the "timed unlabelling" protocol (*vide infra*), the cells were switched to precooled unlabelled M9 medium with 100 mg/L rifampicin for the last 6 h of expression. The "timed unlabelling" is done to reduce background labelled components (mostly referring to free amino acids).

Targeted labelling of Ub in *E. coli* cells using a ^{13}C-, ^{15}N- algal amino acid mix and background deuteration

Cells were grown in 50 mL of LB and up to OD600 of 0.05 in the same manner as described above. The cells were switched to 50 mL unlabelled M9 medium and grown up to OD600 of 1.0. Unlabelled M9 was optionally prepared in its deuterated form- in D_2O and supplemented with deuterated (d_6) D-glucose. Expression was induced by adding 0.5 mM IPTG (final concentration) and incubated at 30°C for 30 minutes. The cells were then moved to a 18°C incubator and incubated for 5 mins to cool them down. Following this, 100 mg/L rifampicin (from 1000x DMSO stock) was added, the flasks were incubated for another 15 mins. Then, 500 mg/L of ^{13}C-, ^{15}N-algal amino acid mixture (16 amino acids- devoid of Cys, Trp, Asn & Gln) was added to the culture and the expression was carried out for a total period of 22 h. The cells were switched to precooled unlabelled M9 medium (optionally deuterated) with 100 mg/L rifampicin in the last 6h of expression.

Preparing cells for DNP-ssNMR

Cells were harvested after expression by spinning down at 4000 G for 20 mins, washed two times with ice-cold buffer of M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 0.5 g/L NaCl). Half of the pellet was frozen for lysate preparation and the other half was resuspended in 50 µL of deuterated DNP juice (60% deuterated d_7- and ^{12}C enriched glycerol, in D_2O) comprising of 15 mM AMUPol in M9 salts (see above). The cells were spun into DNP rotors by centrifugation at 4500 G. The supernatant containing excess DNP juice was removed, and the rotor was frozen in the same manner as before (Chapters 2 & 3).

Lysates for solution-state NMR

The other half of the pellet was thawed on ice. Next, ice-cold buffer of M9 salts prepared in 10% D_2O with protease inhibitor cocktail (Roche). 600 µL of this buffer was used to resuspend the cells, and the cells were sonicated 10 times with 10-15 s long 13 kHz pulses with a 30 s break. After sonication, the contents were passed through a 19G syringe multiple times and finally spun down at 20,000 G for 30 mins. The clear lysate was further spun down for another 30 mins.

DNP Solid-State NMR experiments

DNP ssNMR experiments were conducted under low temperature (100 K) DNP conditions using a 3.2 mm triple-resonance (^1H, ^{13}C, and ^{15}N) magic-angle-spinning (MAS) probe head in a static magnetic field of 9.4 T, corresponding to proton/electron resonance frequencies of 400 MHz/263 GHz. All spectra were recorded at an MAS rate of 8 kHz using SPINAL-64 proton decoupling[12] at an r.f.

field strength of 84 kHz.

1D- ^{1}H-^{13}C and ^{1}H-^{15}N Cross-Polarization experiments

The 1D H-C/ H-N cross-polarization (CP) spectra were recorded using a 2 s recycle delay and a 10 ms acquisition time. Cross-polarization was achieved using contact times of 900 μs and 800-1000 μs for H-C and H-N transfer, respectively unless mentioned otherwise. 100 Hz line-broadening was applied prior to Fourier transformation.

2D- Proton-driven spin diffusion (PDSD) experiments

For 2D proton-driven spin diffusion (PDSD) experiments, a mixing time of 30ms was employed. Acquisition times were set to 10 ms (612 data points used- 405.3 ppm spectral width) and 5 ms (280 data points- 280 ppm spectral width) for the direct and indirect dimensions, respectively. 128 scans were used for samples labelled using ^{13}C Glucose and $^{15}NH_4Cl$. 80 scans were used for other samples. The 2D spectra were processed using a 0.28 π shifted sine squared window function on both dimensions unless explicitly mentioned otherwise. For spectral analysis, ssNMR assignments were obtained in the similar manner as described in Chapter 2.

2D- ^{15}N-^{13}C correlation experiments

^{15}N-^{13}C-correlation experiments were performed using a ^{1}H-^{15}N CP step of 800-100 μs and a SPECIFIC-CP[13] (^{15}N-^{13}C) step of 2.5 ms. The center frequencies were set to 75 ppm and 40 ppm for ^{15}N and ^{13}C dimensions, respectively. A recycle delay of 2 s was used, and 128 scans were acquired (except for the R2 cells where ^{13}C-, ^{15}N- algal amino acid mix was used for labelling, where 256 scans were acquired to compensate for low sensitivity). Acquisition times were set to 10 ms (512 data points- 497 ppm spectral width) and 2.2 ms (32 data points- 180 ppm spectral width) for the direct and indirect dimensions, respectively. The 2D spectra were processed using a 0.5 π shifted sine squared window function on both dimensions. For spectral analysis, ssNMR assignments were obtained in the similar manner as described in Chapter 2.

3D- ^{13}C DQ-SQ-SQ correlation experiment

The pulse program and processing parameters used here were exactly the same as the 3D experiment described in Chapter 2. Acquisition times were set to 10 ms (400 data points used- 307 ppm spectral width) in the F3 1Q direct dimension, 1.5 ms (64 data points- 212 ppm spectral width) for the F2 1Q indirect dimension and 0.85 ms (50 data points- 290 ppm spectral width) for the F1 2Q indirect dimension. 64 scans were acquired with a 2 s recycle delay.

Solution-state NMR experiments on lysates

We conducted solution-state NMR experiments at ambient temperatures (293-298 K) using triple channel (^{1}H, ^{13}C, and ^{15}N) cryogenic probes. Static magnetic fields of 21.1 T and 14.06 T corresponding to proton frequencies of 900 MHz (for R2 Lysates) and 600 MHz (for Lemo Lysates), respectively, were used. 1 s recycle delay was used for all experiments. 8 scans were acquired at 900 MHz and 16 scans at 600 MHz.

For ^{15}N HSQC, at 900 MHz, acquisition times were set to 70 ms (2046 data points- 16 ppm spectral width) and 38 ms (180 data points- 26 ppm spectral with) in direct and indirect dimensions, respectively. For the ^{15}N- HSQC experiment at 600 MHz, acquisition times were set to 53 ms (1024 data points- 16 ppm spectral width) and 26 ms (128 data points- 40 ppm spectral with) in direct and indirect dimensions, respectively.

For ^{13}C- HSQC data obtained at 900 MHz acquisition times were set to 47 ms (1024 data points- 12 ppm spectral width) and 3.5 ms (128 data points- 80 ppm spectral with) in direct and indirect dimensions, respectively. For the ^{13}C- HSQC experiment at 600 MHz, acquisition times were set to 80 ms (1536 data points- 16 ppm spectral width) and 2.8 ms (68 data points- 80 ppm spectral with) on direct and indirect dimensions, respectively.

All solution-state NMR 2D spectra were processed using a cosine squared window function on both dimensions.

Results and Discussion

4.1 Excessive background labelling despite using the rifampicin approach

As our protein of interest here was human- Ub, we used *E. coli* BL21 Rosetta R2 cells which harbor a T7 expression system (DE3) and the pRARE plasmid to further enhance the expression of rare codon tRNAs. We initiated the expression in the exponential growth phase using unlabelled M9 medium and added rifampicin to block native polymerases before switching to rifampicin containing isotope labelled medium (^{13}C Glucose and ^{15}N Ammonium Chloride; see Materials and Methods). This sequential protocol ensured expression of adequate amounts of T7 RNA polymerase that were at the same time not isotope labelled.

As in the case of HeLa cells described in previous chapters, the cells were resuspended in DNP juice after washing. Subsequently, MAS rotors were filled and rapidly frozen for DNP ssNMR (Figure 1A). We obtained a DNP enhancement factor of ~53 on the protein signals (Supporting information, Figure S1), which allowed us to record 2D ^{13}C, ^{13}C, correlated spin diffusion (PDSD) and N-C correlation experiments. The 1D CP experiments (Supporting Information, Figure S1) revealed that background labelling of lipids and possibly sugars had occurred, in line with the notion that these molecules were synthesized via ^{13}C- Glucose which is the sole carbon source in our medium. Notably, lipids, sugars and other building blocks of the cells are constantly under cellular maintenance, notwithstanding the blocking of native protein expression. As visible in Figure 1B, NMR signals stemming from lipids and sugars complicated our analysis of amino acid signals, as the signals from lipid head groups overlap with the Ser/ Thr region, and the lipid tail signals superimpose with the -CH_2 side chain region (Figure 1B: 1D projection and Supporting Information, Figure S1).

In addition to lipid signals, we observed resonances stemming from the isotope labelled unincorporated amino acids (free amino acids) in the N-C correlation experiments (Figure 1C, left panel). Note that such signals do not overlap with the typical protein backbone region (Figure 1C, right panel) because of their different ^{15}N frequencies (NH_3^+ resonate <60 ppm and backbone N-H groups are found above 100 ppm). They are thus usually neglected in solution-state NMR data[6]. Additionally, the NH_3^+ groups can be filtered out in HSQC spectra in solution-state NMR using appropriate NMR schemes. On the other hand, ^{13}C chemical shifts of free amino acids do overlap with aliphatic NMR signals stemming from amino acids in a folded protein. ^{13}C chemical shifts are a primary source of information in DNP-ssNMR experiments, in particular in ^{13}C-^{13}C correlation spectra. As a result, distinguishing between the signals from free amino acids and protein signals becomes complicated.

Notably, free labelled amino acids were also not a concern in previous studies on membrane proteins in their native lipids, as the soluble parts of the cells were discarded during the sample preparation steps[14].

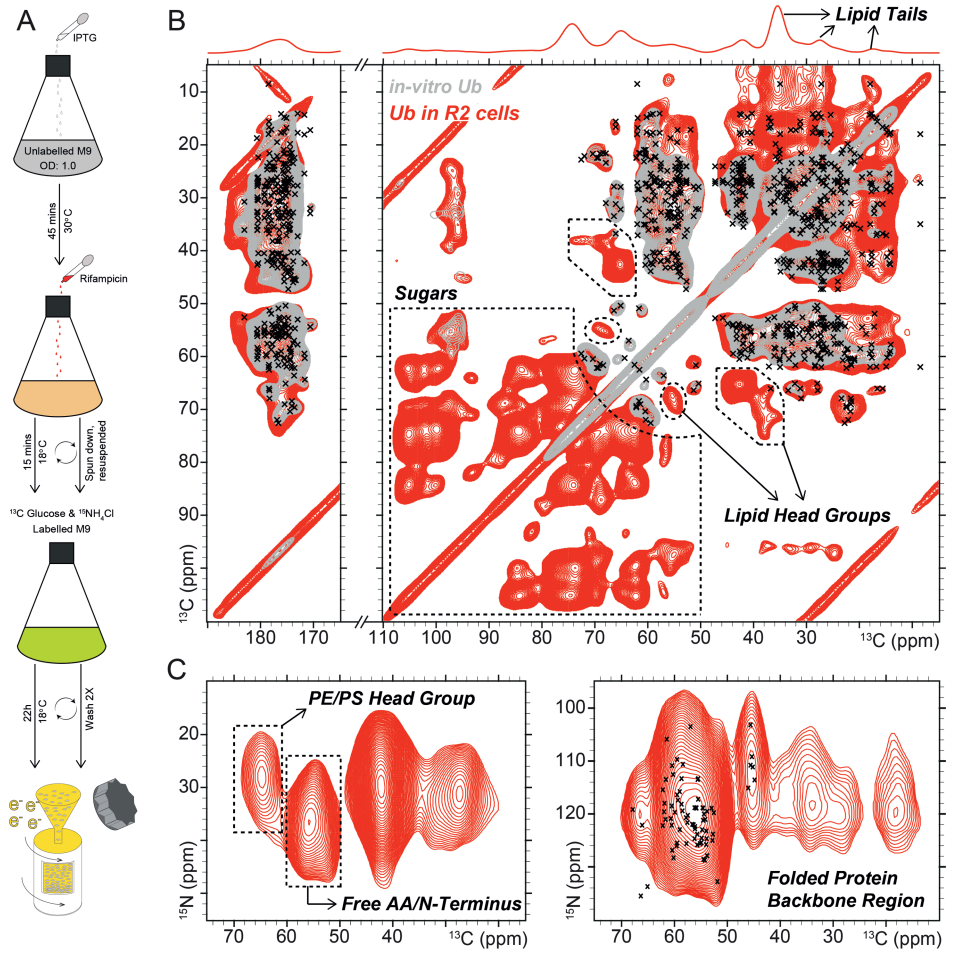


Figure 1: A) Sample preparation scheme for the rifampicin mediated selective protein labelling approach. B) ^{13}C, ^{13}C, correlated spin diffusion (PDSD) spectrum of R2 cells containing Ub (red) produced using the scheme presented in A, overlaid with a similar spectrum of pure Ub (grey). Background signals from sugars and lipids are dominant in the spectrum, in particular in the 1D projection. C) N-C correlation spectrum (divided into two spectral cutouts) showing the presence of labelled free amino acids and lipid head groups derived from amino acids. The black crosses in (B) and (C) represent the resonance assignments of free Ub.

To confirm the presence of free, labelled amino acids, we obtained lysates of the cells from the same cultures as our DNP-ssNMR preparations (See Materials and Methods). While ^{15}N- HSQC spectrum (Figure 2) of these lysates was largely free of background resonances, they readily appeared in the ^{13}C- HSQC spectra (Figure 3) which also showed spectral overlap with backbone and side-chain resonances of folded Ub. In spite of these additional signals, both spectra clearly indicated that the protein is well folded and confirmed the overall good quality of the sample.

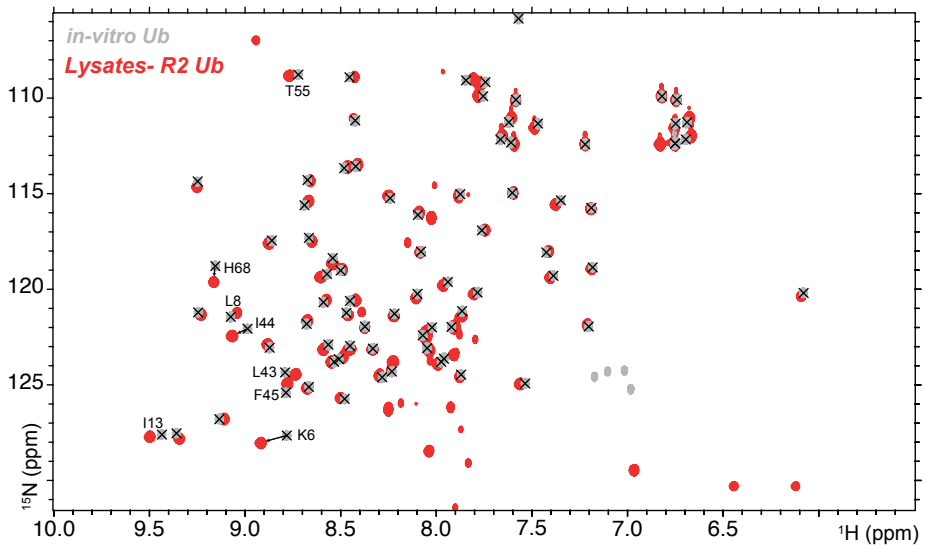


Figure 2: ^{15}N HSQC spectrum (solution-state NMR) of R2 cell lysates containing Ub (red) overlaid with the spectrum of Ub in water (grey) and chemical shift assignments of Ub (black crosses). The lysate spectrum contains minimal background peaks that do not overlap with the protein signals. A few peaks from known Ub recognition sites in its hydrophobic patch display noticeable chemical shift differences, possibly due to random hydrophobic interactions[14] (See also Supporting Information, Figure S2).

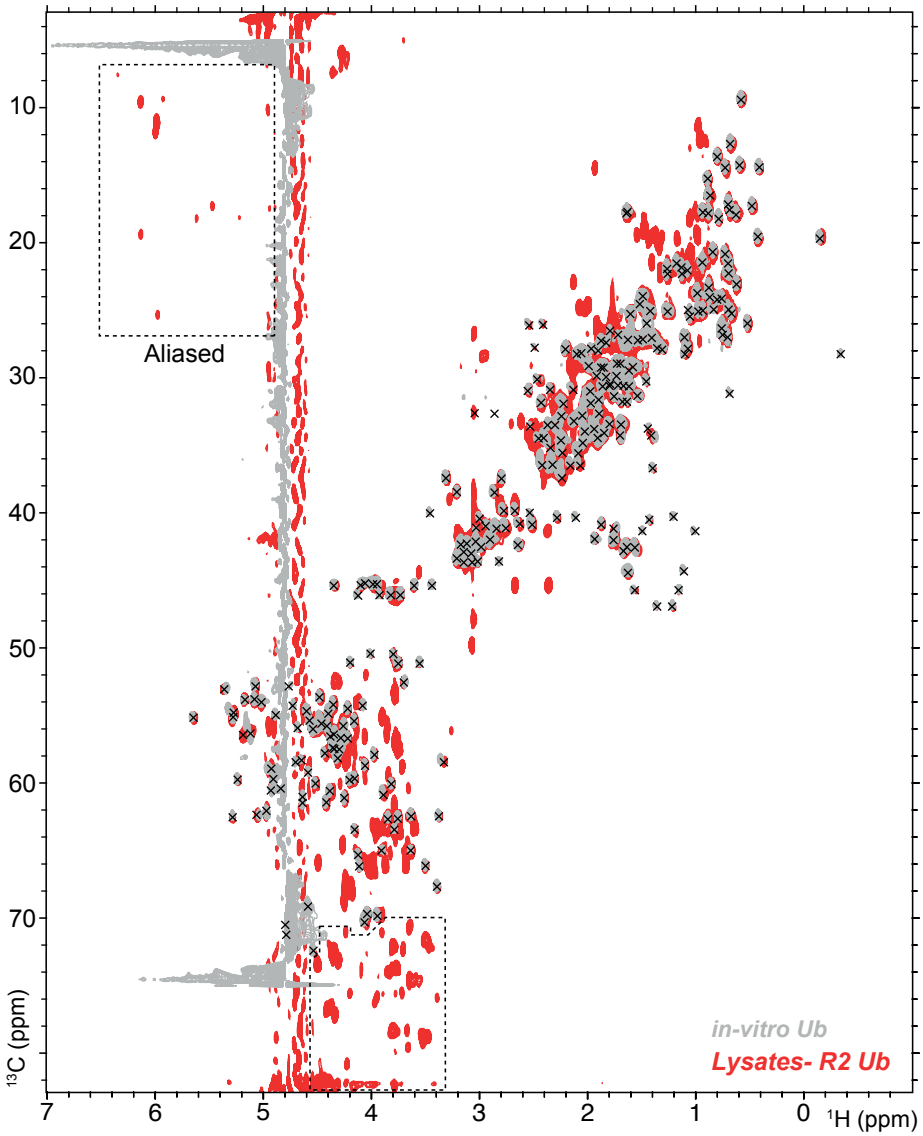


Figure 3: ^{13}C HSQC spectrum (solution-state NMR) of lysates of R2 cells expressing Ub (red) overlaid with the spectrum of Ub in water (grey) and chemical shift assignments of Ub (black crosses). The lysate spectrum contains additional signals, possibly stemming from labelled sugars in RNA (dashed boxes) that do not overlap with proteins. Moreover, other cellular background resonances from free amino acids significantly overlap with protein resonances in the ^{13}C dimension. Note that these correlations entail unmarked peaks in the lysate spectrum that do not match with any Ub assignments or Ub peaks (grey) and lie outside the spectral regions indicated by dashed boxes.

4.2 Reducing free amino acid resonances by "*Timed Unlabelling*"

To eliminate background resonances caused by free (isotope-labelled) amino acids as well as isotope labelled lipids, we speculated that reintroducing the cells into an unlabelled medium during the last few hours of expression (Figure 4A) would replace at least a part of the labelled amino acids and lipids with their unlabelled counterparts. This notion is inspired by the method of pulsed-SILAC labelling[16] used for some mass spectrometry studies. As the aim is to reduce the background without strongly reducing the total amount of isotope labelled Ub, we carried out expression in the isotope labelled medium for the larger time period and switched cells to unlabelled medium at the last quarter (6 h of 22 h) of the total expression time. Subsequently, we recorded 2D NMR experiments (Figure 4B & 4C) using experimental conditions similar to the previous sample and observed comparative values of DNP enhancements (Supporting Information, Figure S3 and Table ST1).

Our timed unlabelling approach showed a clear decrease in the amount of free labelled amino acids (Figure 4C). This was further confirmed in the solution-state NMR analysis of the lysates (Figure 5 & 6). However, we still observed correlations stemming from lipid labelling that dominated the Ser/ Thr region of the ^{13}C, ^{13}C, correlated spin diffusion spectrum (Figure 4B). The high prominence of lipid signals can also be attributed to the proximity of the bulk of the DNP radicals, that perhaps remained on the outside of the cells. In line with this notion, we observed larger DNP enhancements for lipid signals than for protein signals across both samples (Supporting Information, Table ST1).

Notably, we observed a similar effect in our DNP preparations of eukaryotic cells, where naturally abundant ^{13}C spins from ^{2}H- enriched glycerol in DNP juice dominated our ^{13}C carbon detected spectra (see Chapter 3). Such unwanted signal contributions were minimized by using a ^{13}C- depleted version of the ^{2}H- enriched glycerol for future preparations, including the ones pertaining to this chapter.

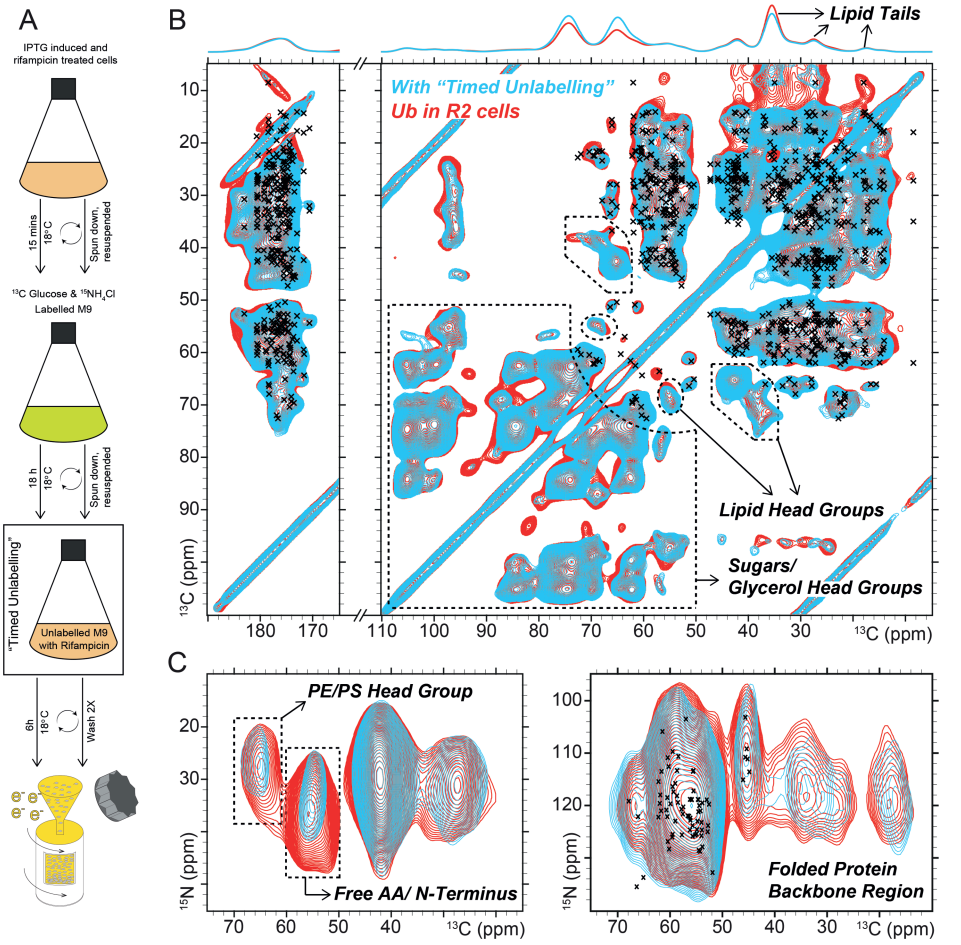


Figure 4: A) Sample preparation scheme for "Timed Unlabelling" of *E. coli* cells. (B) and (C) show comparison of the 2D spectra between cells prepared using the "Timed Unlabelling" approach (sky blue) and the cells prepared with complete expression in labelled medium (red). Crosses in (B) and (C) represent assignments of Ub. B) The ^{13}C, ^{13}C, correlated spin diffusion (PDSD) spectrum shows that the background lipid signals remained prominent as seen for the case of complete expression in labelled medium (see Materials and Methods). On the other hand, the N-C correlation spectrum showed a pronounced reduction of labelled free amino acids and lipid head groups derived from amino acids (C).

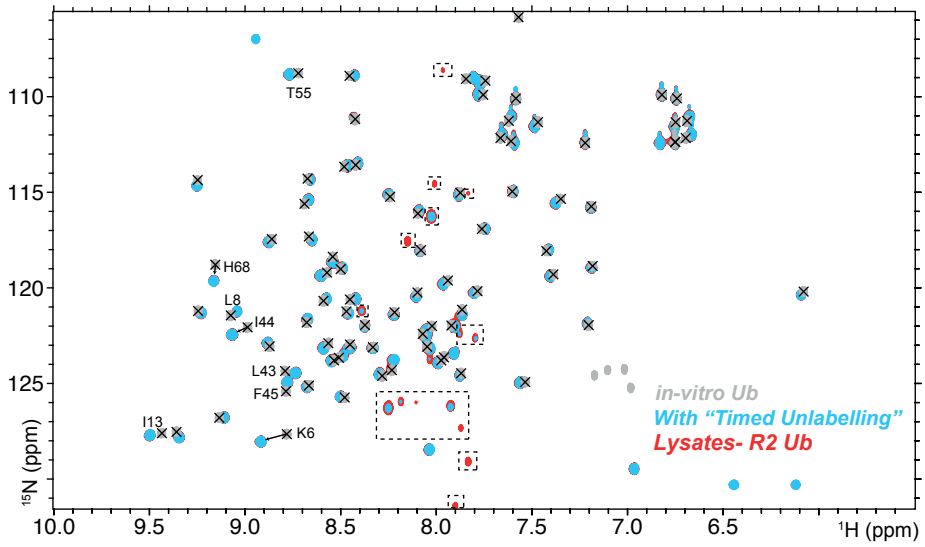


Figure 5: Comparison of ^{15}N HSQC spectra (solution-state NMR) obtained on lysates of cells prepared using the timed unlabelling approach (sky blue) and the lysates of cells prepared with complete expression in labelled medium (red). Overlaid on the two spectra is also the spectrum of Ub in water (grey) and chemical shift assignments of Ub (black crosses). The timed unlabelling approach efficiently reduces several cellular background signals (indicated using dashed boxes).

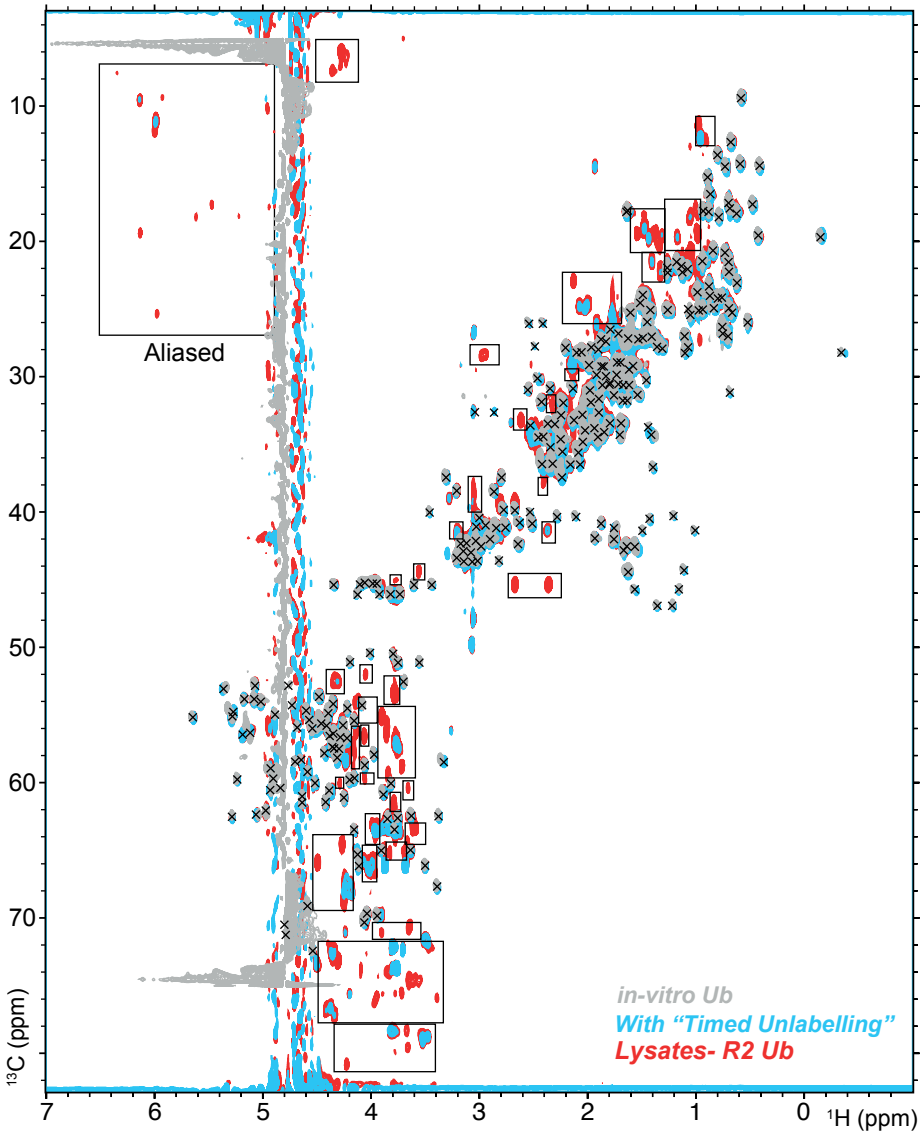


Figure 6: Comparison of the ^{13}C HSQC spectra (solution-state NMR) obtained on lysates of cells prepared using the timed unlabelling approach (sky blue) and the lysates of cells prepared with complete expression in labelled medium (red). Overlaid on the two spectra is also the spectrum of Ub in water (grey) and chemical shift assignments of Ub (black crosses). The timed unlabelling approach reduces background including labelled free amino acids (solid boxes), to a large extent, in line with the DNP-ssNMR N-C spectra of whole cells (Figure 4).

4.3 Isotope labelled amino acid mix further reduces background

From our results described above, we concluded that further modifications to the protocol are needed to fully eradicate lipid signals in our DNP-ssNMR spectra of soluble proteins in cells. Therefore, we further modified our protocol by using algal derived isotope labelled amino acids[17] that have been used previously in ssNMR preparations of eukaryotic membrane proteins in human cells[18] as a labelling source (Figure 7A). Using algal amino acid mixture as the isotope source, we also explored the possibility of increasing DNP enhancements of the protein of interest and, at the same time, reducing the background enhancement by near complete deuteration of the cellular background (see Materials and Methods).

Unexpectedly, using the algal amino-acid mixture to label the protein yielded very poor DNP signal enhancement factors (of ~10) (Supporting Information, Figure S4). However and as intended, using the algal amino-acid mixture removed lipid tail resonances (Supporting Information, Figure S4) along with much of the sugar and head group resonances, except for a strong ^{13}C-^{13}C correlation at 65-40 ppm which was also seen in previous samples and did not overlap with most of the protein peaks (Figure 7B). Notably, this correlation was also observed in previous ssNMR studies of membrane proteins which made use of the rifampicin approach[7]. Therefore, we could tentatively attribute this signal to an amino-acid derived lipid head group such as Phosphotidylethanolamine. In line with this notion, the N-C correlation spectrum showed the presence of a correlation at the shielded (NH_3^+) region in ^{15}N, correlating with a 65 ppm ^{13}C resonance (Figure 7C) (See also Figure 1C & 4C).

The algal amino acid mixture is devoid of Trp, Cys, Gln and Asn amino acids. Since Ub does not contain any Cys and Trp residues, we only expected that the Gln and Asn residues would be missing from the spectra. However, the ^{15}N HSQC spectrum of the lysate (Figure 8) indicated the presence of all backbone Gln and Asn resonances while side-chain resonances were absent. We could attribute this to scrambling from Asp and Glu, which are the precursors of Gln and Asn[19]. Indeed, in the ^{13}C HSQC spectrum (Figure 9), we observed a partial decrease in some Asp and Glu signal intensities, with a concomitant presence of Asn and Gln resonances. At this stage, we cannot exclude the possibility of higher de novo synthesis of Asp and Glu in the unlabelled form. This could lead to a "dilution" effect in the total amount of labelled Asp and Glu incorporated into Ub. Notably, lower intensities of peaks was also seen with Ala: C_α-H_α intensities (A28, A46 in Figure 9). Further analysis using extensive Mass Spectrometry (MS) experiments is required to investigate this observation.

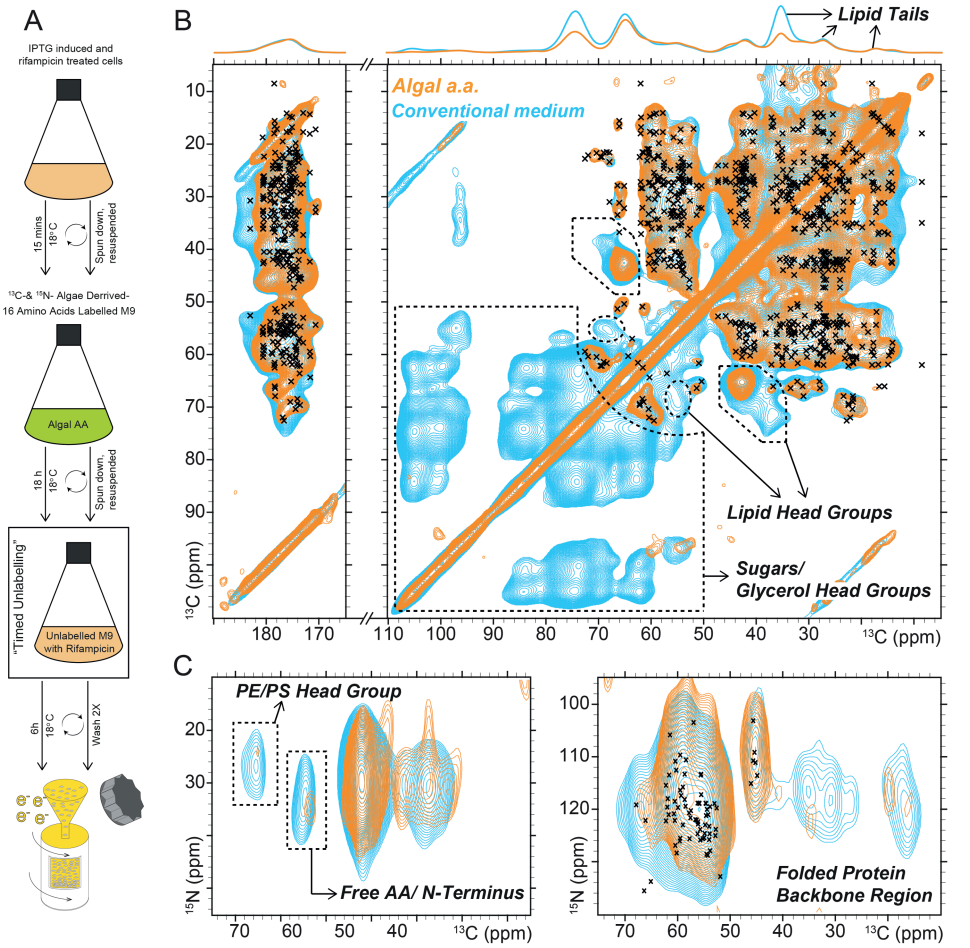


Figure 7: A) Sample preparation scheme where algal amino acid mix was used as the labelling source. (B) and (C) show comparison of the 2D spectra between cells prepared using timed unlabelling approach, with either conventionally ^{13}C-, ^{15}N- labelled medium (sky blue) or a medium containing the ^{13}C-, ^{15}N- algal amino acid mix (orange). Crosses in (B) and (C) represent assignments of Ub B) The ^{13}C, ^{13}C, correlated spin diffusion spectrum shows that the background lipid signals are strongly suppressed when algal amino acid mix is used. Likewise, the N-C correlation spectrum also shows hardly any background (C).

Due to the low overall enhancements, we could not perform 3D experiments within a reasonable experimental time frame. Previous DNP experiments in bacterial cell lysates have indicated that using deuteration in DNP buffers can boost DNP enhancements on the labelled target proteins[20]. In a similar spirit, we attempted to deuterate the cellular background, particularly the aliphatic carbons which have non-exchangeable hydrogens. The expected outcome is to obtain deuterated cells with only the protein of interest being isotope labelled and strongly protonated species in

the cells. We thus prepared fully deuterated (~90%) cellular samples (see Materials and Methods). The DNP enhancement factor of these preparations (~16-19) was still not satisfactory to record 3D spectra (Supporting Information, Figure S5).

Next, we evaluated whether using a shorter H-C CP contact time (≤ 100 μs) reduces NMR background signals from the lipid head groups. We surmised that the lipid background would be surrounded by fewer aliphatic ^{1}H groups than the proteins due to background deuteration. Indeed, we observed lesser DNP enhancements on background peaks at a lower CP contact time, without compromising on the enhancements of proteins (Supporting Information, Figure S6).

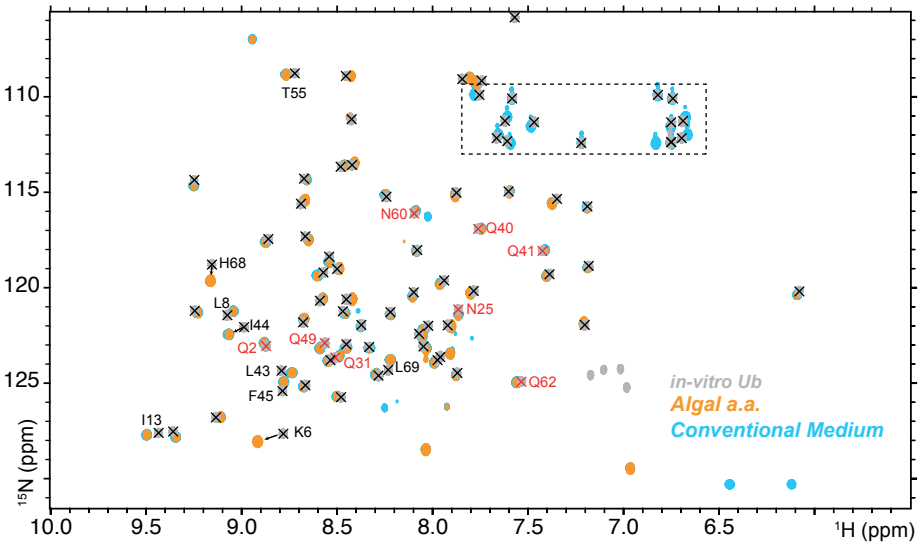


Figure 8: ^{15}N HSQC spectra (solution-state NMR) of R2 cell lysates- where algal amino acid mix was used as a labelling source (orange) overlaid on the spectrum with lysates from cells where a conventional ^{13}C-, ^{15}N- labelled medium was used (sky blue). The spectrum of Ub in water (grey) and chemical shift assignments of Ub (black & red crosses) are also overlaid. Scrambled Asn and Gln residues are indicated using red crosses & peak labels. The absent side-chain amides from Asn and Gln are indicated within the dashed box.

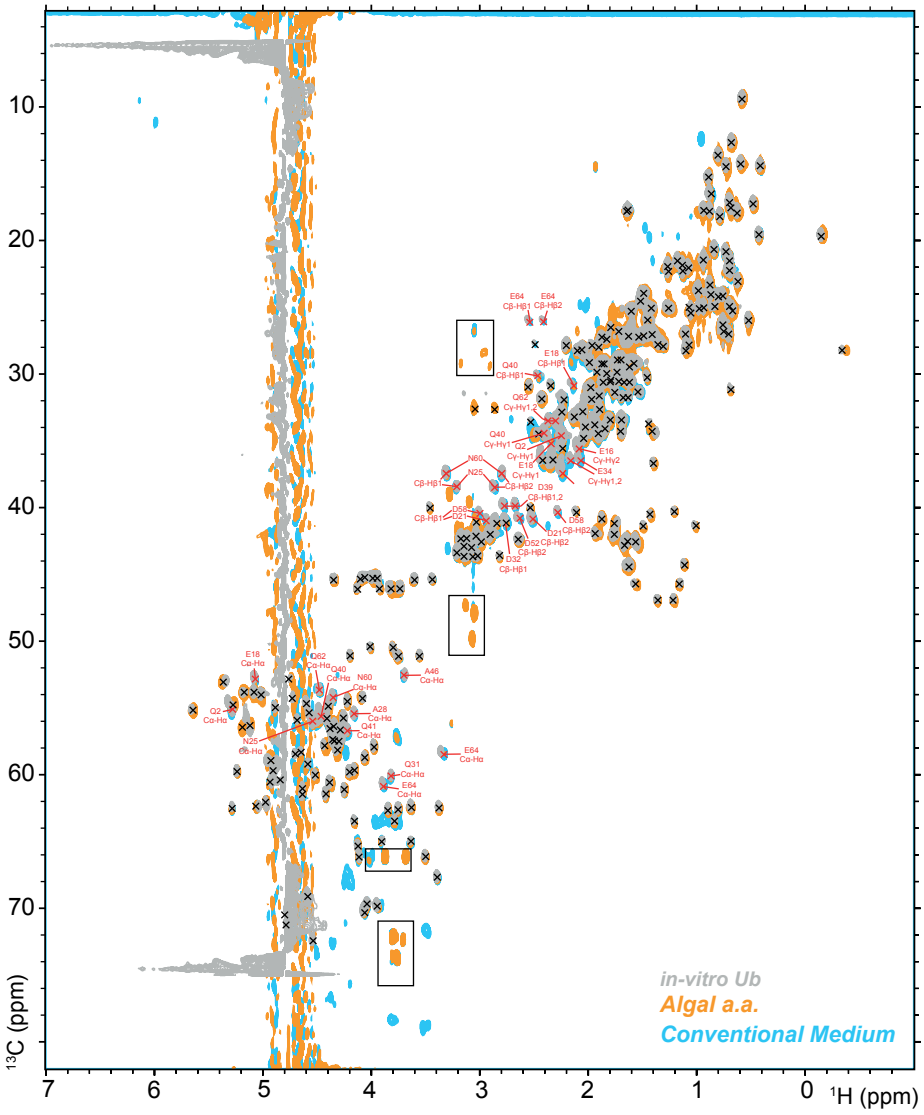


Figure 9: ^{13}C HSQC spectra (solution-state NMR) of R2 cell lysates- where algal amino acid mix was used as labelling source (orange) overlaid over the spectrum with lysates from cells where a conventional ^{13}C, ^{15}N labelled medium was used (sky blue). The spectrum of Ub in water (grey) and chemical shift assignments of Ub (black & red crosses) are also overlaid. Scrambled Asn & Gln residues along with Ala, Asp & Glu peaks that were low in intensity/ absent in the orange spectrum are indicated using red crosses & peak labels. Resonances from the cellular background are indicated within solid boxes.

4.4 Restoring DNP enhancements in algal amino acid labelled medium

Most bacterial proteins can be readily expressed in strains other than BL21 Rosetta 2, as their genes would not likely be composed of the rare codons that eukaryotic proteins contain. Since the final goal is to obtain a method that would be suitable for studies on bacterial proteins in the native bacterial (*E. coli*) environment, we reasoned that using BL21 Rosetta 2 strain is not mandatory. Therefore, we attempted to express Ub in *E. coli* BL21 Lemo cells which have a tunable expression system to regulate the amount of expressed proteins using the pLysS plasmid. This system would also enable the expression of membrane proteins and toxic proteins.

Hence, we tested if Ub is expressed and properly folded in Lemo cells despite lacking the pRARE plasmid, using the algal amino acid mixture as a labelling source in a fully protonated environment. The solution-state NMR spectrum of the soluble fraction revealed that Ub was labelled successfully and was properly folded in the cells without any signs of mistranslation (Figure 10, 11). The 2D ^{13}C, ^{13}C- correlated MAS-ssNMR spectrum confirmed that the insoluble cell fraction obtained after lysis had no traces of isotope labelled proteins (Supporting Information, Figure S7).

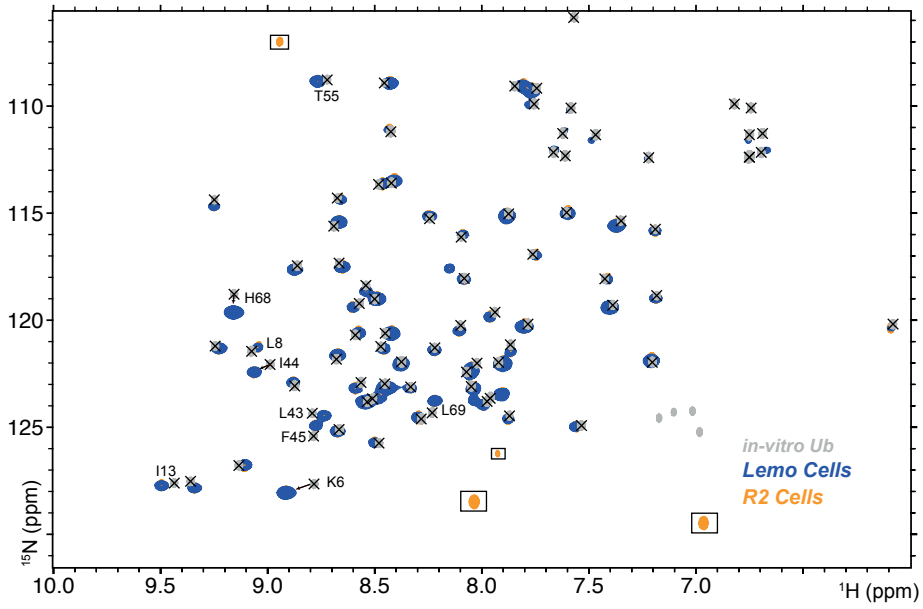


Figure 10: ^{15}N HSQC spectra (solution-state NMR) of R2 cell lysates (orange) vs Lemo cell lysates (blue)- where algal amino acid mix was used as a labelling source. The spectrum of Ub in water (grey) and chemical shift assignments of Ub (black & red crosses) are also overlaid. Background resonances in R2 cells are indicated in boxes.

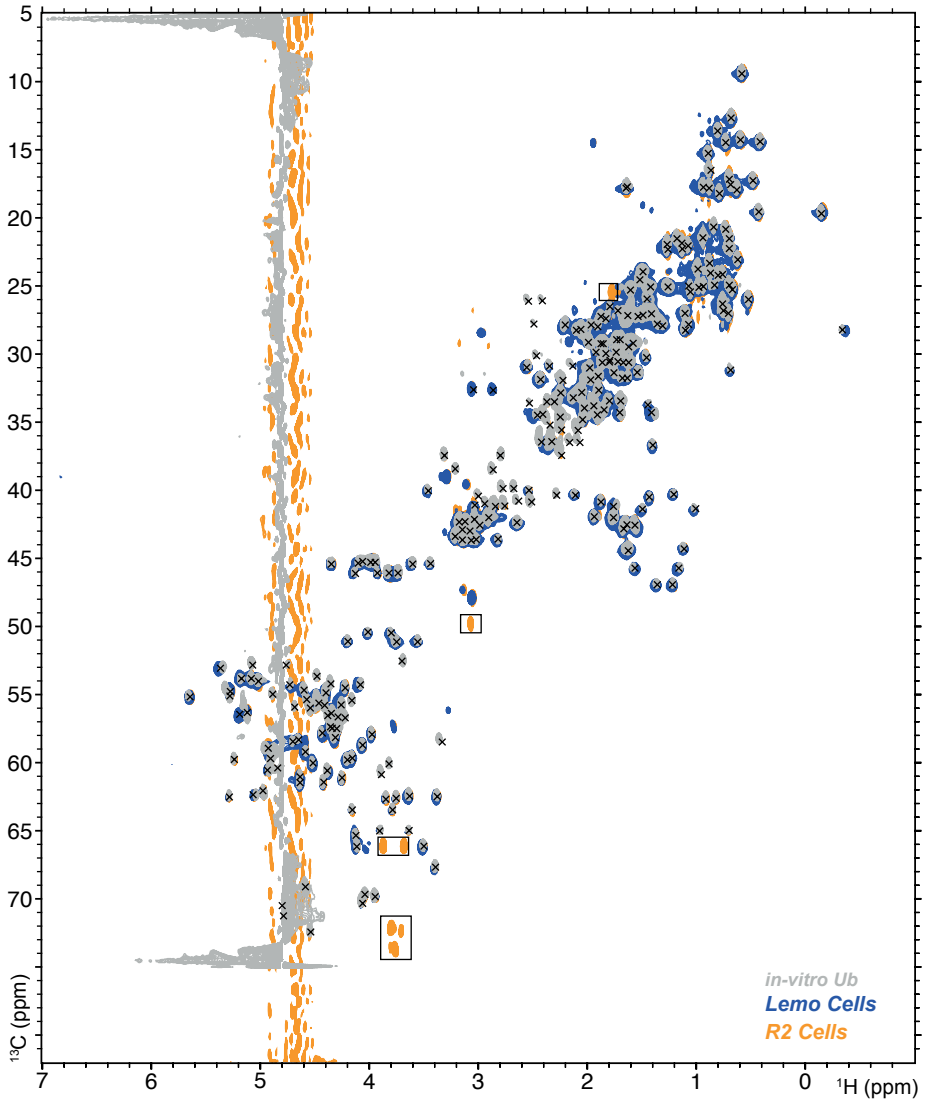


Figure 11: ^{13}C HSQC spectra (Solution-state NMR) of R2 cell lysates (orange) vs Lemo cells (blue)- where algal amino acid mix was used as a labelling source. The spectrum of Ub in water (grey) and chemical shift assignments of Ub (black & red crosses) are also overlaid. Background resonances in R2 cells are indicated in boxes.

As Rosetta R2 strains contain the pRARE plasmid, we speculated that the resulting expression of tRNA in combination with the supplemented algal amino acid mixture causes adverse effects to the observed DNP enhancements. Indeed, we obtained higher enhancements in Lemo cells when ^{13}C-, ^{15}N- algal amino acid mixture was used (vide infra). Elucidating the exact reason for the reduced enhancements when R2 cells are used in combination with algal amino acid mixture is beyond the scope of this current work.

4.5 3D NMR spectra of Ub in deuterated *E. coli* Lemo cells

Having established a means to obtain targeted isotope labelling of Ub in cells, we then produced fully deuterated (<90%) *E. coli* Lemo cells containing protonated and isotope labelled Ub with ^{13}C-, ^{15}N- algal amino acid mixture. We observed DNP enhancement factors between 90 and 108, which are the highest of all preparations discussed in this chapter (Supporting Information, Figure S8 & Table ST1). Note that ssNMR experiments using conventionally labelled cells with protonated background yielded much lower enhancements of ~20 (Supporting Information, Figure S9). These high DNP enhancements not only allowed us to acquire a 2D N-C correlation spectrum (Supporting Information, Figure S10) and a 2D ^{13}C, ^{13}C correlated spin diffusion spectrum (Figure 12), but also enabled the acquisition of a 3D DQ filtered ^{13}C, ^{13}C correlation spectrum. The spectral strips of Ub in HeLa cells shown in Figure 5 of Chapter 2 could also be obtained for Ub in *E. coli* cells (Figure 13). Due to the overlap with the strong background (65 ppm correlating with 40 ppm), tentatively assigned to the head groups of Phosphotidylethanolamine (PE) (see Supporting Information Figure, S10 & S11), some Ub resonances (i.e. I23/ 61) could not be studied.

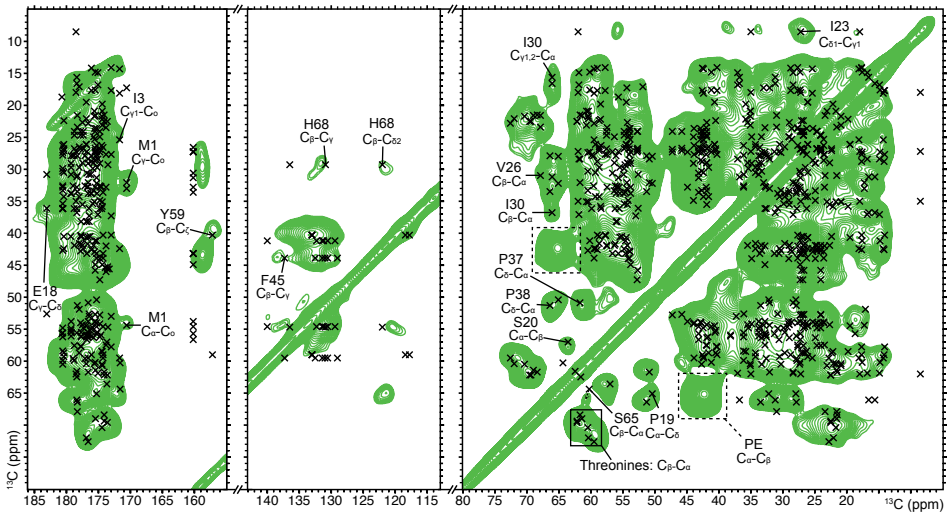


Figure 12: The ^{13}C, ^{13}C, correlated spin diffusion (PDSD) of Ub in deuterated BL21 Lemo cells. Resolved peaks are identified and indicated (at least one per residue identified). The only background peak was from the tentatively identified head group of phosphatidylethanolamine (PE) which is indicated in a dashed box.

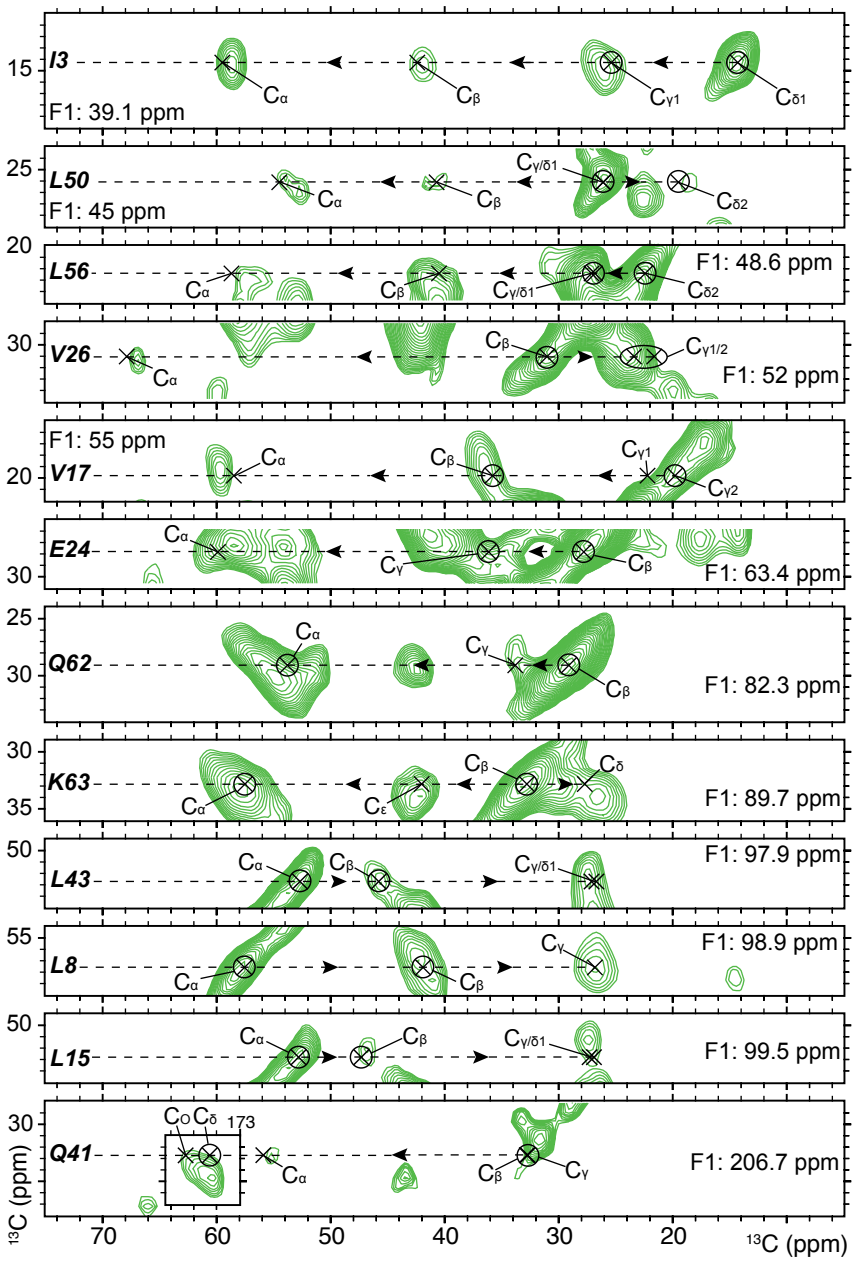


Figure 13: 2D (F_2, F_3) slices from the 3D (2Q-1Q-1Q) ^{13}C, ^{13}C spectrum corresponding to the unambiguously identifiable peaks. The F_1 (2Q) frequency, indicated in the slices corresponds to the chemical shifts of encircled peaks. The correlations in the F_1, F_3 slices are given in the Supporting Information (Figure S12-14)

Conclusions

We have achieved targeted isotope labelling of proteins in bacterial cells for DNP-ssNMR studies by almost completely eliminating cellular background resonances stemming from amino acids and lipids while obtaining best DNP enhancements. The 5 salient features of the sample preparation strategy involve 1) using the rifampicin approach to block native protein expression, 2) switching between labelled/ unlabelled medium at appropriate times, 3) using labelled amino acid mixtures, 4) employing optional cellular background deuteration and 5) finally choosing the right host strain.

During our studies, we discovered that the choice of host strain influences the resulting DNP enhancements when algal amino acid mixtures are used as a labelling source (Supporting Information, Table ST1). We speculate that the low enhancements obtained when algal amino acid mixtures are used in BL21 Rosetta 2 cells are related to the presence of high amounts of tRNA. We propose that similar DNP experiments using a range of *E. coli* BL21 (DE3) strains need to be performed and compared, to unravel the exact parameters that determine DNP enhancements when algal amino acid mix is used for labelling proteins.

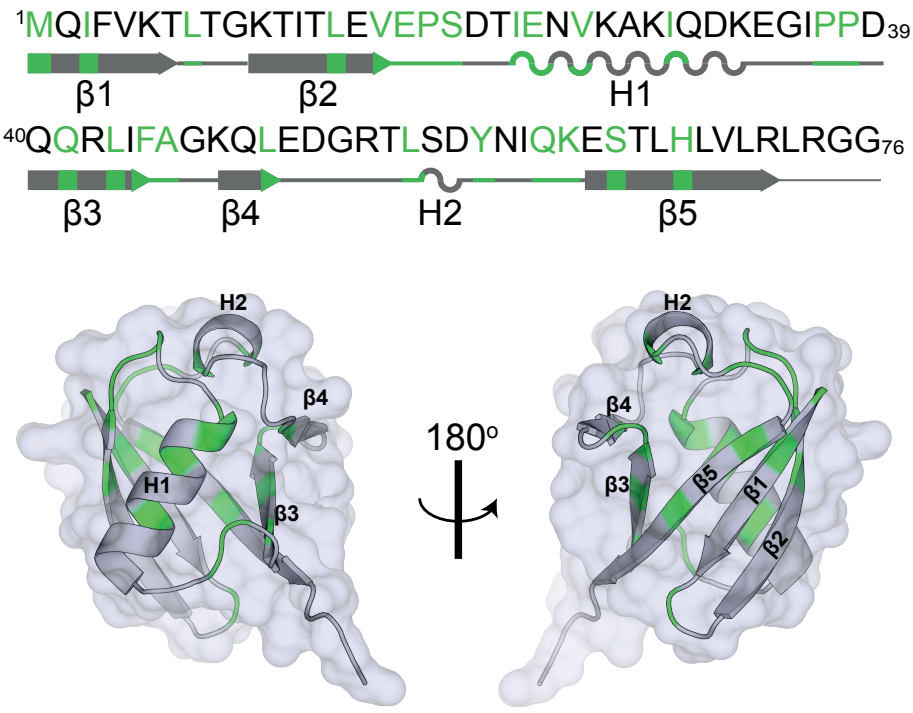


Figure 14: Residues for which correlations were identified for Ub in Lemo cells from 2D (Figure 12, S10) and 3D (Figure 13) NMR spectra

Using the most optimal sample of isotope labelled Ub in Lemo cells with >90% background deuteration, we recorded 2D and 3D NMR experiments (see section 4.5) that allowed us to identify correlations for 25 of 76 residues from all over the Ub structure (Figure 14, indicated in green). Not all residues could be identified due to spectral overlap and the increased linewidths at low-temperature DNP conditions. Amino acid selective labelling and other approaches to reduce this overlap in the future are discussed in chapter 7.

Supporting Information

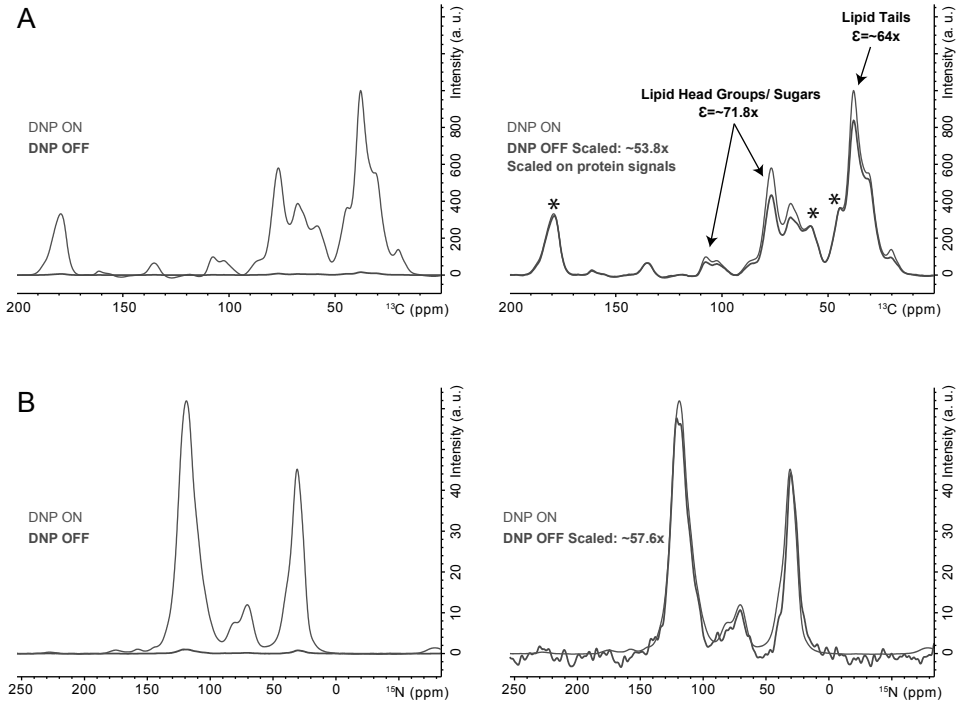


Figure S1: DNP enhancements obtained in different spectral regions on in-cell NMR samples relating to uniformly labelled Ub in BL21 *E. coli* Rosetta 2 cells. A) ^{1}H-^{13}C CP spectra- Lipid signals (both tails, sugars and head groups) clearly dominate the spectra as compared to the protein signals and also show higher enhancements (~53.8 for protein, ~64-71.8 for lipids and/or sugars). B) The ^{1}H-^{15}N CP spectrum also shows a comparable enhancement factor of ~57.6

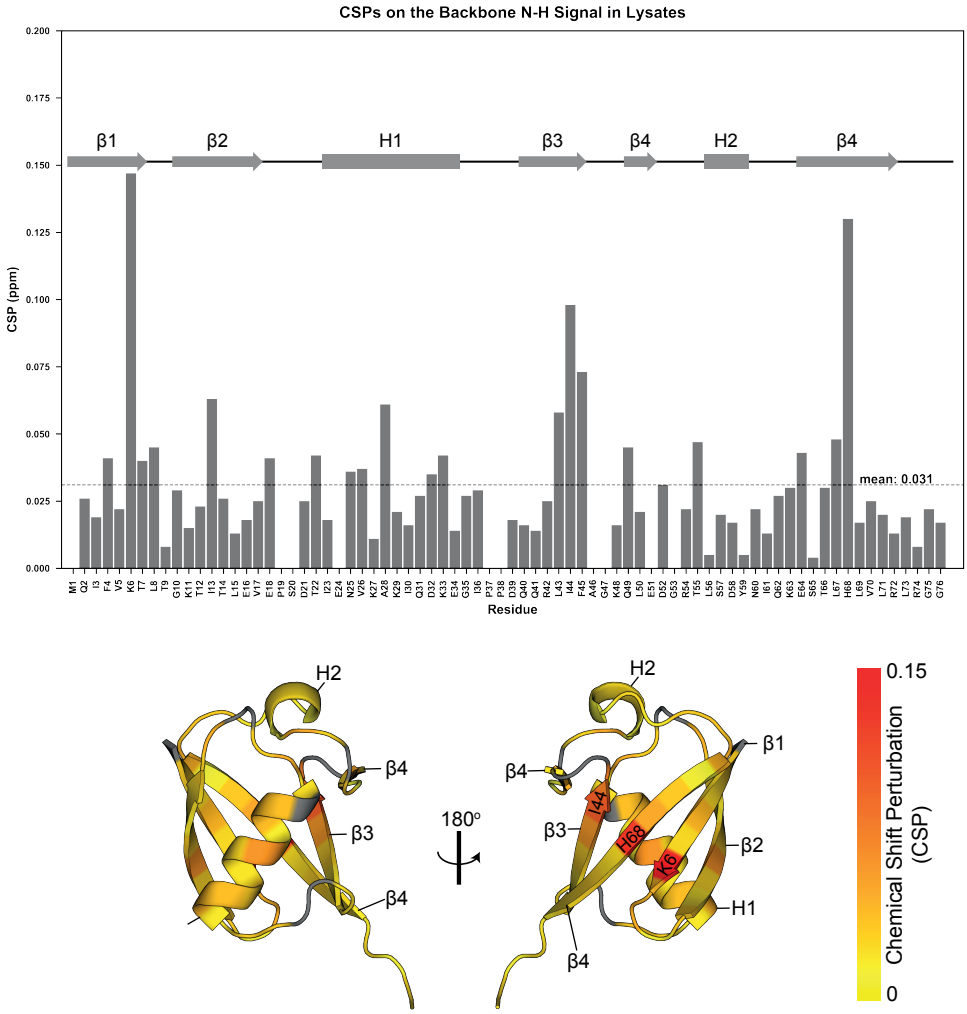


Figure S2: Residues close to the hydrophobic patch on the beta strands show the most significant (backbone amide) CSP in the lysate. Raw spectral data is shown in Figure 2 of main text. CSP was calculated using the following formula where $\Delta\delta_N$ and $\Delta\delta_H$ are the differences in ^{15}N and ^{1}H chemical shifts, respectively:

$$\mathbf{CSP} = \sqrt{(\Delta\boldsymbol{\delta}_{\mathbf{N}}/\mathbf{6.51})^2 + \Delta\boldsymbol{\delta}_{\mathbf{H}}{}^2}$$

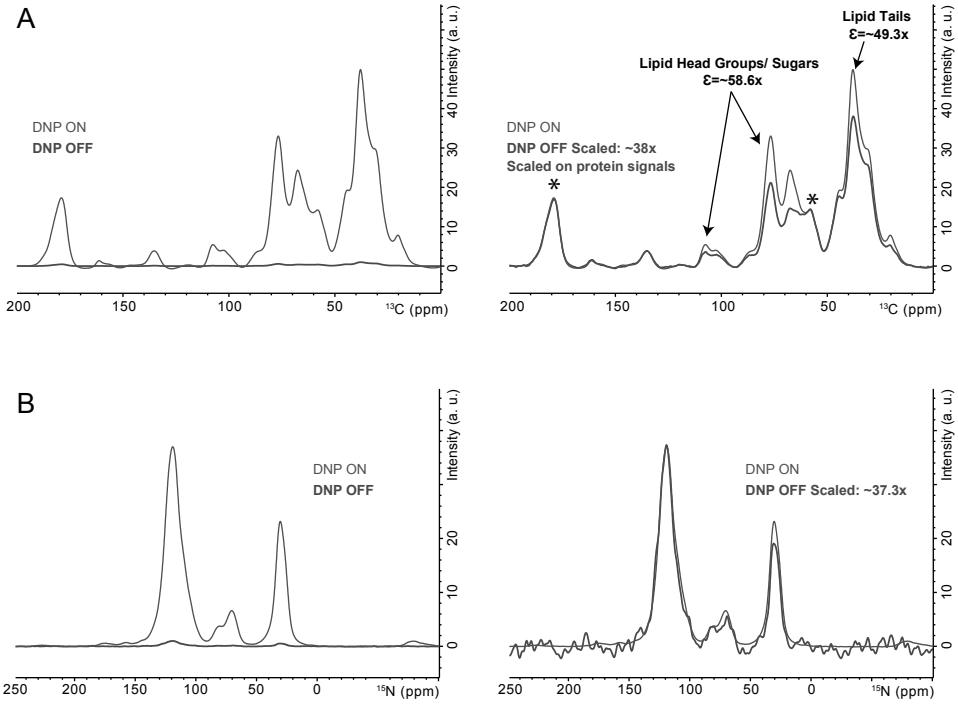


Figure S3: DNP enhancements obtained in different spectral regions on in-cell NMR samples pertaining to uniformly labelled Ub in BL21 *E. coli* Rosetta 2 cells with the "timed unlabelling" strategy. A) ^{1}H-^{13}C CP spectra- Lipid signals (both tails, sugars and head groups) clearly dominate the spectra as compared to the protein signals and also show higher enhancements (~38 for proteins, ~49.3-58.6 for lipids and/or sugars). B) The ^{1}H-^{15}N CP spectrum also shows a comparable enhancement factor of ~37.3

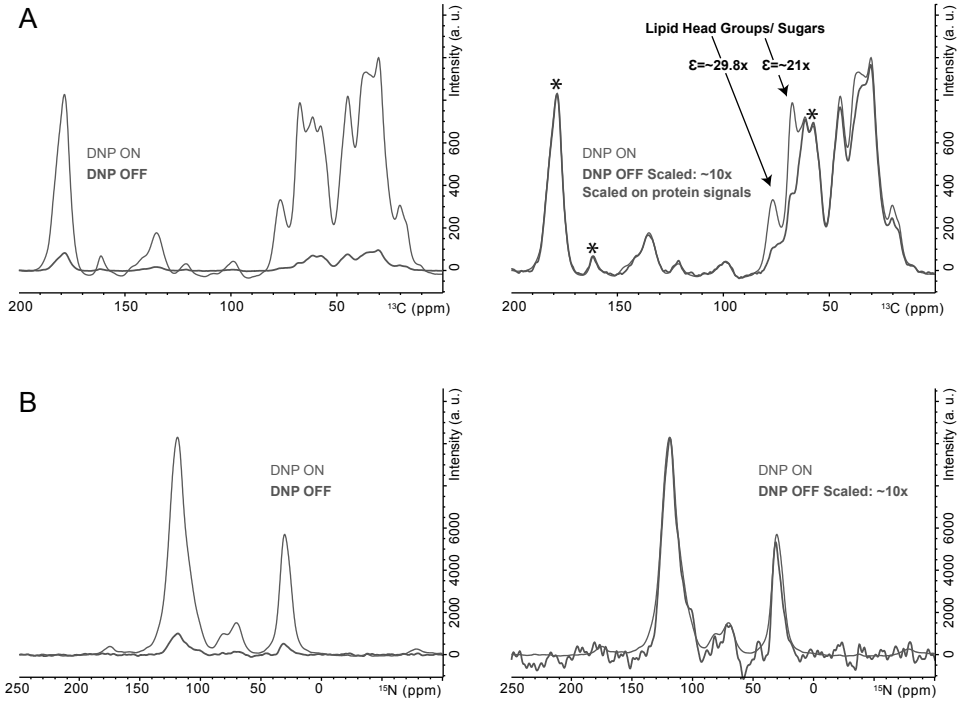


Figure S4: Poor DNP enhancements on in-cell NMR samples pertaining to Ub in BL21 *E. coli* Rosetta 2 cells with the "timed unlabelling" strategy and labelled algal amino acid mix as the labelling source. A) ^{1}H-^{13}C CP spectra- proteins (enhancement factor: ~10) clearly dominate the spectra as compared to the lipid head groups/ sugars (~21-29.8). B) The ^{1}H-^{15}N CP spectrum also shows a comparable enhancement factor of ~10.

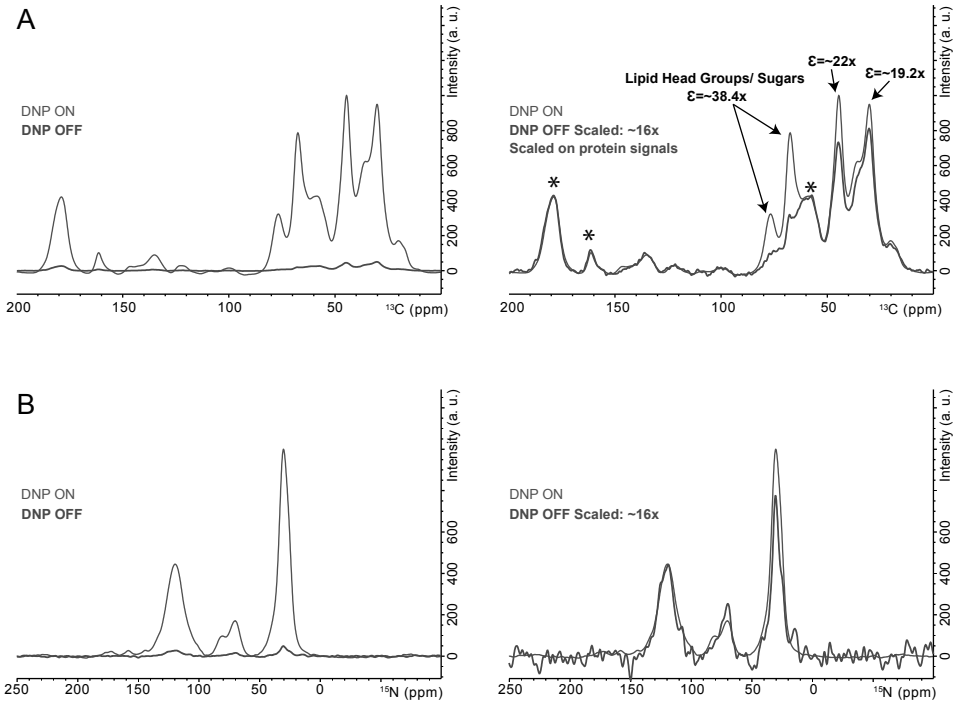


Figure S5: DNP enhancements on in-cell NMR samples pertaining to Ub in background deuterated BL21 *E. coli* Rosetta 2 cells with the "timed unlabelling" strategy and labelled algal amino acid mix as the labelling source. A) ^{1}H-^{13}C CP spectra- Proteins (enhancement factor: ~16) clearly dominate the spectra as compared to the lipid head groups/ sugars (enhancement factor: ~19.2-38.4). B) The ^{1}H-^{15}N CP spectrum also shows a comparable enhancement factor of 16.

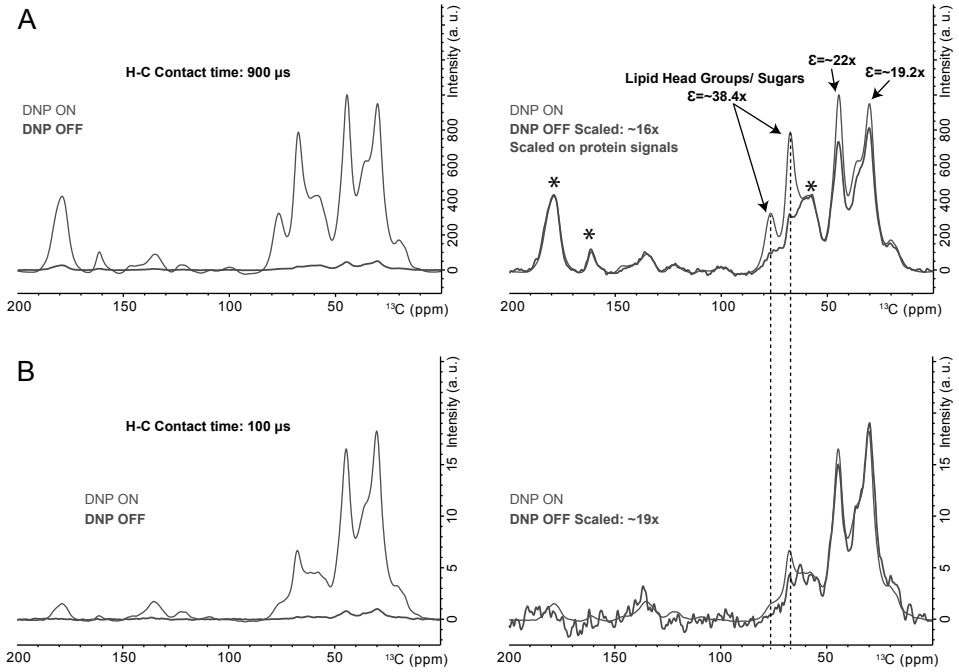


Figure S6: Effect of H-C long (A) and short (B) CP contact times on the reduction in DNP enhancements of background lipid head groups/ sugars. Note that protein enhancement factors (~19) are still comparable to the longer CP times (~16).

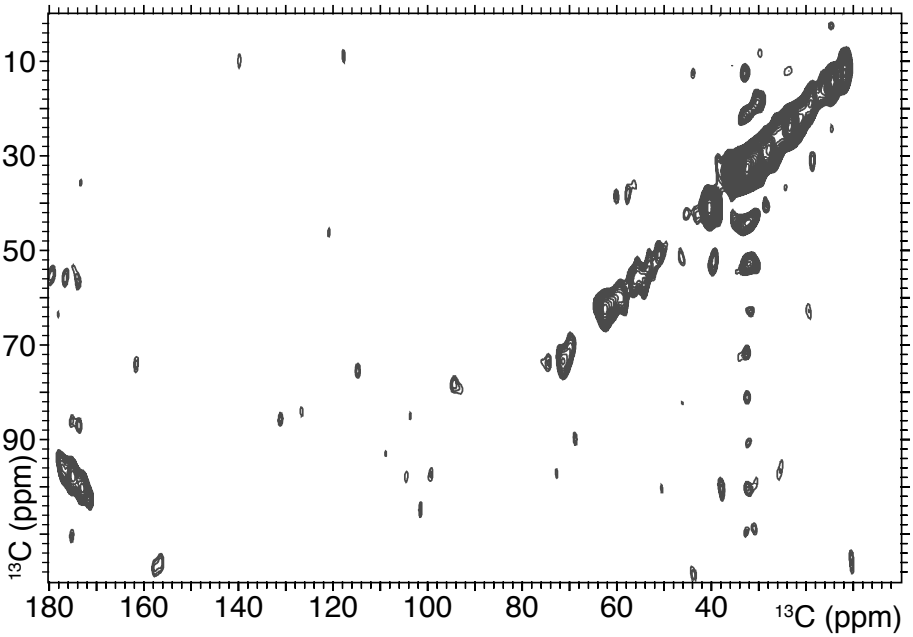


Figure S7: The ^{13}C, ^{13}C, correlated PDSD spectrum shows no labelled proteins in the insoluble fraction of cell lysates prepared from BL21 Lemo cells expressing ^{13}C, ^{15}N labelled Ub with ^{13}C, ^{15}N algal amino acid mix as the labelling source. The solution-state NMR spectrum of the soluble fraction is shown in Figure 10 & 11.

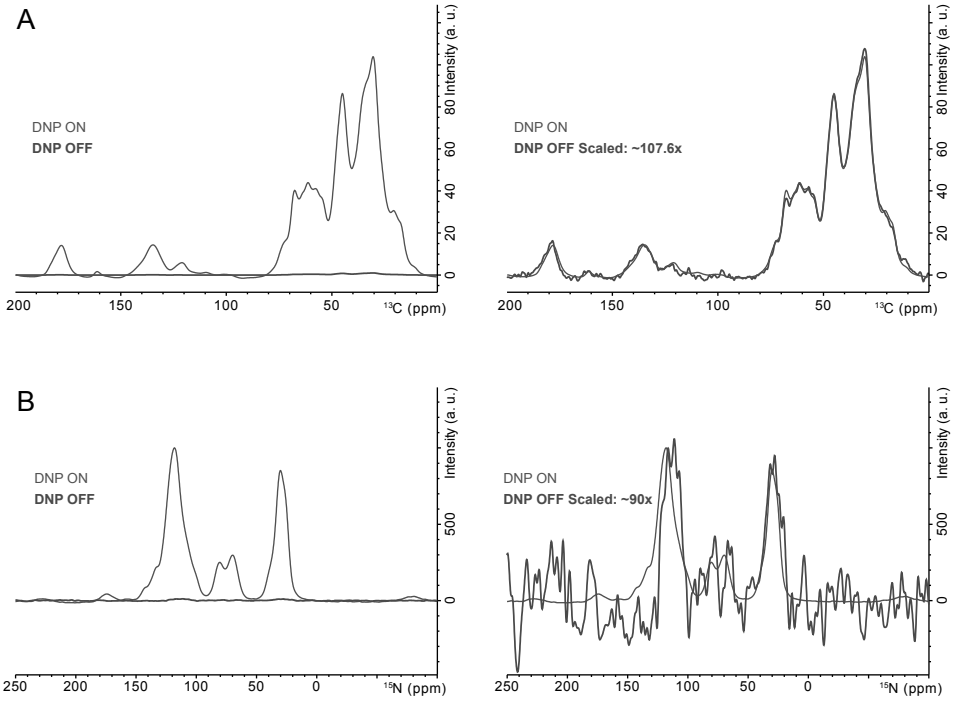


Figure S8: DNP enhancements on in-cell NMR samples pertaining to Ub in background deuterated BL21 Lemo cells with the "timed unlabelling" strategy and algal amino acid mix as the labelling source. A) ^{1}H-^{13}C CP spectra- proteins clearly dominate the spectra as compared to the lipid head groups/ sugars. B) The ^{1}H-^{15}N CP spectrum shows a comparable enhancement factor of ~90 x.

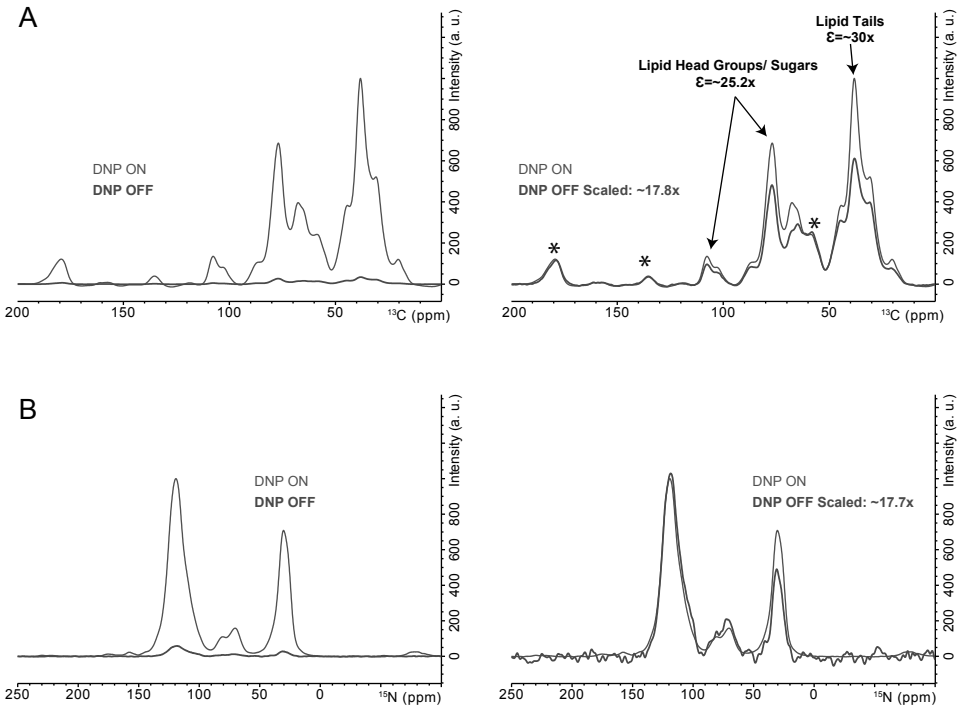


Figure S9: DNP enhancements in different spectral regions on in-cell NMR samples pertaining to uniformly labelled Ub in BL21 Lemo cells with the "timed unlabelling" strategy. A) ^{1}H-^{13}C CP spectra- Lipid signals (both tails, sugars and head groups) clearly dominate the spectra as compared to the protein signals and are also show higher enhancements. B) The ^{1}H-^{15}N CP spectrum also shows a comparable enhancement factor of ~17.7 x.

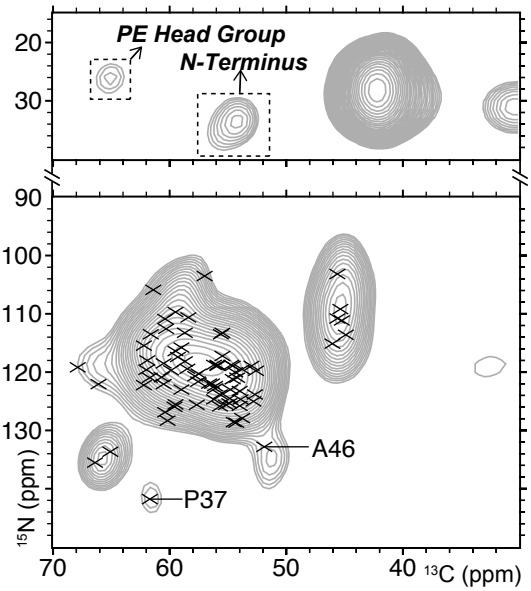


Figure S10: N-C correlation spectrum of Ub in deuterated BL21 Lemo cells. The backbone region matches perfectly with the chemical shift assignments (crosses) of Ub. Resolved peaks P37 and A46 are highlighted.

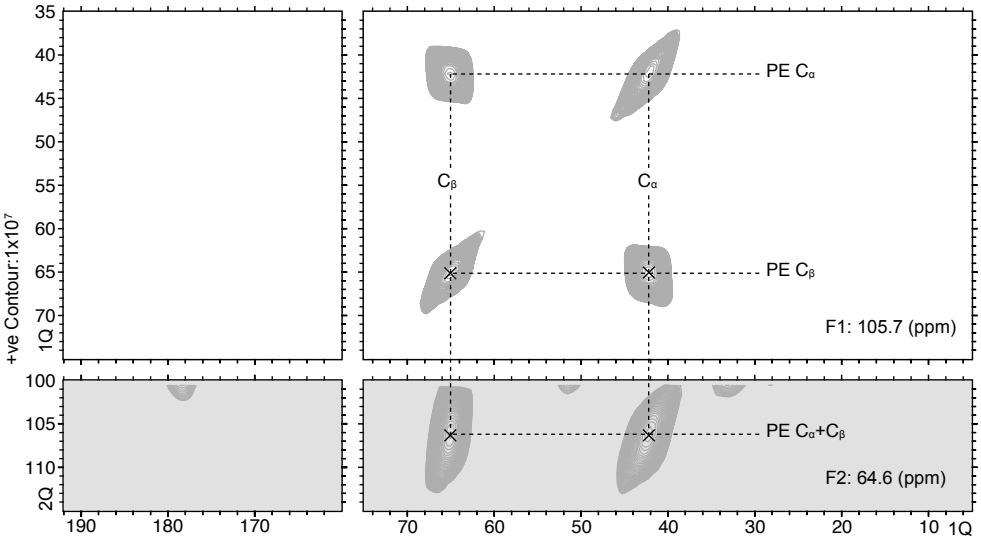


Figure S11: 3D DQ-SQ-SQ correlations for phosphatidylethanolamine.

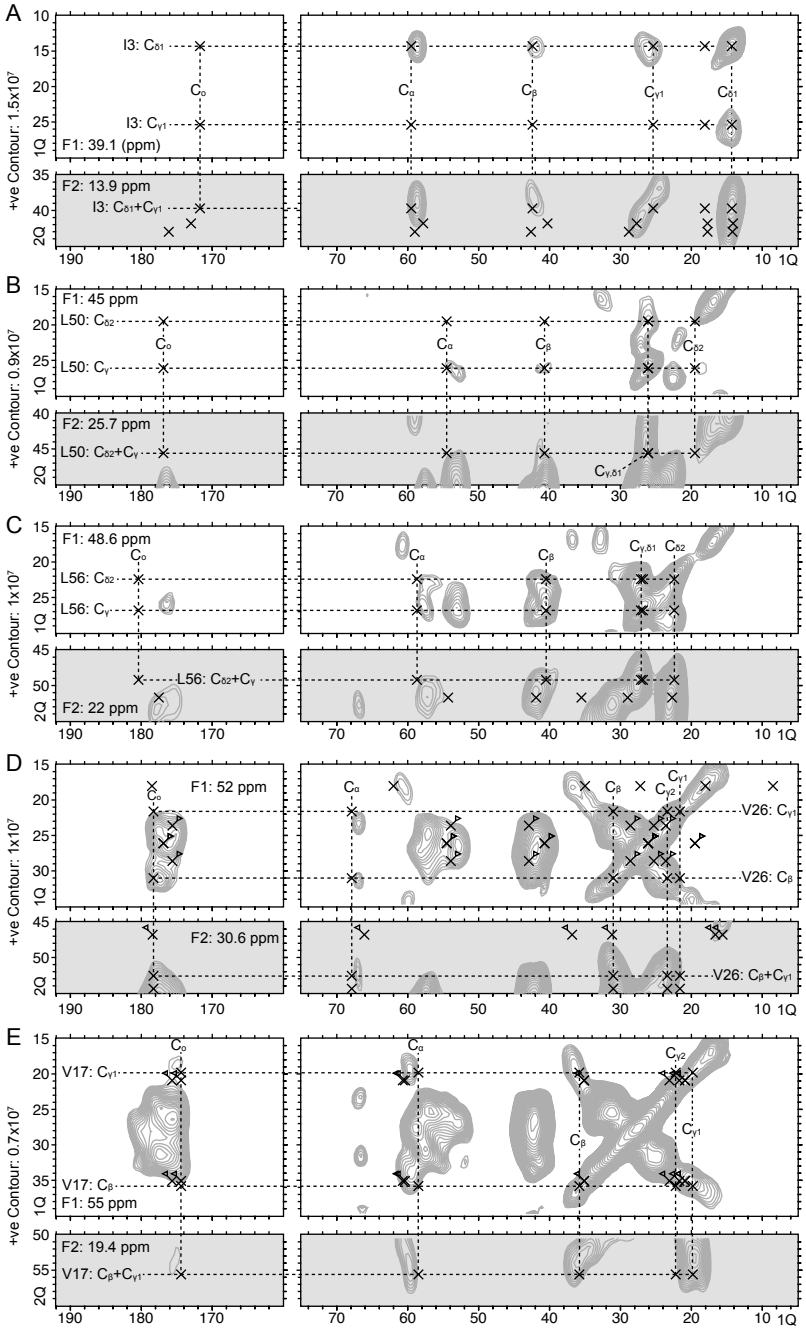


Figure S12: F_1-F_3 planes (grey background) correlated with F_2-F_3 planes for strips shown in Figure 13 for I3 (A), L50 (B), L56 (C), V26 (D) and V17 (E). Contour levels indicated here also apply to the respective slice in Figure 13.

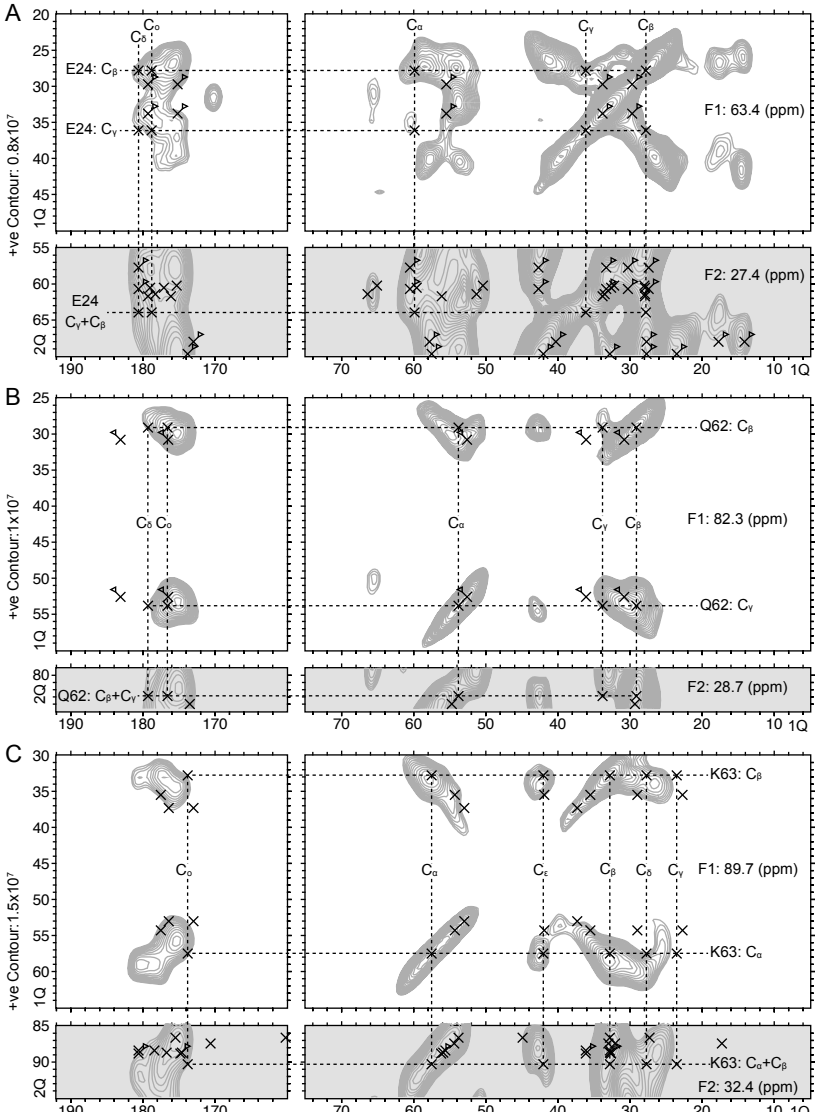


Figure S13: F_1-F_3 planes (grey background) correlated with F_2-F_3 planes for strips shown in Figure 13 for E24 (A), Q62 (B) and K63 (C). Contour levels indicated here also apply to the respective slice in Figure 13.

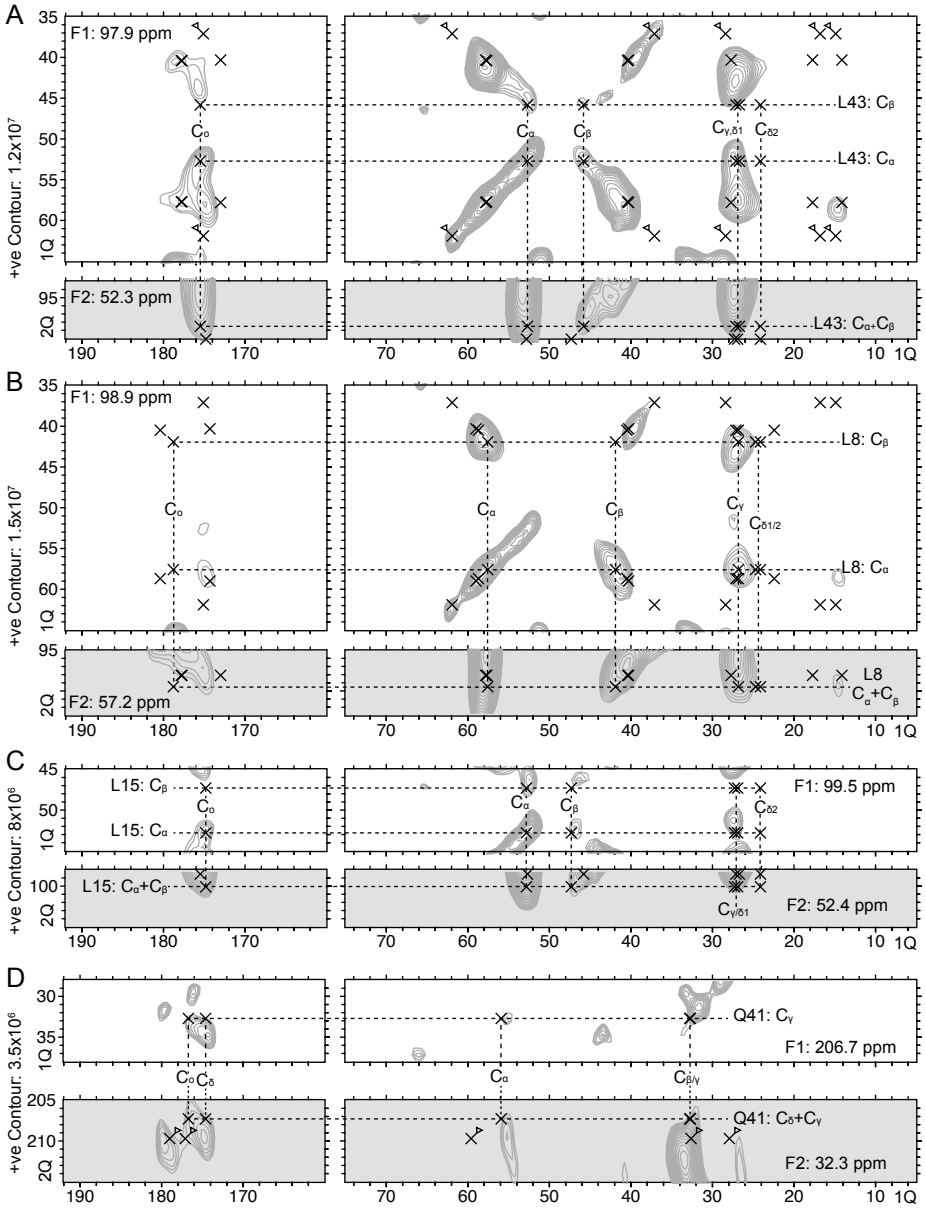


Figure S14: F_1-F_3 planes (grey background) correlated with F_2-F_3 planes for strips shown in Figure 13 for L43 (A), L8 (B), L15 (C) and Q41 (D). Contour levels indicated here also apply to the respective slice in Figure 13.

Table ST1: DNP enhancements (seen on H-C CP experiments) of various samples used in the chapter. "*" indicates timed unlabelling and "**Algal Mix+deut**." denotes background deuteration and use of algal amino acid mix as labelling source (see Main Text). Protein peaks were scaled on the C_α and C_α peaks, for chemical shift range of background peaks such as Lipid head groups, tails and sugars, refer to Figures S1, S3, S4, S5, S6, S8 and S9.

BL21(DE3) Strain	Labelling Scheme	DNP Enhancement Factor		
		Proteins	Lipid head groups/ Sugars	Lipid Tails
Rosetta 2	Conventional	53.8	71.8	64
	Conventional (*)	38	58.6	49.3
	Algal Mix (*)	10	21 - 29.8	-
	Algal Mix+deut. (*)	16 - 19	38.4	-NA-
Lemo	Conventional (*)	17.8	25.2	30
	Algal Mix+deut. (*)	107.6	-NA-	-NA-

References

1. Tabaka, M., Sun, L., Kalwarczyk, T. & Hołyst, R. Implications of macromolecular crowding for protein–protein association kinetics in the cytoplasm of living cells. *Soft Matter* **9**, 4386 (2013).
2. Siegal, G. & Selenko, P. Cells, drugs and NMR. *J. Magn. Reson.* **306**, 202–212 (2019).
3. Albert, B. J. *et al.* Dynamic Nuclear Polarization Nuclear Magnetic Resonance in Human Cells Using Fluorescent Polarizing Agents. *Biochemistry* **57**, 4741–4746 (2018).
4. Narasimhan, S. *et al.* DNP-Supported Solid-State NMR Spectroscopy of Proteins Inside Mammalian Cells. *Angew. Chemie - Int. Ed.* **58**, 12969–12973 (2019).
5. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
6. Serber, Z. & Dötsch, V. In-cell NMR spectroscopy. *Biochemistry* **40**, 14317–14323 (2001).
7. Baker, L. A., Daniëls, M., van der Cruijsen, E. A. W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62**, 199–208 (2015).
8. Medeiros-Silva, J. *et al.* 1 H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angew. Chem. Int. Ed. Engl.* **55**, 13606–13610 (2016).
9. Almeida, F. C. L. *et al.* Selectively labeling the heterologous protein in Escherichia coli for NMR studies: A strategy to speed up NMR spectroscopy. *J. Magn. Reson.* **148**, 142–146 (2001).
10. Galvão-Botton, L. M. . *et al.* High-throughput screening of structural proteomics targets using NMR. *FEBS Lett.* **552**, 207–213 (2003).
11. Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *J. Am. Chem.*

Soc. **123**, 8895–8901 (2001).
12. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *J. Magn. Reson.* **142**, 97–101 (2000).
13. Baldus, M., Petkova, A. T., Herzfeld, J. & Griffin, R. G. Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **95**, 1197–1207 (1998).
14. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4863–8 (2012).
15. Winget, J. M. & Mayor, T. The Diversity of Ubiquitin Recognition: Hot Spots and Varied Specificity. *Molecular Cell* **38**, 627–635 (2010).
16. Schwanhäusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205–209 (2009).
17. Kigawa, T. *et al.* Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **442**, 15–19 (1999).
18. Kaplan, M. *et al.* EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **167**, 1241-1251.e11 (2016).
19. Lacabanne, D., Meier, B. H. & Böckmann, A. Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *J. Biomol. NMR* **71**, 141–150 (2018).
20. Viennet, T. *et al.* Selective Protein Hyperpolarization in Cell Lysates Using Targeted Dynamic Nuclear Polarization. *Angew. Chemie Int. Ed.* **55**, 10746–10750 (2016).

Chapter 5:

Detecting the in-vivo assembly of an artificial metalloenzyme

This chapter is adapted from a part of:
"An Artificial metalloenzyme catalyzes a new to nature reaction in the cytoplasm of bacteria"
(Manuscript in Preparation)
Shreyans Chordia, Siddarth Narasimhan, Alessandra Lucini Paioni, Marc Baldus and Gerard Roelfes

Abstract

Artificial Metalloenzymes (ArMs) have enabled the catalysis of reactions which were previously thought to be infeasible or prohibitively expensive. Alternatively, synthetic biologists have been attempting to perform novel or cost-effective catalytic reactions using ArMs in live cells. In this study we examine the assembly of one such ArM consisting of a copper phenanthroline bound LmrR protein in *E. coli* cells using DNP enhanced solid-state NMR. Next our solid-state NMR studies, we have conducted high-resolution solution state NMR spectroscopy of cellular lysates. The latter experiments provided a validation for our in-cell data and yielded novel insights into the dynamics of the protein in *E. coli* cells.

Introduction

LmrR is a *Lactococcus lactis* derived transcription repressor that controls the transcription of the ABC type multidrug transporter, LmrCD[1]. It consists of a typical DNA binding domain comprising a helix-turn-helix (wHTH) motif and a long alpha helix in the C-terminus which aids in homodimerization and formation of a hydrophobic pore[2]. The core of the dimeric interface houses the multidrug recognition region (MDR) with two indole groups of W96 from both monomers that aid in stacking a range of planar aromatic compounds[2,3] (Figure 1). Drug binding in the MDR region prevents DNA binding, and thereby inhibits LmrR's functioning as a transcription repressor[2,4,5]. NMR studies have revealed that LmrR samples a variety of conformations in free solution enabling drug and DNA binding to occur through the conformation selection model, in which LmrR can adopt a slightly different conformation depending on the specific molecule it is bound to[3,5]. The promiscuous nature of LmrR also extends to DNA binding, wherein it not only binds to the operator sequence of the lmrCD gene of *Lactococcus lactis* but also non-specifically to fragments of DNA in the *E. coli* genome[5]. Current models suggest that the DNA bound conformation of LmrR is different from the confirmations it adapts when bound to different drugs, particularly in reference to the relative orientations of the H4 helix with respect to the wHTH region[3,5].

By carefully designing planar compounds that bind at the multidrug recognition region (MDR), synthetic biologists have been able to transform LmrR into artificial metalloenzymes, capable of catalyzing new-to-nature reactions[6–11]. One such metalloenzyme, comprising LmrR in complex with $Cu^{(II)}$ -phenanthroline (Cu^{II}-Phen) has been shown to catalyze the enantioselective Friedel-Crafts alkylation of indoles[6,10]. Thus far, such catalytic reactions have only been reported to occur in-vitro where the ArM is allowed to self-assemble in a suitable buffer. Performing the

same reaction inside cells (in-vivo) would provide the possibility to reduce costs associated with carrying out the aforementioned reaction by avoiding cumbersome purification and assembly steps. To carry out in-vivo catalysis, it is necessary to first confirm if the metalloenzyme can indeed functionally assemble inside cells. The in-vivo assembly of the LmrR Cu^{II}-Phen complex in *E. coli* is further complicated by the fact that Cu^{II} could readily be quenched to Cu^{I} in *E. coli*[12]. Here we describe a combined DNP-ssNMR and solution state NMR study that enabled us to probe the assembly of LmrR inside bacterial cells. Our results suggest that Cu^{II}-Phen bound LmrR properly assembles into a functional ArM inside *E. coli* cells. Our studies reveal interesting details about the nature of LmrR in the cellular environment that could potentially aid in future directed evolution efforts to improve catalysis.

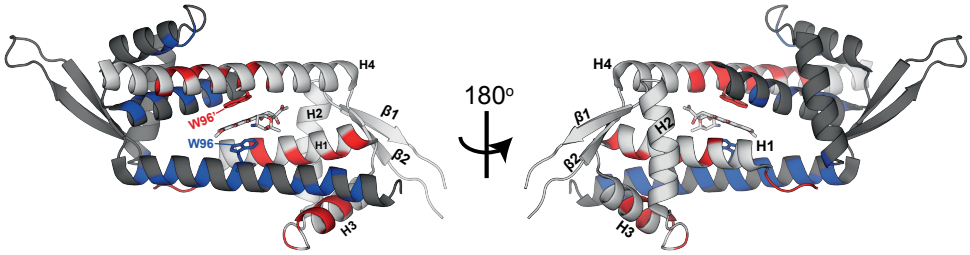


Figure 1: Crystal structure of LmrR complexed with the drug Daunomycin (PDB ID: 3F8F) where the different structural elements are indicated. The two monomeric subunits are colored in dark and light shades of gray, respectively. Residues that are located at the dimeric interface (5Å away from the other chain) are indicated in red and blue on either monomer of the two chains. DNA recognition occurs on helix H3 and the ligand as shown in the figure, is sandwiched between the indole groups of W96 by means of π-stacking.

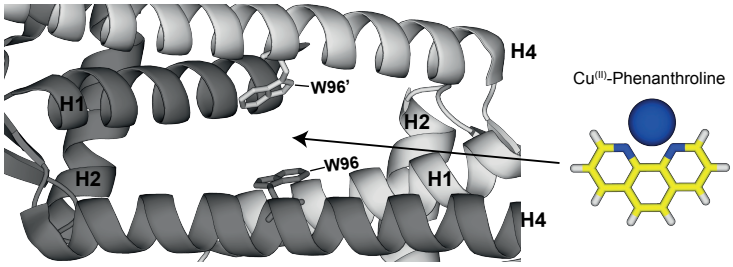


Figure 2: $Cu^{(II)}$-Phen binding site indicated in the crystal structure of LmrR (PDB ID: 3F8F)

Materials and Methods

Expression and purification of isotope labelled LmrR

E. coli BL21 Lemo cells were transformed with pET17b_LmrR_LM_A92E including a C-terminal Strep-Tag where LM stands for "Lysine mutant", which is a variant of LmrR with the K55D and K59Q mutations to reduce DNA binding[11], and A92E mutation that enables stronger binding to Cu^{II}-Phen (unpublished work). This variant will be referred to as LmrR in the following, for simplicity. A single colony of the fresh transformant was incubated in LB supplemented with 100 mg/L ampicillin. The culture was scaled up to 50 mL in LB and transferred to 1 L isotope labelled M9 medium when an OD600 of 0.5 was reached. For the samples used for in-vitro ssNMR studies, the M9 medium consisted of ^{13}C- glucose and ^{15}N- ammonium chloride. For the samples used to perform in-vitro solution-state NMR experiments we produced perdeuterated proteins using an M9 medium containing D_2O with ^{13}C-deuterated (d_6) glucose and ^{15}N- ammonium chloride. The cultures were induced at OD600 of 0.8-0.9 using 1 mM IPTG and expression was done overnight at 30°C. Cells were harvested by pelleting and resuspending in the lysis buffer containing 50 mM $NaHPO_4$ pH 8.0, 150 mM NaCl along with a tablet of EDTA-free protease inhibitor cocktail (Roche).

For purification, the cell suspension was first freeze-thawed and subjected to sonication. Sonication was done 10 times using 10-15 s long 13kHz pulses with a 30 s gap to prevent heating. The lysates were then passed through a 19G syringe multiple times and spun down at 20,000 G for 30 mins. The clear lysate was then loaded on to a 5 ml Strep-Tactin Superflow cartridge pre-equilibrated with the lysis/wash buffer- 50 mM $NaHPO_4$ pH 8.0, 150 mM NaCl. The column was washed with the wash buffer until no protein was seen in the UV-Vis output. The protein was eluted using 5 mM desthiobiotin in 50 mM $NaHPO_4$ pH 8.0, 150 mM NaCl. Prior to any subsequent experiments, the protein was buffer exchanged in 20 mM MOPS pH 7.0 with 150 mM NaCl by dialysis. Concentration of the yield was determined using a Nanodrop by measuring the absorbance at 280 nm ($\varepsilon_{280} = 25440\ M^{-1}\ cm^{-1}$).

Assembling the metalloenzyme in-vitro

DMSO stocks of 100-120 μM Cu^{II}-Phen (synthesized using an approach reported previously[6]) was freshly prepared. To assemble the metalloenzyme, a 1:10 molar ratio of LmrR:Cu^{II} -Phen mixture was prepared and incubated for an hour at room temperature. The excess Cu^{II}-Phen was subsequently washed away using multiple steps of concentrating and diluting in 20 mM MOPS pH 7.0 with 150 mM NaCl to 10 times the original volume using 15,000 Da centrifugal filters.

Protein precipitation for in-vitro ssNMR studies

20 mL of 5 mg/mL LmrR (with/ without Cu^{II}-Phen) was added to an equal volume of 50% w/v Polyethylene Glycol (PEG) 2200 Monomethyl Ether (MME) prepared in 20 mM MOPS pH 7.0 with 150 mM NaCl. After overnight incubation (16-18 hours) at 4°C, the microcrystalline LmrR (precipitated) was spun down, resuspended in 1 mL of fresh 25% PEG 2200 MME in 20 mM MOPS pH 7.0 with 150 mM NaCl and spun into 3.2 mm zirconia rotors for ssNMR experiments at ambient temperatures.

Targeted labelling of LmrR in *E. coli* cells with background deuteration

A single colony of freshly transformed BL21 Lemo cells harboring the plasmid containing LmrR was isolated and grown in an overnight LB culture with 100 mg/L ampicillin. 50 µL of the overnight culture was incubated in 50 mL of LB and grown up to OD600 of 0.05. The cells were then switched to an equal volume (50 mL) of unlabelled M9 medium prepared in D_2O using deuterated d_6 D- glucose and grown to an OD600 of 1.0. Expression was induced by adding 1 mM IPTG (final concentration) and the culture was incubated at 30°C for 30 minutes. The cells were then moved to an 18°C incubator and incubated for 5 mins to cool them down. Subsequently, 100 mg/L rifampicin (from a 1000x DMSO stock) was added and the flasks were incubated for another 15 mins. 500 mg/L of an ^{13}C-, ^{15}N- labelled algal amino acid mixture (devoid of N, Q, C & W amino acids) was then added to the culture medium to achieve protein selective labelling. Alternatively, for the sample preparations that required labelling of Tryptophan (W), the cells were switched to precooled (18°C) fully labelled M9 medium prepared in D_2O (supplemented with 100 mg/L rifampicin, 2 g/L of ^{13}C-D-glucose and 0.5 g/L of $^{15}NH_4Cl$). After 16h of expression at 18°C, the cells were switched to a precooled, unlabelled and deuterated M9 medium containing a reduced concentration of rifampicin (50 mg/L) for 4h. The final step involved transferring the cells to unlabelled and deuterated M9 medium devoid of rifampicin for the last 2h of expression. The rifampicin concentration was thus gradually reduced, to remove bulk rifampicin from the cells which might otherwise interfere with Cu^{II}- Phen.

Cu^{II}- Phen incubation in cells for DNP-ssNMR

After expression, the 50 mL culture was spun down and resuspended in 10 mL of precooled (4°C) 20 mM MOPS pH 7.0 with 150 mM NaCl prepared in D_2O with 10 µL of Cu^{II}-Phen stock (3.5 mM in deuterated d_6 DMSO). 10 µL DMSO was used instead of a Cu^{II}-Phen stock solution for the control samples. The incubation was carried out for 2h at 4°C. The cells were then pelleted, washed twice using ice cold 20 mM MOPS pH 7.0 with 150 mM NaCl prepared in D_2O. Half the pellet was

frozen for solution state NMR analysis of lysates and the other half was resuspended in 50 µL of deuterated DNP juice (60% deuterated d_7- and ^{12}C enriched glycerol, in D_2O) comprising of 15 mM AMUPol in 20 mM MOPS pH 7.0 with 150 mM NaCl. Washing buffers were prepared in D_2O, to prevent back-exchange of deuterated amide protons, thereby maximizing cell deuteration.

Lysates for solution-state NMR

The other half of the harvested cell pellet, which was frozen during DNP sample preparation, was thawed on ice. This was followed by the addition of ice-cold buffer of 20 mM MOPS pH 7.0 with 150 mM NaCl prepared in 10% D_2O. Subsequently, cells were resuspended and sonicated at 13 kHz with 10-15 s long pulses with a gap of 30 s to avoid excessive heating. After sonication, the cell debris was passed through a 19G syringe multiple times and finally spun down at 20,000 G for 30 mins. The pellet was discarded, and the clear lysate was spun down for another 30 mins to remove insoluble components.

DNP solid-state NMR experiments

DNP ssNMR experiments were conducted under low temperature (100 K) DNP conditions using a 3.2 mm triple-resonance (^1H, ^{13}C, and ^{15}N) magic-angle-spinning (MAS) probe head in a static magnetic field of 9.4 T, corresponding to proton/electron resonance frequencies of 400 MHz/263 GHz. All spectra were recorded at a MAS rate of 8 kHz using SPINAL-64 proton decoupling[13] at an r.f. field strength of 84 kHz.

Solid-state NMR experiments on in-vitro LmrR

We conducted in-vitro solid-state NMR experiments using a 3.2 mm triple-resonance (^1H, ^{13}C, and ^{15}N) E-free flip-angle MAS probe head in a static magnetic field of 16.4 T, corresponding to a proton frequency of 700 MHz. The set temperature was 253K and all spectra were recorded at a MAS rate of 13 kHz using a 77 kHz SPINAL-64 proton decoupling[13].

1D- ^1H-^{13}C Cross-Polarization experiments

The 1D H-C (CP) spectra were recorded using a 2 s recycle delay and 10 ms acquisition time. Cross-polarization was achieved using contact times of 300-800 µs for in-vitro ssNMR experiments and 900-1200 us for DNP-ssNMR experiments. 100 Hz line-broadening was applied prior to Fourier transformation.

2D- Proton-driven spin diffusion (PDSD) experiments in cells (DNP)

For experiments using cells labelled with the algal amino acid mixture, a H-C CP contact time of 1.2 ms and a mixing time of 30 ms was used. Acquisition

times were 15 ms (760 points used- 405.3 ppm spectral width) and 3.5 ms (200 data points- 280 ppm spectral width) for direct and indirect dimensions respectively. FIDs from three different experiments with 192, 64 & 64 scans were added.

For all the other cellular samples, the following experimental conditions were employed. A short H-C CP contact time of 100 μs was used to avoid excitation of background lipids/ sugars. In addition, a mixing time of 60 ms was employed. Acquisition times were set to 15 ms (600 data points used- 405.3 ppm spectral width) and 2.8 ms (160 data points- 280 ppm spectral width) for the direct and indirect dimensions, respectively. 128 scans were recorded.

All in-cell spindiffusion spectra were processed using a 0.25 π shifted sine squared window function on both dimensions.

2D- ^{13}C-^{13}C PARIS experiments for microcrystalline LmrR

A 300 μs contact time was used for H-C CP, followed by PARIS[14] recoupling for a mixing time of 30 ms using a ^{1}H radio frequency amplitude of 7 kHz. Acquisition times were set to 10 ms (1302 points used- 354.98 ppm spectral width) and 5.4 ms (380 data points- 200 ppm spectral width) for direct and indirect dimensions respectively. 64 scans were recorded and a 0.33 π shifted sine squared window function was applied on both dimensions.

Calculating T_1 relaxation times

T_1 relaxation rates for ^{1}H were calculated from the data obtained from a ^{1}H-^{13}C CP Saturation-Recovery experiment. 100 Hz line-broadening was applied prior to Fourier transformation of the series of 1D spectra. Peaks were manually selected and integrated, and the T_1 relaxation tool in TOPSPIN 4.0.7 was used to estimate the T_1 relaxation rates.

2D- ^{15}N-^{13}C correlation experiments

Conventional ssNMR on samples prepared in-vitro involved the following experimental parameters. A ^{1}H-^{15}N CP step of 1.1 ms and a SPECIFIC-CP[15] ^{15}N-^{13}C transfer time of 1.5 ms were employed. On-resonance frequencies were set to 115 ppm and 56 ppm for ^{15}N and ^{13}C dimensions, respectively to select for N-C_α transfer. A recycle delay of 2 s was used, and 512 scans were acquired. Acquisition times were set to 10 ms (800 data points- 338 ppm spectral width) and 9.2 ms (42 data points- 32 ppm spectral width) for the direct and indirect dimensions, respectively. In addition, linear forward prediction using 4 points and 4 coefficients were employed in the indirect dimension. The 2D spectra were processed using a 0.33 π shifted sine squared window function on both dimensions.

For DNP-ssNMR experiment on intact bacterial cells, the following

experimental conditions were used. A ^{1}H-^{15}N CP step of 800 μs and a SPECIFIC-CP[15] ^{15}N-^{13}C transfer time of 2 ms were employed. On-resonance frequencies were set to 75 ppm and 40 ppm for ^{15}N and ^{13}C dimensions, respectively. A recycle delay of 2 s was used, and 256 scans were acquired. Acquisition times were set to 10 ms (512 data points- 497 ppm spectral width) and 2.5 ms (36 data points- 180 ppm spectral width) for the direct and indirect dimensions, respectively. Linear forward prediction using 8 points and 8 coefficients was applied in the indirect dimension. The 2D spectra were processed using a 0.25 π shifted sine squared window function on both dimensions.

3D- ^{13}C DQ-SQ-SQ correlation experiment

The pulse program and processing parameters used here were exactly the same as the 3D experiment described in Chapter 2. Acquisition times were 10 ms (460 data points used- 307 ppm spectral width) in the F_3 1Q direct dimension, 1.5 ms (64 data points- 185 ppm spectral width) in the F_2 1Q indirect dimension and 0.9 ms (50 data points- 280 ppm spectral width) for the F_1 2Q indirect dimension. 64 scans were acquired. The spectrum was processed using a 0.28 π shifted sine squared window function in the direct dimension and a 0.25 π shifted sine squared window function in the two indirect dimensions.

Solution-state NMR experiments

We conducted solution-state NMR experiments at ambient temperatures (293-298 K) using triple channel (^{1}H, ^{13}C, and ^{15}N) probes. A static magnetic field of 21.1 T corresponding to a proton frequency of 900 MHz, with cryogenic probe was used to record 2D spectra. A static magnetic field of 17.6 T corresponding to a proton frequency of 750 MHz was used to record 3D experiments serving to obtain backbone assignments. We acquired 3D HNCA & HNCACB (TROSY) spectra of perdeuterated LmrR. We used the spectral assignments for ^{15}N and amide ^{1}H for wild type LmrR[3,5] to assign the ^{13}C chemical shifts in our spectra. The ^{15}N-^{1}H chemical shifts were kindly provided to us by the Shimada group (University of Tokyo).

For the 2D in-vitro ^{15}N TROSY experiments, 16 scans were recorded with a recycle delay of 0.3 s. On-resonance frequencies were set to 8.2 and 118.508 ppm for ^{1}H and ^{15}N respectively. Acquisition times were set to 80 ms (1024 data points used- 14 ppm spectral width) and 70 ms (512 data points used- 40 ppm spectral width) in the indirect and direct dimensions respectively. The 2D spectra were processed using a cosine squared window function on both dimensions.

For the 2D ^{15}N HSQC experiments, 16 and 32 scans were recorded in-vitro and in lysate samples respectively, with a recycle delay of 1 s. On-resonance frequencies were set to 4.714 and 118.508 ppm for ^{1}H and ^{15}N respectively. Acquisition times were

set to 71 ms (800 data points used- 16 ppm spectral width) and 20 ms (128 data points used- 35 ppm spectral width) in the indirect and direct dimensions respectively. The 2D spectra were processed using 0.4 π and 0.33 π shifted sine squared window function on both dimensions for in-vitro and in lysate spectra respectively.

Spectral visualization and analysis

All spectra were processed using TOPSPIN 4.0.7 except the 3D solution-state NMR spectra which were processed using NMRPipe[16]. Spectra were visualized and analyzed using the NMRFAM-SPARKY[17] software, and the peaks pertaining to every experiment was generated using FANDAS 2.0[18] software package. Weighted chemical shift perturbations (CSPs) were calculated using the following formula:

$$\mathbf{CSP} = \sqrt{(\Delta\boldsymbol{\delta}_{\mathbf{N}}/\mathbf{6.51})^2 + \Delta\boldsymbol{\delta}_{\mathbf{H}}{}^2}$$

where $\Delta\delta_N$ and $\Delta\delta_H$ are the differences in ^{15}N and ^{1}H chemical shifts, respectively.

Results and Discussion

5.1 in-vitro reference spectra of the metalloenzyme

Since we used a triple mutant version of LmrR (see Materials and Methods), we first obtained backbone assignments (Supporting Information, Table ST1) using 3D solution state NMR experiments on the perdeuterated protein. Spectral assignments remained incomplete due to the spectral overlap as well as the high α-helical content and rigidity (such as in helices H1 and H4) of the protein. Nevertheless, the obtained assignments were sufficient to analyze the in-vitro spectra of LmrR before and after addition of Cu^{II}-Phen in solution-state NMR ^{15}N- TROSY experiments (Supporting Information, Figure S1).

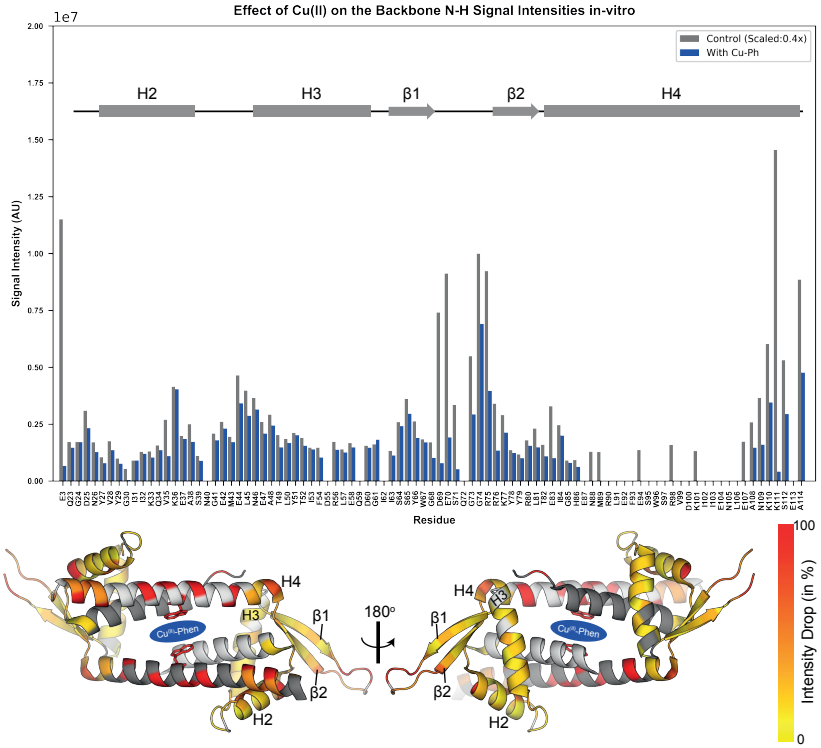


Figure 3: Residue-specific drop in NMR signal intensity of the backbone amides in LmrR due to paramagnetic quenching effects (in solution-state NMR). Data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). Residues that are not assigned are shown in gray on the structure, with the exception of W96, where we observed quenching of side-chain resonance. See Supporting Information, Figure S1 for the 2D spectrum.

Addition of Cu^{II}-Phen lead to a significant drop in signal intensities for backbone correlations throughout the protein peaks due to paramagnetic quenching

effects and chemical-shift perturbations (CSPs) (Figure 3&4). When plotted on the structure of LmrR, we found that the strongest quenching was observed for residues located in the H4, followed by the loop connecting the beta strands (Figure 3). While quenching of the residues in the H4 was expected due to the proximity to the Cu^{II}, we attribute signal loss for loop residues between the beta strands to changes in protein dynamics induced by the Cu^{II}-Phen binding. Similarly, the CSPs observed for residues located in the H2 & H3 (Figure 4) may be related to structural changes occurring upon binding, as seen before with LmrR bound to other drugs[2,3,5].

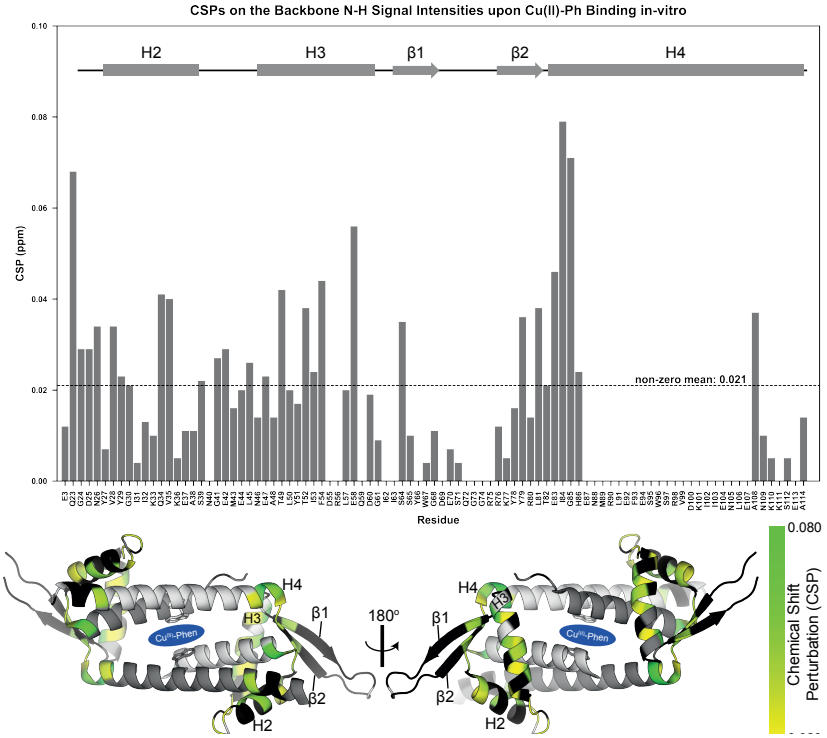


Figure 4: CSPs per residue of LmrR due to the binding of Cu^{II}-Phen (in solution-state NMR). Residues in black exhibit CSPs of less than 0.02 ppm on the backbone amide. Data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). See Supporting Information, Figure S1 for the 2D spectrum.

Next to the solution-state NMR data, we also recorded ssNMR experiments using microcrystalline LmrR. In Figure 5, 2D- ^{13}C,^{13}C correlated PARIS[14] ssNMR spectra are shown before and after the addition of Cu^{II}- Phen. In line with the NMR data in solution, we observed quenching of residues in the spectrum. In addition, several residues that remained unassigned in solution-state NMR due to their high rigidity or spectral overlap, could be tentatively assigned in the ssNMR spectrum.

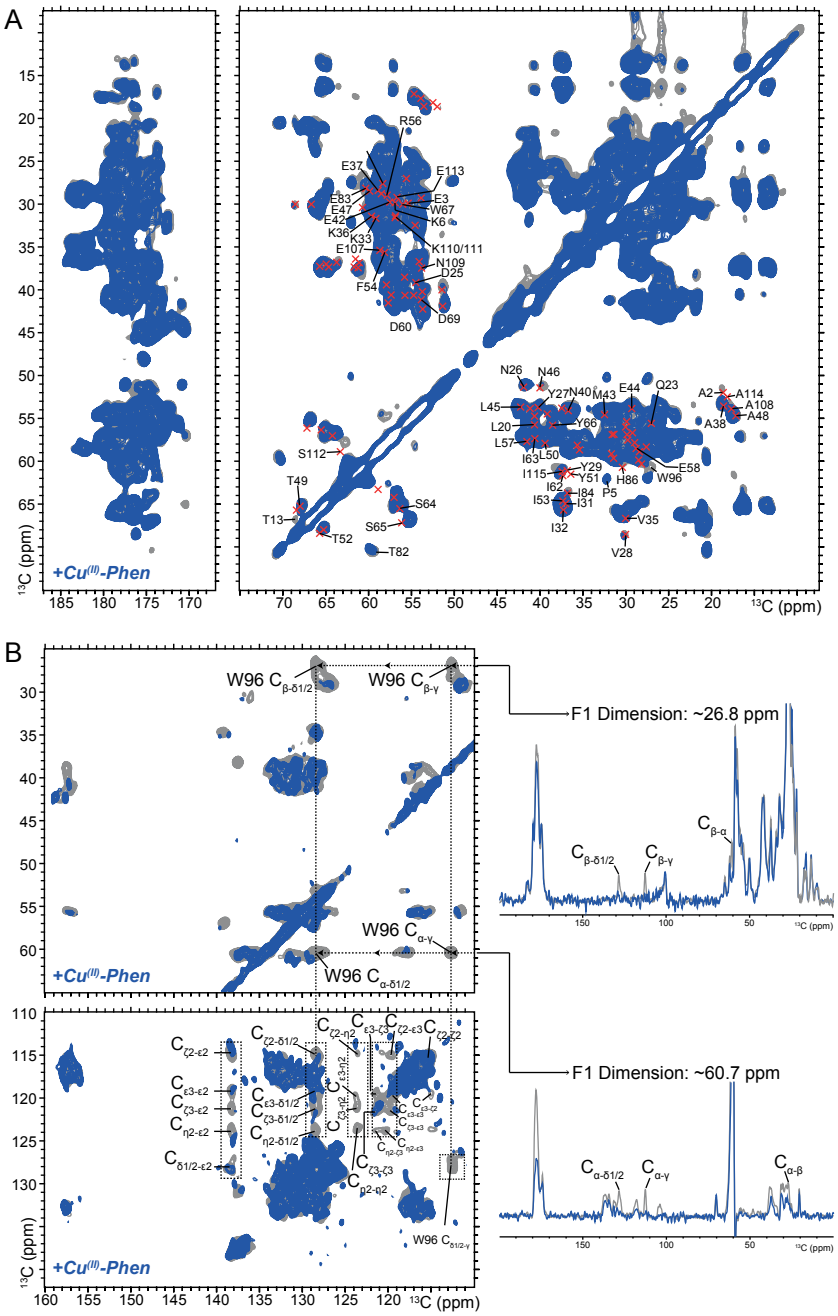


Figure 5: A) 2D $^{13}C,^{13}C$ correlated PARIS spectrum of LmrR in the microcrystalline state (with and without Cu^{II}-Phen). Red crosses are in-vitro solution NMR derrived assignments of the C_α-C_β correlations. The residues tentatively assigned in the ssNMR spectra are indicated without crosses. B) Analysis of the aromatic region of the spectrum shows selective and partial quenching of aromatic side chains of Trp, with the complete quenching of the backbone resonances (of W96) correlating with the aromatics.

These include T13 (the only unassigned helical Thr residue), T82 (the only Thr in the unstructured regions) and W96 (the only alpha-helical Trp) (see Figure 5). In fact, for both aromatic and backbone resonances of W96, we observed a clear decrease in the intensities, in the presence of Cu^{II} (Figure 5B). Besides the quenching of W96 peaks, which is correlated to the T_2 relaxation rates, we also observed faster T_1 relaxation rates in the presence of Cu^{II} (Figure 6).

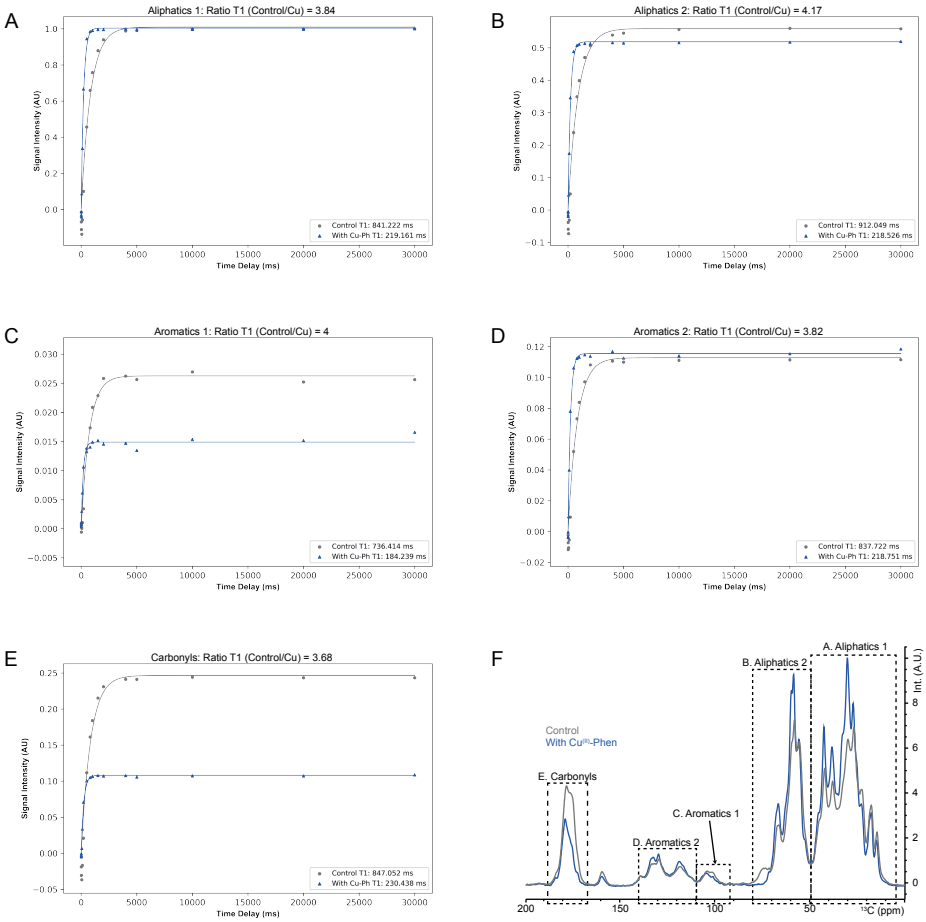


Figure 6: Relaxation profiles (A-E) of different spectral regions of LmrR in a ^{1}H-^{13}C cross-polarized saturation recovery experiment (F), before and after addition of Cu^{II}-Phen. In the presence of Cu^{II}, the relaxation times are between 3.68-4.17 times faster.

While we could identify several additional residues in our 2D ssNMR spectrum (Figure 5), a complete analysis of the entire protein was hampered by the reduced spectral resolution compared to the solution-state NMR data on perdeuterated LmrR (Figure 3&4). To address this shortcoming, future proton

detected ssNMR experiments using tailored deuteration schemes[19,20] could aid in improving the resolution and provide complementary information to solution-state NMR studies.

Taken together, the 2D ssNMR experiments served as a valuable in-vitro reference to our DNP-ssNMR experiments in cells described below as well as to the lysate studies using solution state NMR. We of course exclude the contribution of changes in dynamics affecting the T_1/ T_2 relaxation rates (also see part 1.2 of chapter 1).

5.2 in-cell ssNMR spectrum of LmrR confirms that the protein is well folded

In Chapter 4, we have demonstrated that using ^{13}C, ^{15}N- algal amino acid mixture as a labelling source largely eliminates all background signals from the cell. Therefore, to confirm that LmrR is properly folded in the cells and that it can be identified from the NMR spectra, we applied the same approach in the following (see also Materials and Methods). In line with our results on Ub discussed in the previous chapter, we obtained strong DNP enhancements of up to a factor 100 (see Supporting Information, Figure S2A) at 400 MHz DNP conditions. The high enhancements allowed us to acquire a 2D- ^{13}C, ^{13}C correlated PDSD spectrum (Figure 7), a 2D ^{15}N-^{13}C correlation spectrum (Figure 8) and a 3D ^{13}C correlated DQ-SQ-SQ spectrum (Figure 9). The 2D NMR spectra (Figure 7 & 8) were in very good agreement with the in-vitro spectrum of LmrR. However, the spectral resolution was reduced compared to our findings on Ub (Chapter 4) which we attribute to the larger size and the highly dynamic nature of LmrR.

In line with our 2D data sets, analysis of the 3D data sets suggested qualitative agreement between our in-vitro assignments and the backbone correlations observed in cells. Due to the lack of complete assignments, we could however obtain only four spectral strips in the 3D experiment which matched unambiguously with the backbone assignments determined in-vitro (Figure 9). We correlated the C_α, C_β assignments to the side chains and the carbonyl chemical shifts to confirm the respective amino acid type (Figure 9). To seek additional high-resolution confirmation that we indeed observed signals from LmrR in cells, we analyzed the lysates of the same batch of cells used for DNP analysis (see Materials and Methods). The solution state NMR spectrum confirmed that well folded LmrR is the only labelled molecule in the sample and additionally showed no visible signs of degradation (Supporting Information, Figure S3).

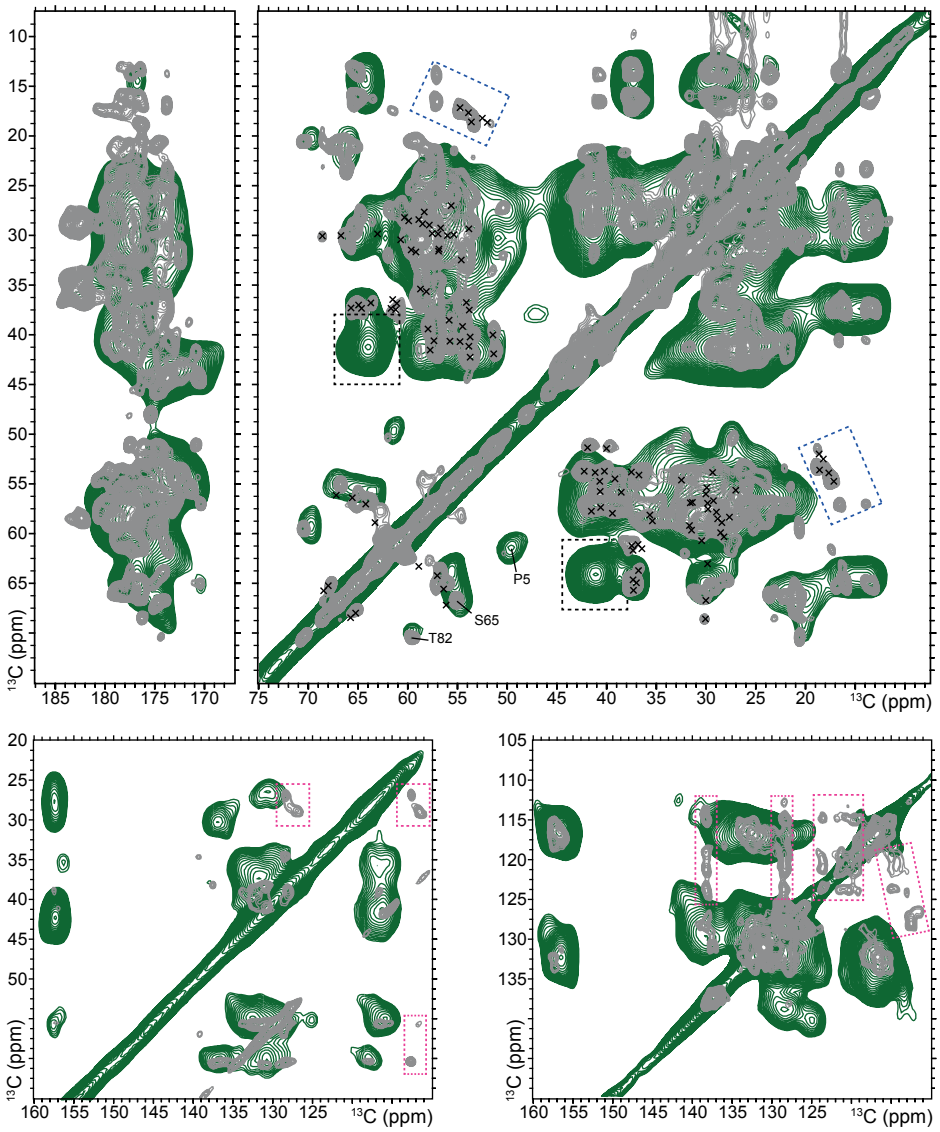


Figure 7: 2D ^{13}C-^{13}C, proton driven spin-diffusion experiment on LmrR in *E. coli* (green) overlaid with the 2D ^{13}C-^{13}C spectrum of microcrystalline LmrR (grey). C_α-C_β assignments obtained from solution NMR are superimposed (black crosses). The spectra are in good overall agreement with each other, and the most resolved peaks are indicated. Trp resonances were missing (red dashed boxes) because the algal mixture does not contain labelled Trp (see Materials and Methods). As often seen in DNP data sets, Ala methyl correlations were strongly reduced due to methyl group rotation at low temperatures (blue dashed boxes). Also, background labelling in the cells is strongly suppressed, except for a correlation marked using black dashed box, corresponding to the head group C_α-C_β correlation of phosphatidylethanolamine[21] (PE) (see also Figure 8 and Chapter 4- Supporting Information for tentative assignments).

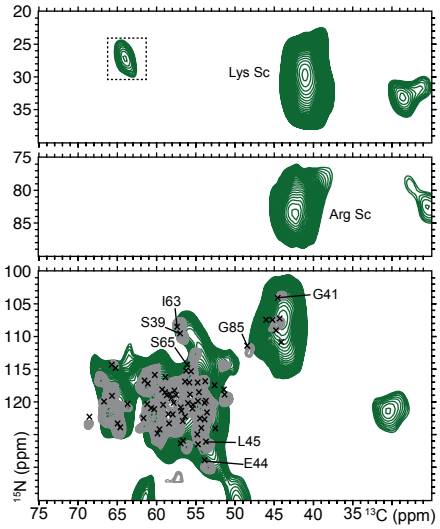


Figure 8: 2D aliphatic ^{15}N-^{13}C correlation experiment on LmrR in *E. coli* labelled using the ^{13}C-, ^{15}N-labelled algae sourced amino acid mixture (green). N-C_α assignments obtained from solution NMR are overlaid (black crosses). The overlaid spectrum in grey corresponds to the analogous in-vitro microcrystalline sample (only backbone region shown). The spectra are in good agreement with each other, strongly suggesting that the dominant species of LmrR in cells which is observed in NMR is LmrR that is properly folded. As with Figure 7, background labelling or free isotope labelled amino acids is strongly suppressed with the exception of a correlation that may stem from the lipid head group of phosphatidylethanolamine[21] (PE) marked by a black dashed box.

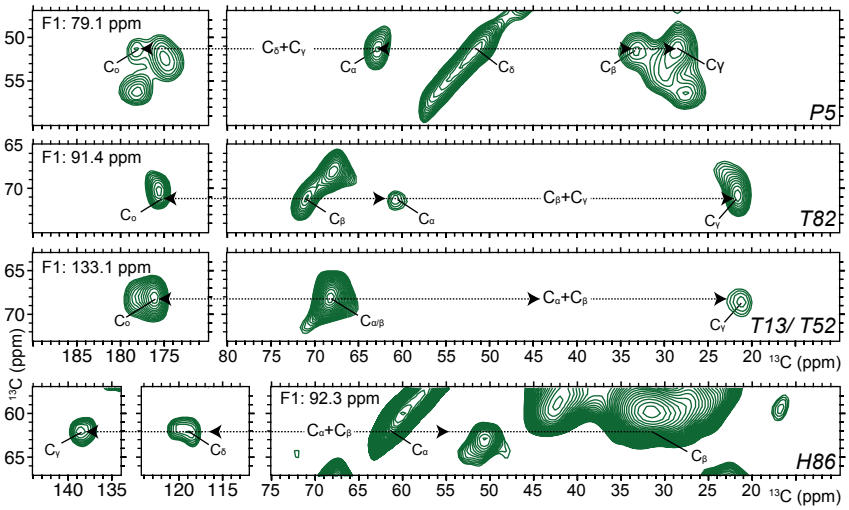


Figure 9: Most resolved 2D (SQ-SQ:F_2-F_3) slices of 3D ^{13}C (DQ-SQ-SQ) DNP-ssNMR spectrum recorded on LmrR in *E. coli* cells. The signals observed are in agreement with chemical shift assignments derived from solution NMR. The correlating resonances in the F_1, F_3 dimensions for the slices are shown in Supporting Information, Figures S4-6

5.3 Conventional labelling confirmed the assembly of the metalloenzyme

From the in-vitro ssNMR spectrum of LmrR (Figure 5), it is clear that the aromatic region, along with the changes in T_1 relaxation times (Figure 6) are the best indicators of Cu^{II}-Phen binding at the LmrR dimeric interface. Since the algal amino acid mixture was devoid of Trp, we proceeded with the conventional labelling scheme (with ^{13}C- glucose and $^{15}NH_4Cl$- see Materials and Methods) but using cellular background deuteration to detect Trp resonances.

In the control cells where Cu^{II}-Phen was not added, we obtained DNP enhancements of 107 on protein signals and enhancements ranging between 122 and 145 on the lipid head group/ sugar signals stemming from the cellular background (Supporting Information, Figure S2B). The DNP enhancement factors in the Cu^{II}-Phen treated cells unexpectedly dropped to a value of 32 on the proteins and to values between 66 and 93 on the lipid head group/ sugar background (Supporting Information, Figure S2C). To our knowledge, no DNP experiments have so far been reported in which both bi-nitroxide radicals as well as paramagnetic Cu^{II} are present. Thus, additional work may be needed to pinpoint any interference effects during preparation or DNP measurements.

Next we measured T_1 relaxation rates using the ^1H-^{13}C saturation recovery experiments. We find that the parts of the spectrum dominated by the protein signals exhibited a much faster T_1 relaxation rate (at least ~4x, as seen in the aromatics) than the parts of the spectrum dominated by non-protein components such as lipids or sugars (~1.62 times) (Figure 10). This finding provided a first hint that the metalloenzyme is properly coordinating the paramagnetic ligand in cells.

Next we focused on the aromatic region of the ^{13}C, ^{13}C correlated spindiffusion (PDSD) spectrum and observed that the aromatic resonances corresponding to Trp show partial quenching similar to the microcrystalline case (Figure 11). In the entire ^{13}C, ^{13}C correlated spindiffusion spectrum (Supporting Information, Figure S7), we found that the regions except aromatics and carbonyls were dominated by the background from the cells, preventing the detection of any resolved aliphatic backbone residues from the protein.

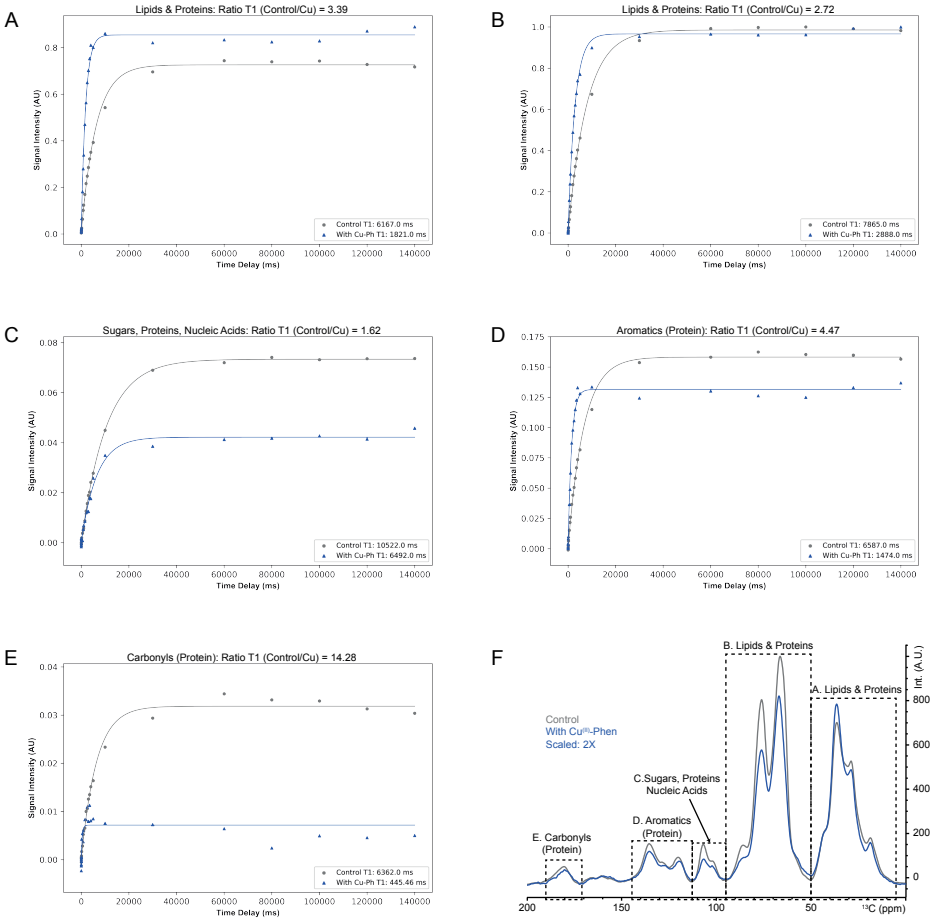


Figure 10: Relaxation profiles (A-E) of different spectral regions of the cellular samples in a ^{1}H-^{13}C cross-polarized saturation recovery experiment (F), before and after addition of Cu^{II}-Phen. In the presence of Cu^{II}, the relaxation times are fastest in the protein rich signals such as aromatics & carbonyls.

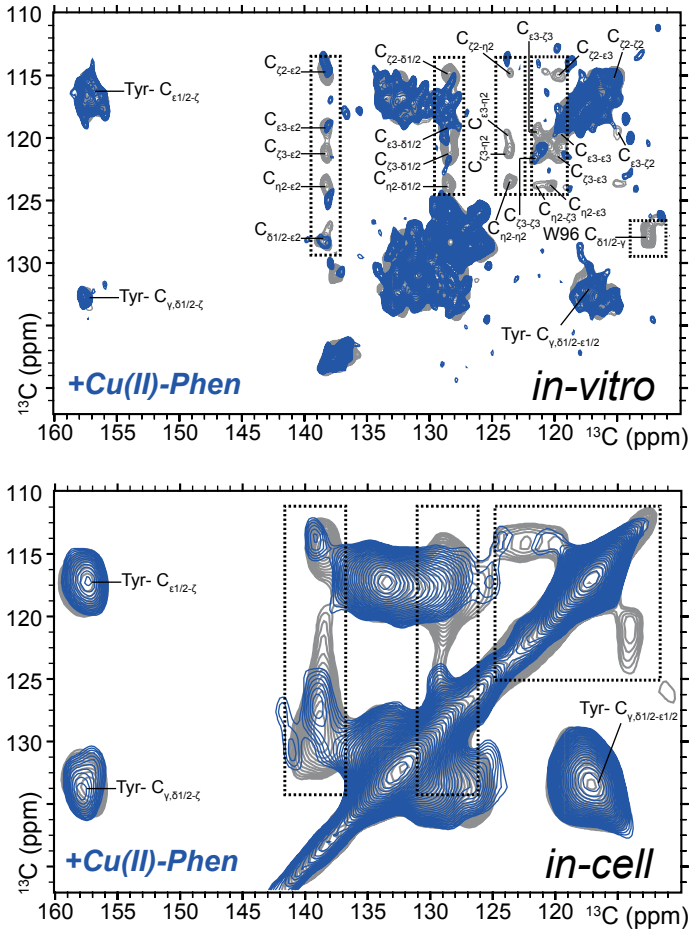


Figure 11: In-cell assembly of Cu^{II}-Phen was confirmed by selective partial (Cu^{II} induced paramagnetic) quenching of Trp aromatic resonances (indicated with dashed boxes) in the ^{13}C-^{13}C spindiffusion experiments, similar to the reference in-vitro experiments on the microcrystalline LmrR. See supporting information, Figure S7 for 1D slices and the complete in-cell spectrum.

To seek a complementary and a more high-resolution validation for our in-cell analysis, we performed ^{15}N-HSQC experiments on the lysates of identical cells used in DNP-ssNMR (Supporting Information, Figure S8A). The spectra overlaid with assignments confirmed that the protein signals were stemming from LmrR and we also observed partial quenching of several residues (Figure 12). Despite the quenching of resonances from the H4, we also saw a large-scale quenching of residues far away from the binding site, such as in the winged- HTH domain, H2 and H3 (Figure 12) that was not seen in-vitro (Figure 3). Notably, the CSP profile of the protein in lysate (Figure 13) was largely similar to the one in-vitro (Figure 4).

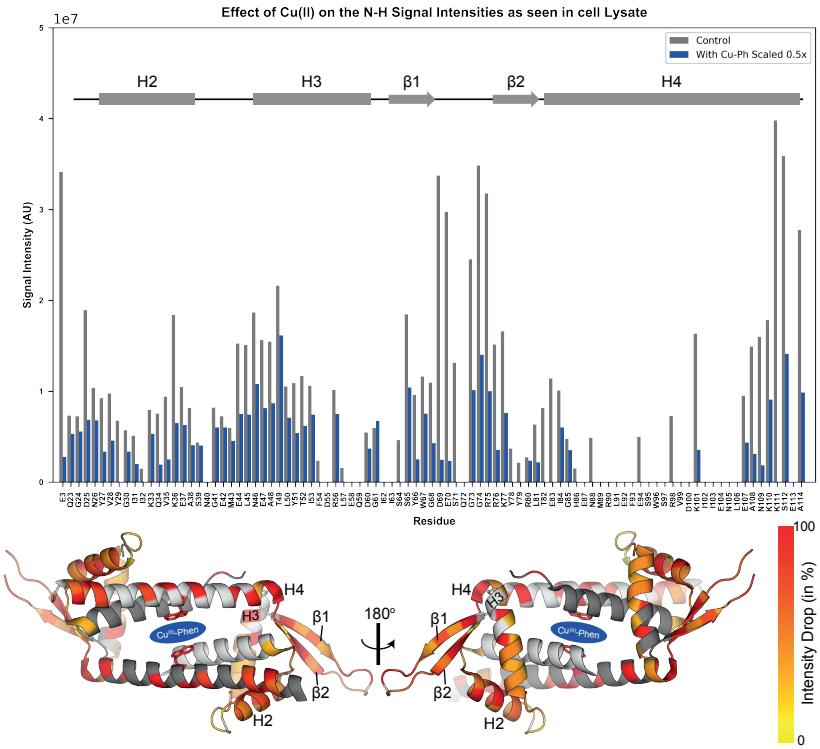


Figure 12: Residue-specific drop in NMR signal intensity of the backbone amides in LmrR in cell lysates due to Cu^{II}-Phen (in solution-state NMR). Data is plotted on the crystal structure of LmrR (PDB ID: 3F8F). Residues that are not assigned are shown in gray on the structure, with the exception of W96, where we observed quenching of side-chain resonance. See Supporting Information, Figure S8A for the 2D spectrum.

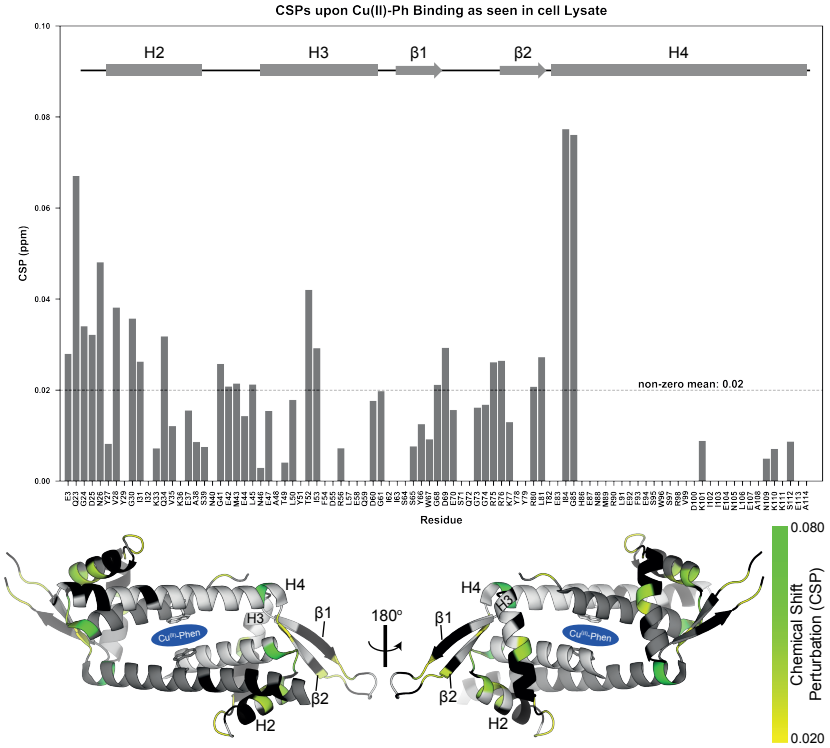


Figure 13: CSPs per residue of LmrR due to the $Cu^{(II)}$-Phen binding in lysates (in solution-state NMR). Residues in black exhibit CSPs of less than 0.02 ppm on the backbone amide. Data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). See Supporting Information, Figure S8A for the 2D spectrum.

A simple explanation for the significant loss in NMR signal intensity could be the presence of unbound Cu^{II} Phen around the protein. However, we expected that during the extensive washing steps (see Materials and Methods), the unbound bulk Cu^{II}-Phen would at least in part diffuse out of the cell or would be reduced to the diamagnetic Cu^{I} which would not quench signals. The bound Cu^{II}-Phen on the other hand would not be reduced to Cu^{I} because it is buried in the hydrophobic pocket of LmrR. Moreover, the aromatic Tyr signals in both in-vitro and in-cell ssNMR spectra did not show significant quenching (Figure 11), while they do decrease in intensity in the solution-state NMR data of the lysate (see intensities of Y27, Y29, Y66, Y78 & Y79 peaks in Figure 12). An alternative explanation for the reduction in intensity may hence be that rather than paramagnetic quenching, the binding of Cu^{II}-Phen leads to change in protein dynamics. This effect cannot be observed in the in-cell DNP-ssNMR experiments which are performed at 100K.

5.4 Lysate spectra of LmrR exhibit differences from in-vitro data

To distinguish the effect of ligand binding in Figure 12 & 13 from spectral alterations due to the presence of cell lysate components, we compared the spectra of LmrR devoid of Cu^{II}-Phen in solution to data obtained in lysate (Supporting Information, Figure S8B). We plotted the (absolute) intensity differences and CSPs on the structure (Supporting Information, Figure S9 & S10). We then set a cut off for significant intensity loss and CSP (Figure 14). We did not detect significant CSPs (Figure 14B). However, the intensities of a subset of residues that we had identified in Figure 12 showed significant changes in NMR signal intensity (Figure 14A). These included several residues located in the beta strands as well as part of helices H2 and H3 (Figure 14A) even before the addition of the ligand.

Among the residues that exhibited the strongest change in NMR signal intensity were Tyrosines (Y78 & Y79), the charged amino acids (R80 & E58) and the hydrophobic amino acids (I32, F54, L57 & I63). These residues are located distant from the Cu^{II}-Phen binding site but maybe sensitive to the cellular surrounding that may affect protein structure or dynamics either by direct binding or by unspecific interactions. Hence the even stronger changes observed in the NMR signal intensities upon ligand binding in lysates (Figure 13) may be the consequence of a combination of local binding and (possibly allosteric) interactions with components in the cellular milieu (see Figure 14A for comparison).

If future studies reveal that the observed weak intracellular interactions hinder catalysis, our findings may serve as useful leads to the rational (re)design of the metalloenzyme for optimized in-cell catalysis.

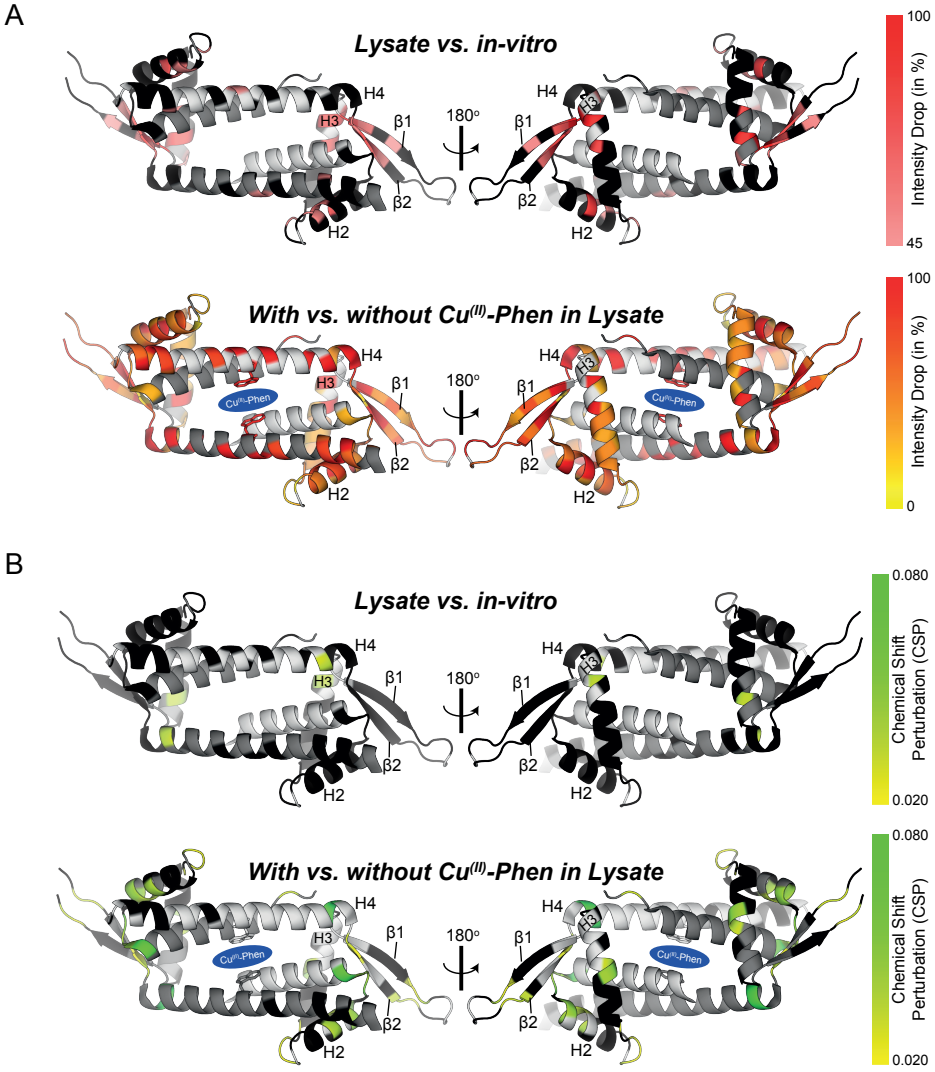


Figure 14: A) Residues displaying a significant drop in intensity (>45%) due to cellular components in the lysate (top). Shown in black are residues that show <45% drop in absolute signal intensities. Residue-specific losses in NMR intensity in lysates due to Cu^{II}-Phen addition (bottom, same as Figure 12) is shown for comparison. Residues that are not assigned are shown in gray on the structures. B) Residues displaying significant CSPs (>0.02 ppm) due to cellular components in the lysate (top). Residue specific CSPs in lysates due to Cu^{II}-Phen addition (bottom, same as Figure 13) is shown for comparison. Residues in black exhibit CSPs of less than 0.02 ppm on the backbone amide. In (A) & (B), data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). Also see Supporting Information S8-10 for raw data.

Conclusions

Artificial Metalloenzymes (ArMs) attract increased attention in the context of optimizing existing or developing new chemical reactions in-vitro and in-vivo. As for e.g. demonstrated by F. Arnold *et al.* such reactions may directly be designed in bacterial cells in a process coined directed evolution[22]. To fully exploit the potential of such approaches, suitable cellular structural biology methods that directly report on ArM structure and function in living cells are highly desirable and are currently lacking.

Here we have introduced a novel approach that combines the benefits of in-cell DNP-ssNMR and solution NMR on cell lysates to study a metalloenzyme that is based on LmrR as a protein scaffold and binds Cu^{II}-Phen in bacterial cells. Our data provide strong evidence of a proper folding and assembly of the complex in cells but also point to possible interactions between the protein and the cell interior. Further NMR and biochemical studies may provide additional insight on whether these interactions would hinder/ interfere with catalysis in whole cells. Such information may aid in further optimizing the protein in terms of catalytic conversion or enable a more precise control of protein interactions with the cellular background, in the spirit of directed evolution.

Supporting Information

Table ST1: Backbone assignments of LmrR used in this work.

Residue	N	H	C_α	C_β
A2	-	-	52	18.64
E3	120.47	8.3	55.37	29.91
P5	-	-	63.04	29.85
K6	126.35	7.92	56.92	29.85
L21	118.87	7.67	55.75	40.65
K22	123.28	8.11	58.41	-
Q23	115.3	7.35	55.61	27.01
G24	107.23	7.35	44.32	-
D25	122.56	8.45	54.48	39.16
N26	118.9	8.39	51.34	41.92
Y27	119.86	9.36	53.71	40.22
V28	122.29	9.34	68.56	30.04
Y29	117.18	7.7	61.1	36.84
G30	107.43	5.71	46.05	-
I31	123.28	7.83	64.95	37
I32	119.1	7.72	65.7	37.28
K33	120.44	7.61	59.16	31.66
Q34	118.85	7.97	58.33	27.63
V35	119.97	8.33	66.7	30.02
K36	124.24	8.1	59.64	31.44
E37	118.52	8.08	58.9	28.41
A38	121.62	8.3	53.58	18.6
S39	109.53	7.66	57.02	64.23
N40	-	-	54.11	36.73
G41	104.11	8.74	44.57	-
E42	118.97	7.42	57.53	29.78
M43	118.53	8.16	54.59	32.48
E44	128.94	8.37	53.86	29.32
L45	126.07	8.37	53.7	42.26
N46	118.12	7.92	51.44	40
E47	124.82	9.02	59.88	28.51
A48	119.92	8.06	54.72	17.14
T49	114.82	7.65	65.23	67.99
L50	121.93	7.47	57.94	39.42
Y51	116.73	8.1	61.52	36.43

Table ST1 (continued): Backbone assignments of LmrR used in this work.

Residue	N	H	C_α	C_β
T52	114.31	7.48	65.72	68.43
I53	123.89	7.57	64.65	37.34
F54	119.45	8.86	58.12	35.66
R56	120.17	7.04	57.83	28.97
L57	118.26	8.28	57.72	41.53
E58	121.88	8.78	58.51	28.81
D60	117.03	7.79	54.75	40.66
G61	107.46	7.8	45.23	-
I62	122.47	8.1	61.67	37.34
I63	108.5	6.8	57.37	40.59
S64	116.88	9.29	56.37	65.56
S65	114.31	8.13	56.13	67.2
Y66	117.07	8.32	55.82	38.54
W67	120.19	8.66	56	30.02
G68	110.86	9.33	44.1	-
D69	120.14	8.35	53.86	41.14
E70	122.23	8.5	56.37	-
S71	118.11	8.38	59.3	-
G73	109.1	8.12	44.77	-
G74	109.76	7.9	76.64	-
R75	121.04	8.22	55.45	-
R76	124.01	8.63	54.3	-
K77	126.51	8.22	54.74	-
Y78	125.95	8.44	56.62	-
Y79	115.22	8.49	56.25	-
R80	117.44	8.52	52.6	-
L81	124.99	8.52	54.81	-
T82	115.86	7.87	60.24	-
E83	121.12	9.02	60.32	28.17
I84	120.3	7.89	63.7	36.78
G85	111.42	8.05	48.46	-
H86	120.95	8.55	60.71	30.42
E107	118.98	7.79	58.7	35.38
A108	122.67	7.94	53.88	17.66
N109	116.85	8.13	53.81	37.49
K110	121.38	7.65	56.86	31.38
K111	120.88	8.01	56.91	31.57
S112	116.24	8.08	58.89	63.3
E113	122.94	8.16	56.68	29.23
A114	124.08	8	52.51	18.19
I115	120.36	7.87	61.22	37.45

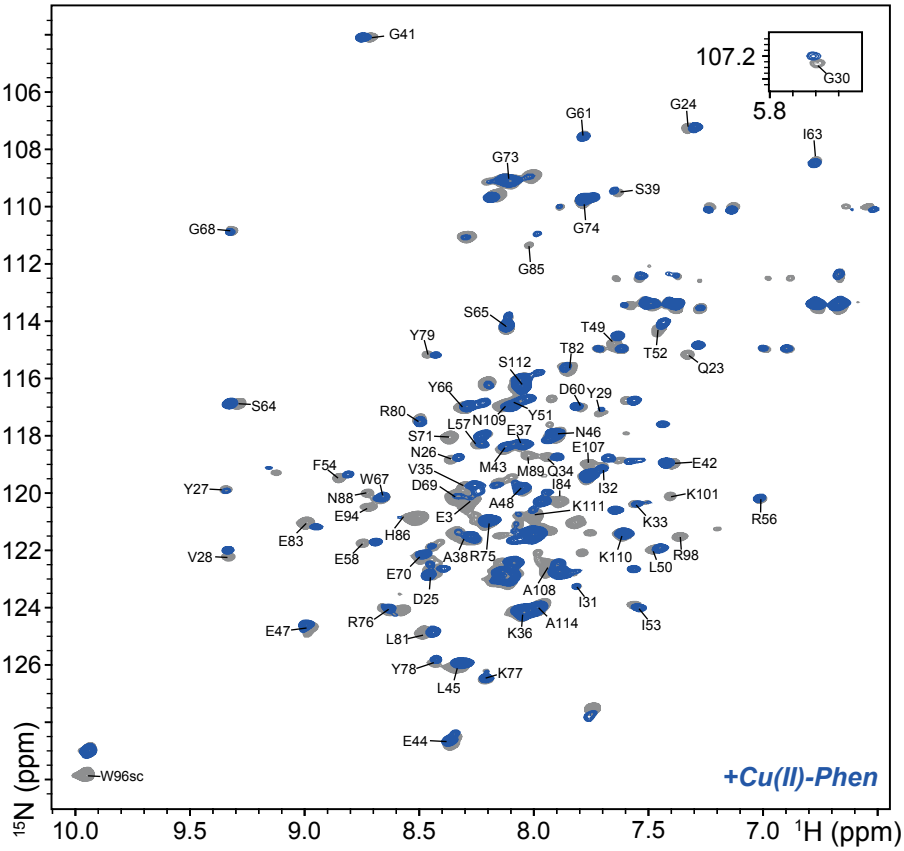


Figure S1: Solution state ^{15}N- TROSY experiment on perdeuterated LmrR with and without Cu^{II}-Phen. Note that W96 side chain is a tentative assignment on the basis of strong signal loss after the addition of Cu^{II}-Phen.

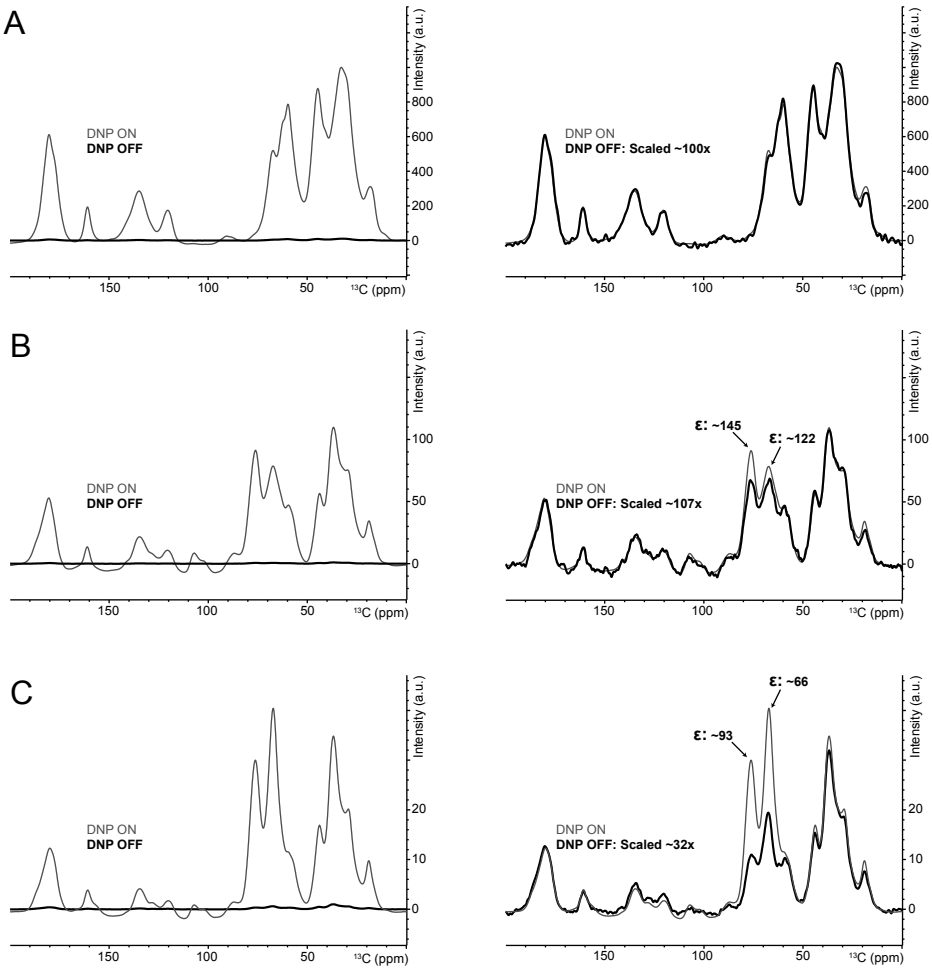


Figure S2: DNP enhancements, scaled on protein signals in the in-cell DNP samples referred to in the main text. A) Obtained on deuterated cells where an algal amino acid mixture was used as the labelling source. B) & C) Measured on deuterated cells where ^{13}C Glucose and ^{15}N Ammonium Chloride were used as an isotope labelling source. In addition, when Cu^{II}-Phen was introduced to the cells (C), enhancement dropped by a factor ~3 on the proteins and a factor ~2 on the non-protein background.

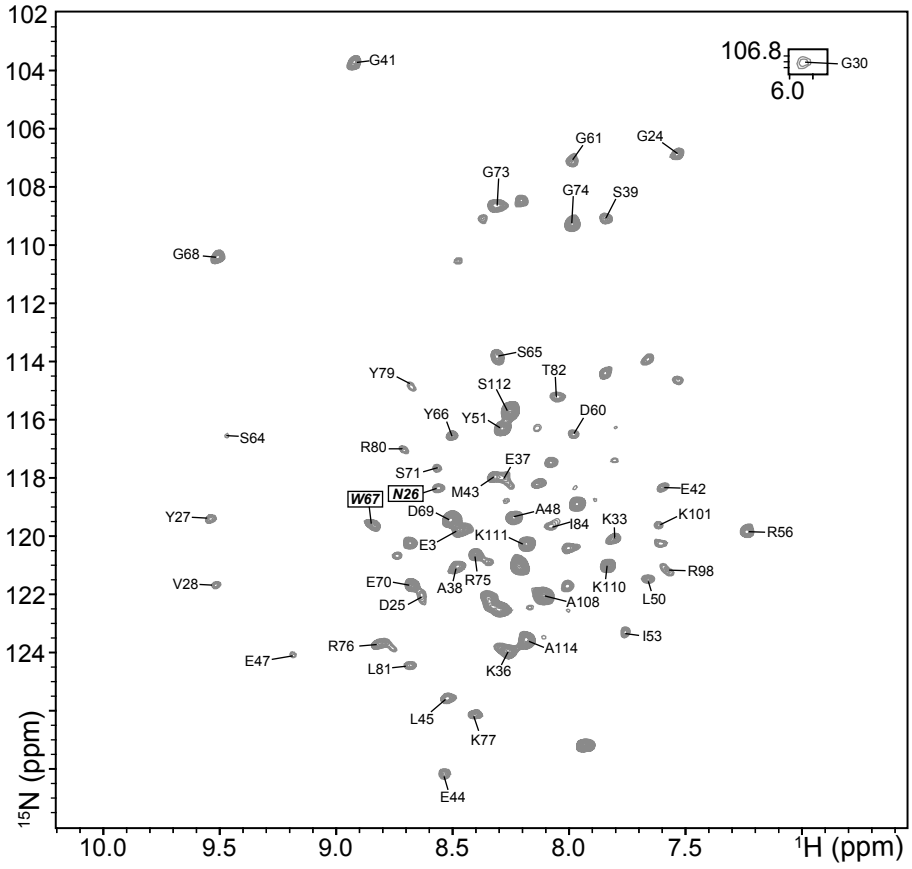


Figure S3: ^{15}N-^{1}H HSQC Experiment on cell lysates containing LmrR in-vivo labelled using the ^{13}C-, ^{15}N- labelled algae sourced amino acid mixture. The lysates were prepared from identical cells used in DNP-ssNMR experiments used for the 2D ^{13}C- spin diffusion and 3D ^{13}C- DQ-SQ-SQ experiment. All Asn, Gln and Trp resonances except for the W67 and N26 backbone amide are absent. Labelling of W67 and N26 can be attributed to scrambling, in specific transamination reactions from Glu and Asp that have been reported to occur[23].

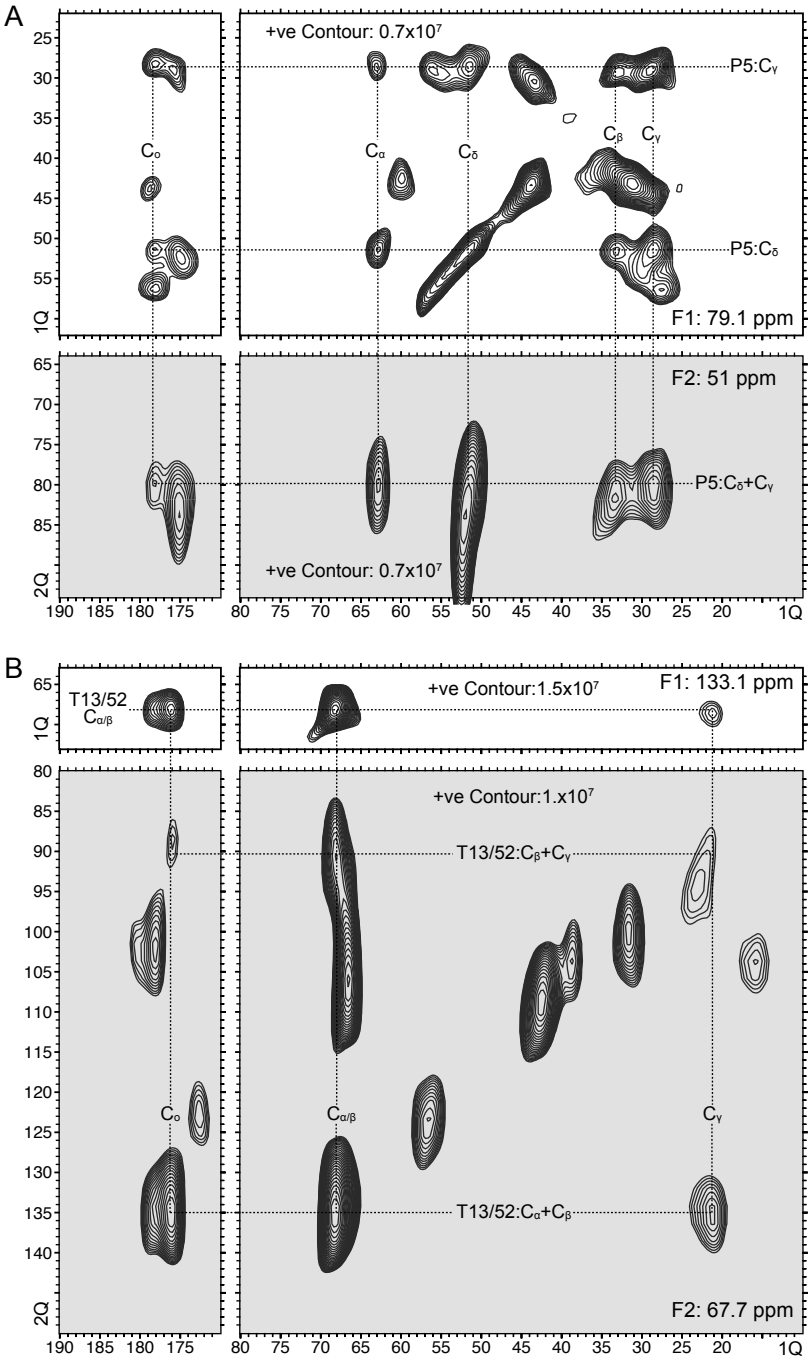


Figure S4: Signals in F_1-F_3 dimension (grey background) correlating with signals in the F_2-F_3 dimension (white background) for strips corresponding to P5, & T13/T52 shown in Figure 9.

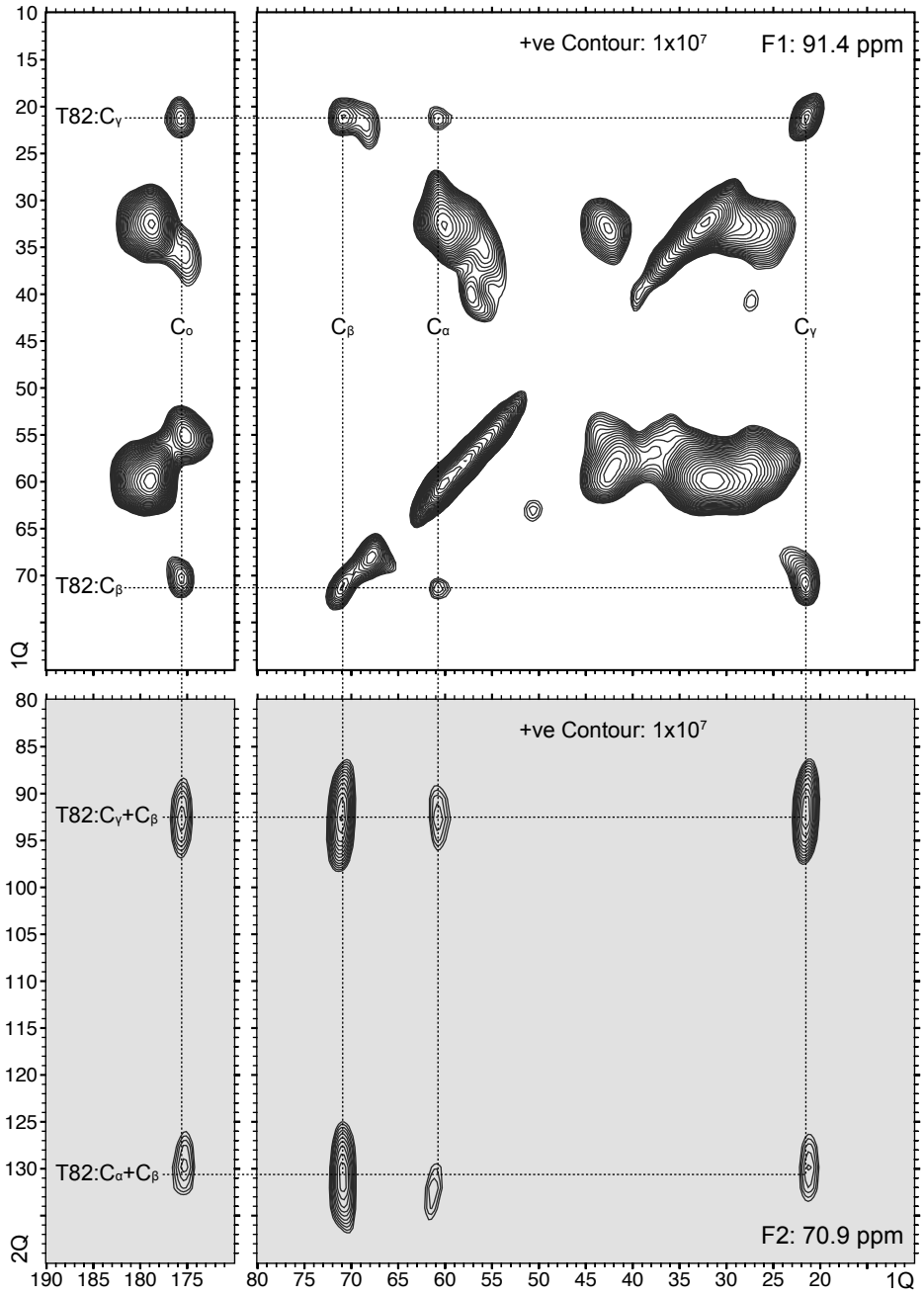


Figure S5: Signals in F_1-F_3 dimension (grey background) correlating with signals in the F_2-F_3 dimension (white background) for strips corresponding to T82 shown in Figure 9.

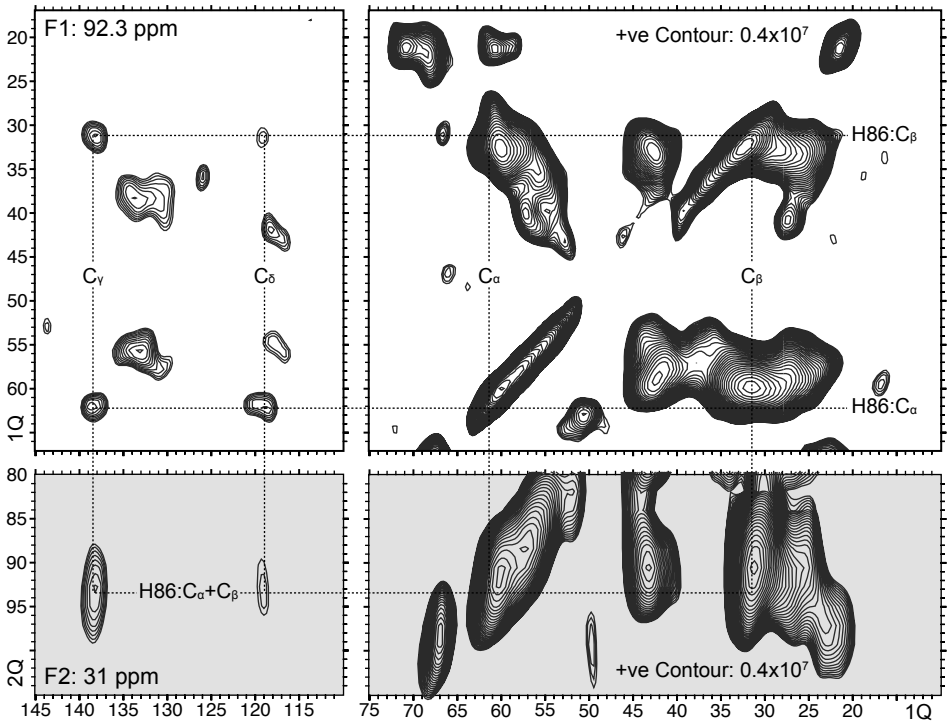


Figure S6: Signals in F_1-F_3 dimension (grey background) correlating with signals in the F_2-F_3 dimension (white background) for strips corresponding to H86 shown in Figure 9.

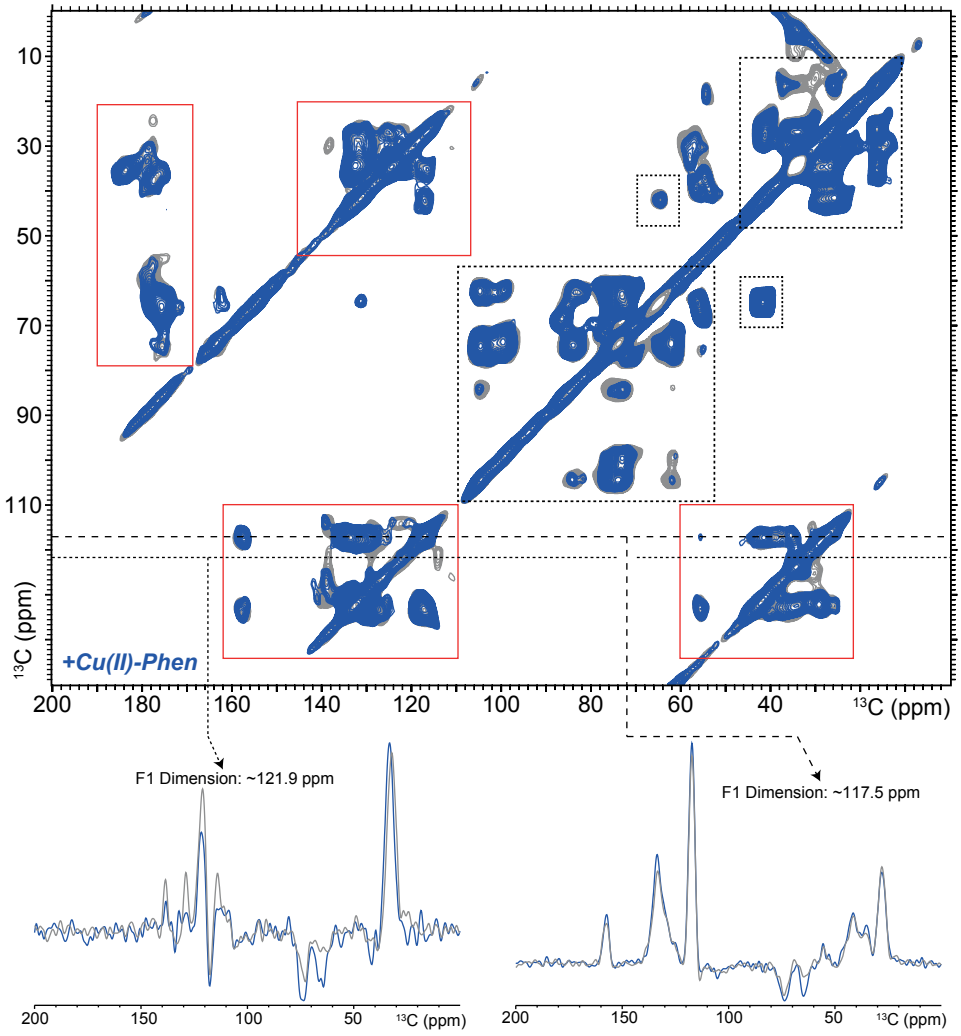


Figure S7: ^{13}C-^{13}C spin diffusion (PDSD) spectra on ^{13}C, ^{15}N labelled LmrR in *E. coli* cells. The aliphatic region was heavily dominated by lipid and sugar resonances (dashed boxes). The aromatic region between 110-160 ppm and the carbonyl region, solely consisted of protein resonances (shown in red boxes). 1D slices from purely Trp aromatic resonances (bottom left) and other aromatic resonances like Phe & Tyr (bottom right) are scaled to show the selective quenching of the former over the latter.

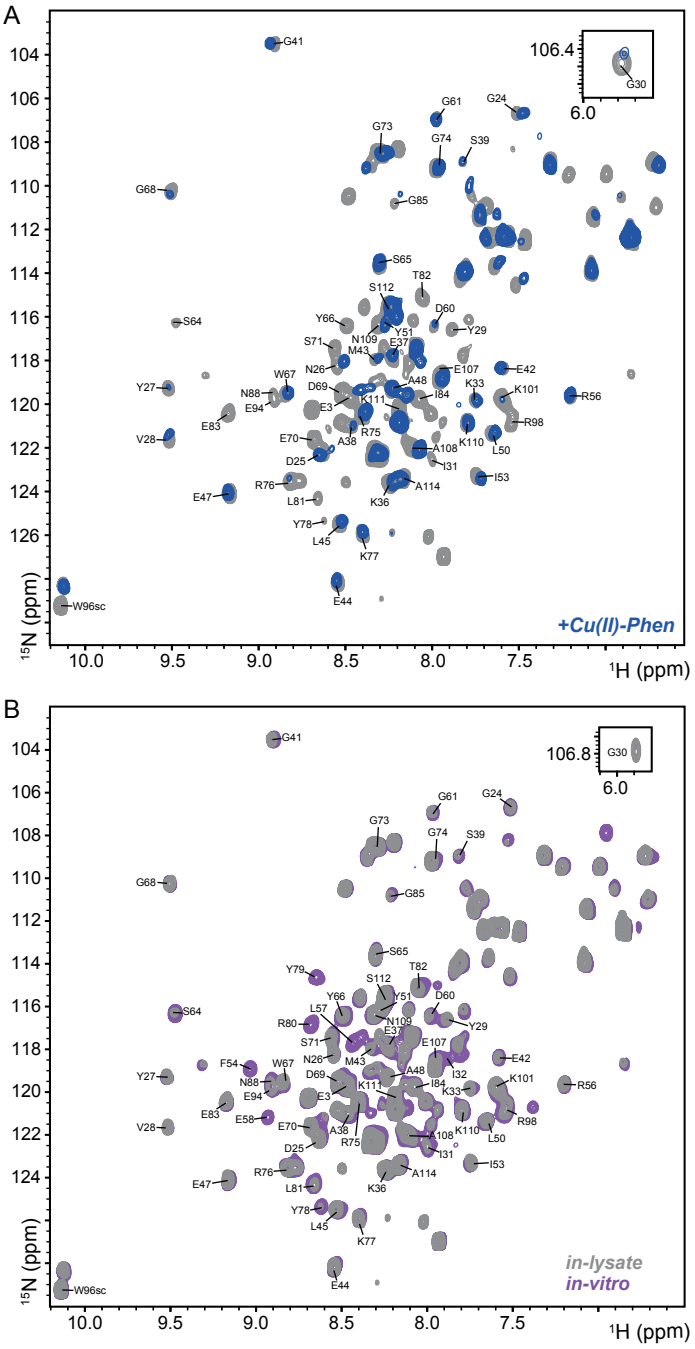


Figure S8: A) ^{15}N- HSQC spectra of cell lysate containing LmrR with and without Cu^{II}-Phen. B) ^{15}N-HSQC spectra of LmrR in-vitro (magenta) vs. in-lysate (grey).

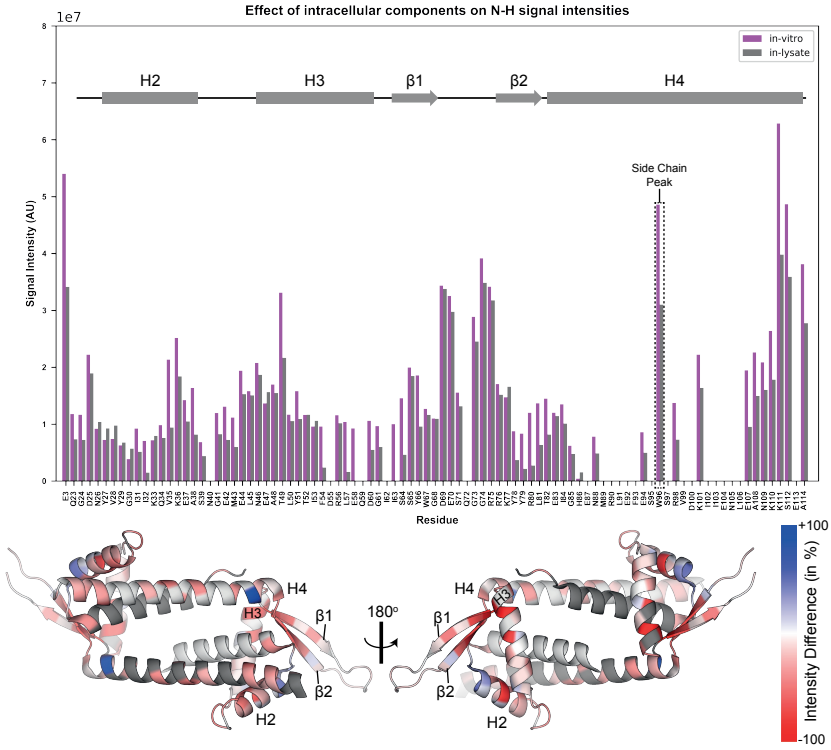


Figure S9: Residue-specific change in NMR signal intensity of the backbone amides in LmrR in cell lysates (in solution-state NMR). Data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). Note that the color scale shown in the figure in the main text is different from here. (See Figure S8B for the 2D spectrum).

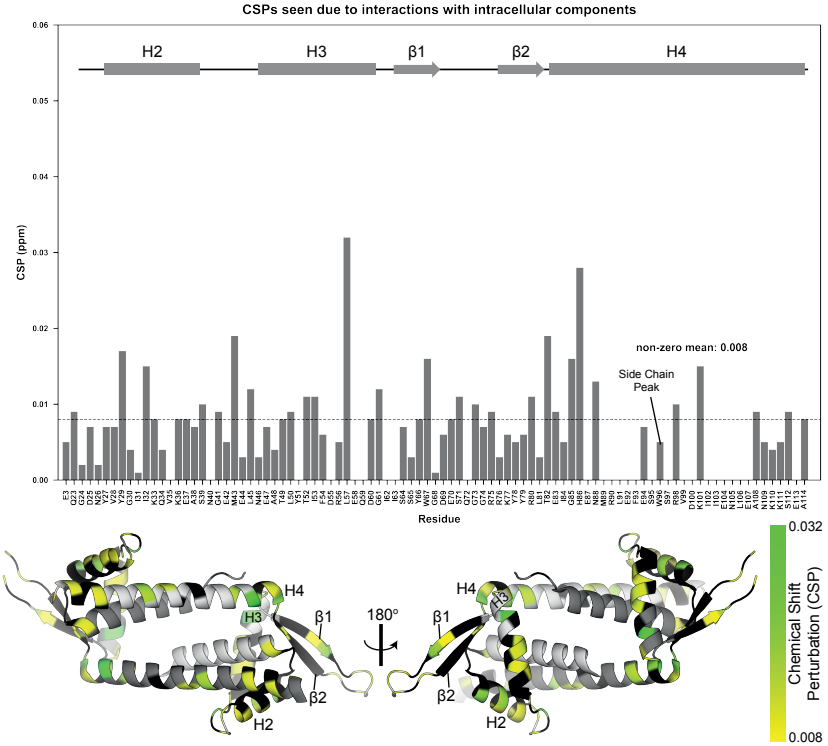


Figure S10: CSPs per residue of LmrR due to weak interactions from cellular components in cell lysates (in solution-state NMR). Residues in black exhibit CSP's of less than 0.08 ppm on the backbone amide, which is the averge CSP. Note that the color scale shown in the figure in the main text is different from here. Data are plotted on the crystal structure of LmrR (PDB ID: 3F8F). See Figure S8B for the 2D spectrum.

References

1. Agustiandari, H., Lubelski, J., van den Berg van Saparoea, H. B., Kuipers, O. P. & Driessen, A. J. M. LmrR Is a Transcriptional Repressor of Expression of the Multidrug ABC Transporter LmrCD in Lactococcus lactis. *J. Bacteriol.* **190**, 759–763 (2008).
2. Madoori, P. K., Agustiandari, H., Driessen, A. J. M. & Thunnissen, A. M. W. H. Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition. *EMBO J.* **28**, 156–166 (2009).
3. Takeuchi, K., Tokunaga, Y., Imai, M., Takahashi, H. & Shimada, I. Dynamic multidrug recognition by multidrug transcriptional repressor LmrR. *Sci. Rep.* **4**, (2014).
4. van der Berg, J. P., Madoori, P. K., Komarudin, A. G., Thunnissen, A.-M. & Driessen, A. J. M. Binding of the Lactococcal Drug Dependent Transcriptional Regulator LmrR to Its Ligands and Responsive Promoter Regions. *PLoS One* **10**, e0135467 (2015).
5. Takeuchi, K., Imai, M. & Shimada, I. Dynamic equilibrium on DNA defines transcriptional regulation of a multidrug binding transcriptional repressor, LmrR. *Sci. Rep.* **7**, (2017).
6. Bos, J., Browne, W. R., Driessen, A. J. M. & Roelfes, G. Supramolecular Assembly of Artificial Metalloenzymes Based on the Dimeric Protein LmrR as Promiscuous Scaffold. *J. Am. Chem. Soc.* **137**, 9796–9799 (2015).
7. Drienovská, I. *et al.* Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid. *Chem. Sci.* **8**, 7228–7235 (2017).
8. Ségaud, N., Drienovská, I., Chen, J., Browne, W. R. & Roelfes, G. Artificial Metalloproteins for Binding and Stabilization of a Semiquinone Radical. *Inorg. Chem.* **56**, 13293–13299 (2017).
9. Villarino, L. *et al.* An Artificial Heme Enzyme for Cyclopropanation Reactions. *Angew. Chemie - Int. Ed.* **57**, 7785–7789 (2018).
10. Roelfes, G. LmrR: A Privileged Scaffold for Artificial Metalloenzymes. *Acc. Chem. Res.* **52**, 545–556 (2019).
11. Bos, J., García-Herraiz, A. & Roelfes, G. An enantioselective artificial metallo-hydratase. *Chem. Sci.* **4**, 3578–3582 (2013).
12. Volentini, S. I., Farías, R. N., Rodríguez-Montelongo, L. & Rapisarda, V. A. Cu(II)-reduction by Escherichia coli cells is dependent on respiratory chain components. *BioMetals* **24**, 827–835 (2011).
13. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *J. Magn. Reson.* **142**, 97–101 (2000).
14. Weingarth, M., Demco, D. E., Bodenhausen, G. & Tekely, P. Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem. Phys. Lett.* **469**, 342–348 (2008).
15. Baldus, M., Petkova, A. T., Herzfeld, J. & Griffin, R. G. Cross polarization in the tilted frame: Assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **95**, 1197–1207 (1998).
16. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
17. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
18. Narasimhan, S. *et al.* Rapid prediction of multi-dimensional NMR data sets using FANDAS. in *Methods in Molecular Biology* **1688**, 111–132 (2018).
19. Medeiros-Silva, J. *et al.* 1 H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angew. Chem. Int. Ed. Engl.* **55**, 13606–13610 (2016).
20. Lecoq, L. *et al.* 100 kHz MAS Proton-Detected NMR Spectroscopy of Hepatitis B Virus Capsids. *Front. Mol. Biosci.* **6**, 58 (2019).

21. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
22. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chemie - Int. Ed.* **57**, 4143–4148 (2018).
23. Lacabanne, D., Meier, B. H. & Böckmann, A. Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *J. Biomol. NMR* **71**, 141–150 (2018).

Chapter 6

Rapid prediction of multi-dimensional NMR data sets using FANDAS

This chapter is adapted from a published chapter in a book:
"Rapid prediction of multi-dimensional NMR data sets using FANDAS"
Protein NMR: Methods in Molecular Biology, Volume 1688, 2018
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 labelling schemes have become a powerful approach to tackle spectral crowding. In fact, selecting the appropriate isotope-labelling 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 labelling 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 (^{1}H) 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.

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[6] (FANDAS) 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 accommodate a variety of inputs and input modifications to accurately predict the peak pattern for given NMR experiments (Figure 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 labelling 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 labelling schemes by offering improved choices for forward labelling (see, ref.[10] for a further discussion of forward and reverse isotope labelling) such as specific ^{13}C-only or ^{15}N-only labelling which were not available before. These new options speed up the process of predicting NMR correlation data for different forward labelling 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 which 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 which can be found at 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.

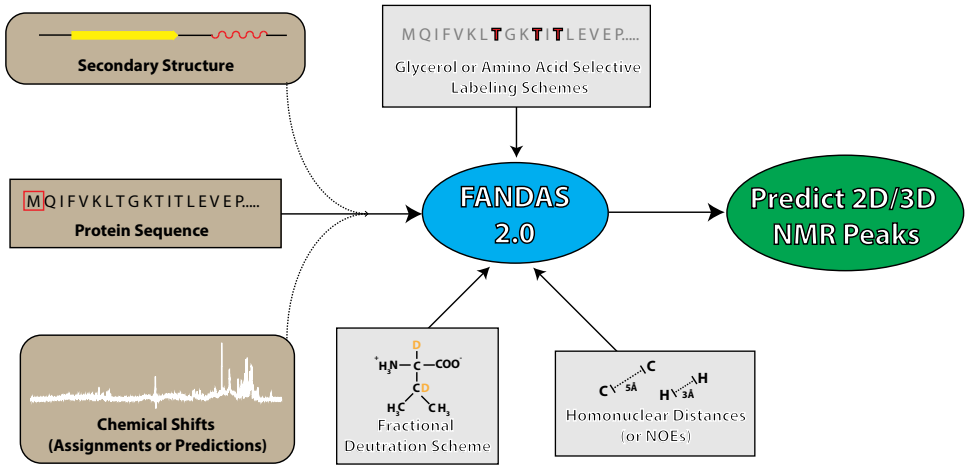


Figure 1: Overall scheme of FANDAS, describing the different types of data used to refine the final prediction of peaks for multi-dimensional NMR experiments. The different types of inputs that FANDAS accepts are given in brown boxes and input modification options such as labelling schemes and distances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

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.

Materials and Prior Knowledge

Introduction to SPARKY peak lists

FANDAS has been geared towards 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 which appears as an annotation on the spectrum, B) the peak coordinates which represent chemical shifts of the correlating nuclei and C) the notes section, where the user is free to add any notes regarding the peak (Figure 2). The output of FANDAS is composed of peak files that the user can then superimpose on the spectra.

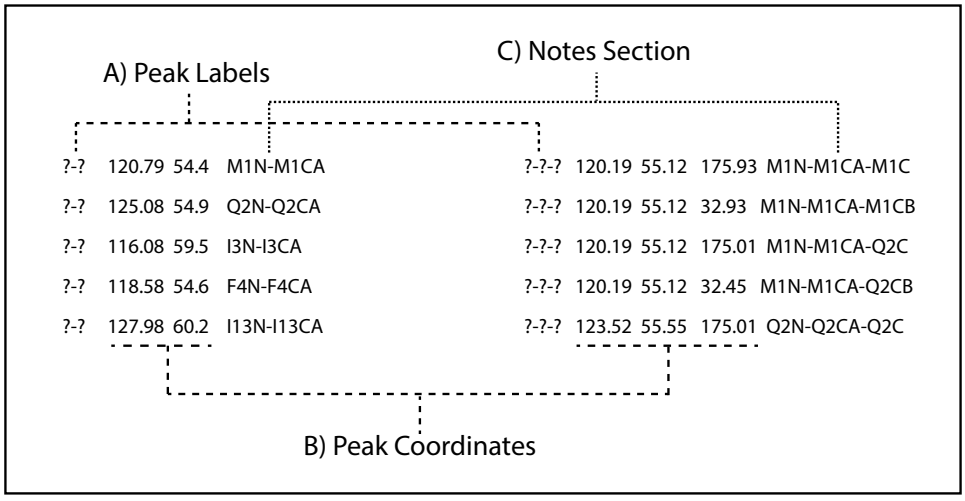


Figure 2: Description of SPARKY 2D (left) and 3D (right) peak lists generated using FANDAS. A) The peak labels, which would appear on the spectrum when the peak list is loaded. The default "? -?" for 2D peak list and "?-?-?" for 3D peak list imply that the peaks do not have any labels. B) Peak coordinates, 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.

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

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 which 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 on to 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

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.

6.1 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:

i. *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.

ii. *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.

iii. *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 Figure 3, the secondary structure assignments were done using command line version of STRIDE on the PDB structure 1UBQ and 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 "APAPMLQSMVSLLQSLV" 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 Figure 3.

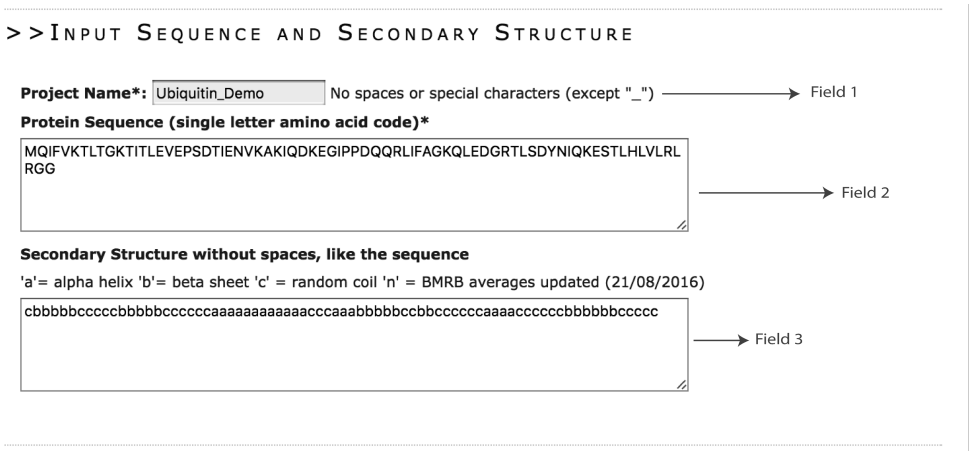


Figure 3: Example input of the "Input Sequence and Secondary Structure" block.

2. Input Chemical Shifts as BMRB Table Block: FANDAS also accepts user defined chemical shift assignments which 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:

i. *BMRB Tables:* It is recommended for the assignments to be in a tabular NMR-STAR format (for a detailed description, see: http://www.bmrb.wisc.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 (Figure 4). In the example (Figure 5), the command line version of SHIFTX2 was used to generate the BMRB tables using the crystal structure of Ubiquitin (1UBQ).

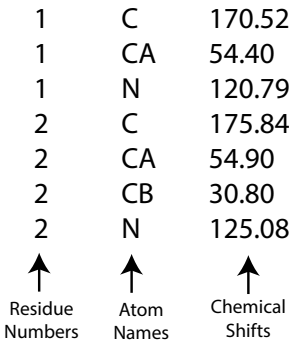


Figure 4: An example of BMRB tables that is accepted in FANDAS.

ii. 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.

iii. 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.

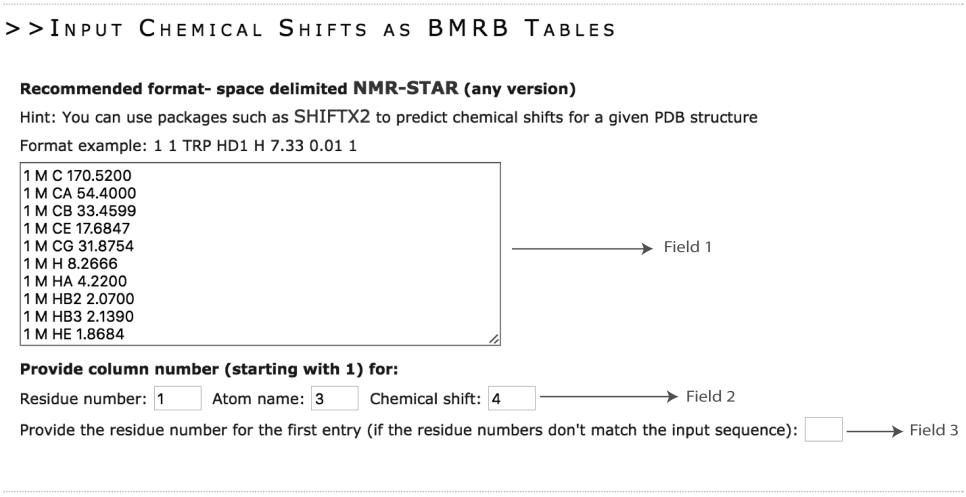


Figure 5: Filled out sample of "Input Chemical Shifts as BMRB Tables" block.

3. Amino-Acid Selective Labelling Schemes Block: By default, it is assumed that the protein of interest is fully isotope labelled (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 labelling schemes (Figure 6). If forward or reverse labelling schemes are used, the user must specify a list of amino acids for the chosen labelling scheme. This section is particularly useful for assessing the result of different labelling schemes on the resulting ssNMR spectra. A typical application for features in this section is discussed later in this chapter.

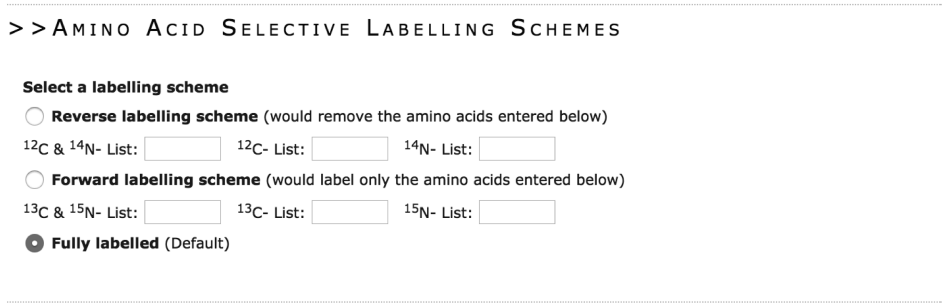


Figure 6: "Amino-Acid Selective Labelling Schemes" block consists of options to choose the appropriate labelling scheme and a provision to enter the desired labelled amino acids in double or single labelled forms.

4. Glycerol Labelling Schemes and Fractional Deuteration: To reduce spectral crowding among the various carbon positions, glycerol-based labelling schemes have been introduced[21–23]. Such schemes have been implemented in FANDAS and choosing any of the glycerol labelling schemes would disable amino acid selective (forward or reverse) labelling schemes or fully labelled schemes. Additionally, this section is set up to include fractional deuteration (dashed box in Figure 7) scheme, which is intended to fully deuterate $^{13}C_{\alpha}$ nuclei as well as specific side chain carbons[13,24].

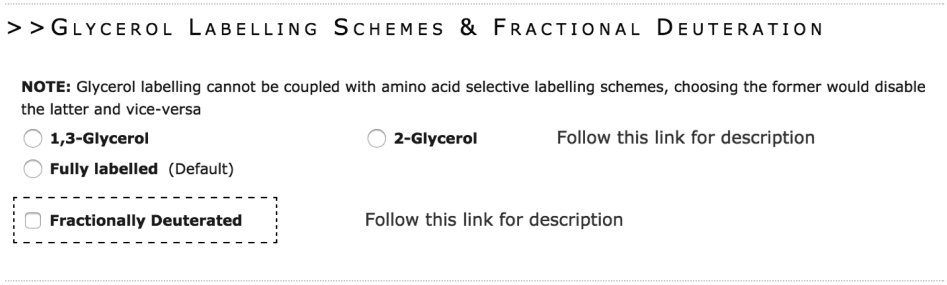


Figure 7: Highlighting how different labelling schemes are handled in FANDAS and the option for fractional deuteration (dashed box).

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 cut-off distance (in Å). To exemplify such a distance cut-off, a sample output for a distance edited CC Spin diffusion experiment is shown for two cut off values in (Figure 8).

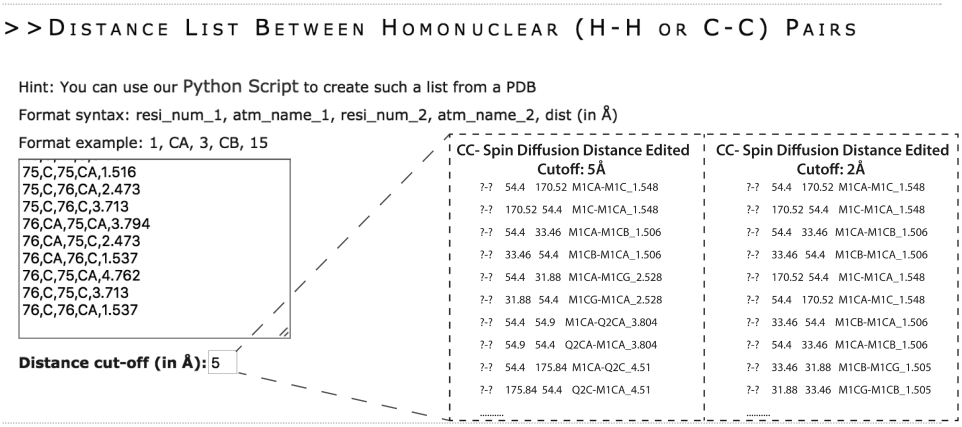


Figure 8: Example outputs (dashed boxes) suited for the analysis of distance-dependent CC correlation experiments (such as protein-driven spin diffusion) for two different cut off values. Distance lists were calculated using the crystal structure of Ubiquitin (PDB: 1UBQ).

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 which the user can simply select using the appropriate check boxes (Figure 9).

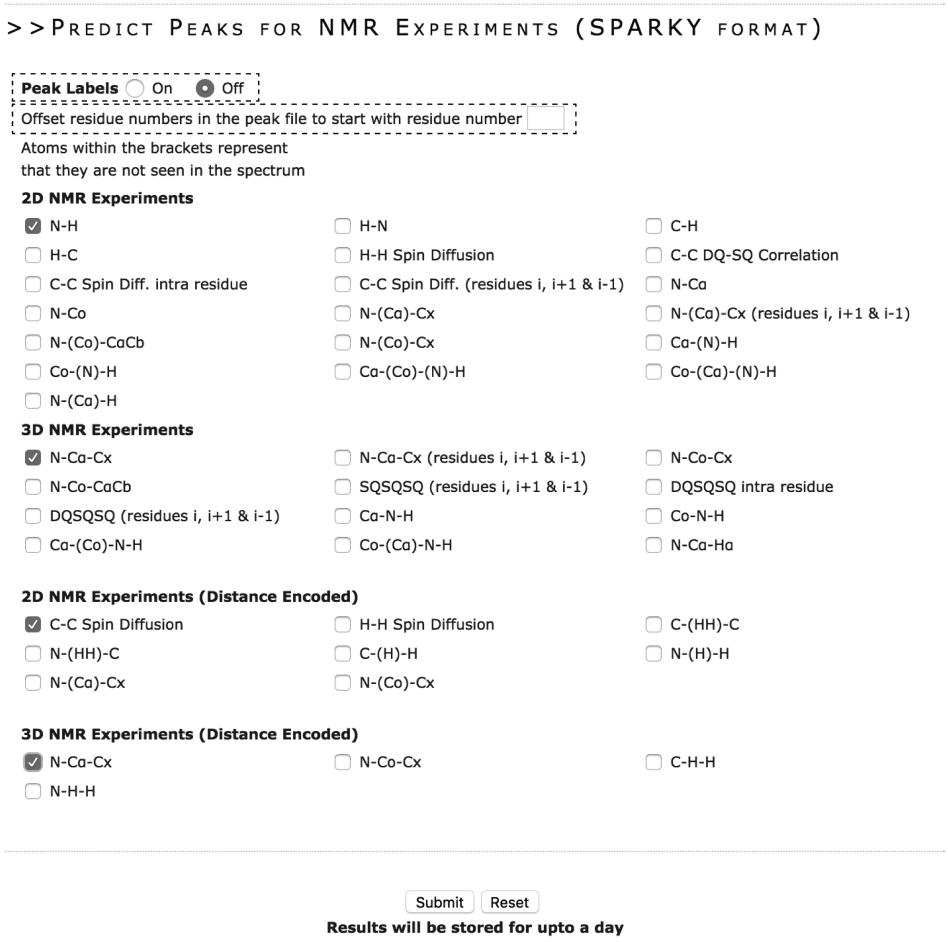


Figure 9: The final input block, where the user can select experiments for which predictions are to be made, specify if peak labels are required and if the residue numbers need to be offset (highlighted using dashed boxes).

i. Peak Labels: Along with the list of experiments, this section features an option to turn on peak labels (highlighted with a dashed box in Figure 9). As described in Section 6.1, 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 (Figure 10).

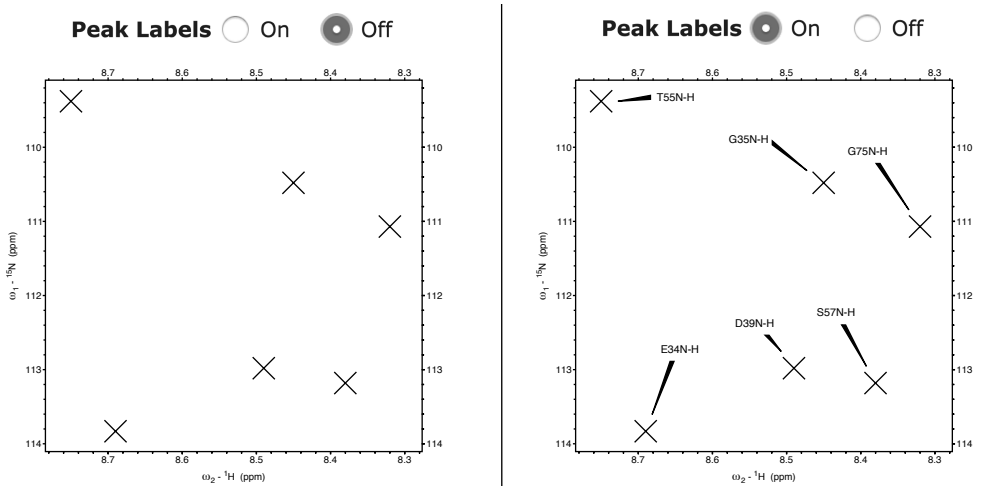


Figure 10: Sample peak outputs overlaid on empty N-H Spectra in SPARKY. The left panel shows the default output when Peak Labels are inactive and the right panel shows the output when Peak Labels are turned on.

ii. Residue number offset: This field (highlighted with a dashed box in Figure 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 1, 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, where in the starting residue number may not be 1.

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 (Figure 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.

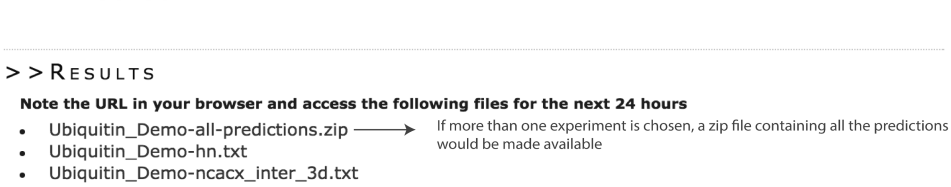


Figure 11: Prediction results as seen on the results page.

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 on to 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 on to 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, 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:

i. distance_calculator.py: This script is used to calculate the distance between homonuclear pairs from the supplied PDB coordinates.
ii. fandas.py: This is the FANDAS_2.0 application.
iii. README.md: A github mandated readme file for the program.
iv. standard.dat: A file containing the standard chemical-shift tables for all amino acids.
v. test_dataset.tar: Contains a test dataset which has been used to demonstrate the use of FANDAS in this manual. It contains a readme file named "readme. txt" that describes all 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 are 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 which are not listed within square brackets are the obligatory arguments, which in our case is 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,
                        gl13 = 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 angstorms 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 angstorms, default: 5 Angstorm
  -o O                  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 1ubq-seq.txt -ss 1ubq-ss.txt -bt 1ubq-bmrb-tables.txt
-btc 1 3 4 -dlist 1ubq-dist.txt -dlim 5 -exp_2d cc_spindiff_intra -exp_2dd
chh -exp_3d ncacx -exp_3dd ncacx -sl
```

Table 1: Description of all parameters in a sample FANDAS input

Flag + Argument(s)	Description
-i 1ubq-seq.txt	input file (or path) is "**1ubq-seq.txt**"
-ss 1ubq-ss.txt	file (or path) containing secondary structure is "**1ubq-ss.txt**"
-bt 1ubq-bmrb-tables.txt	file (or path) containing the BMRB tables is "**1ubq-bmrb-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 1ubq-dist.txt	file (or path) containing the distance list is "**1ubq-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	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

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

Table 2: Alternative and supplementary labelling schemes to the default- uniformly ^{13}C & ^{15}N default labelling 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 (`-ls`)
`-cl <residue list>`	list of forward ^{13}C labeled or reverse ^{12}C (unlabeled) amino acids as per the chosen labeling scheme (`-ls`)
`-nl <residue list>`	list of forward ^{15}N labeled or reverse ^{14}N (unlabeled) amino acids as per the chosen labeling scheme (`-ls`)
`-fd`	include fractional deuteration; this can be combined with other labeling schemes

3. 3 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 Figure 12. 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.

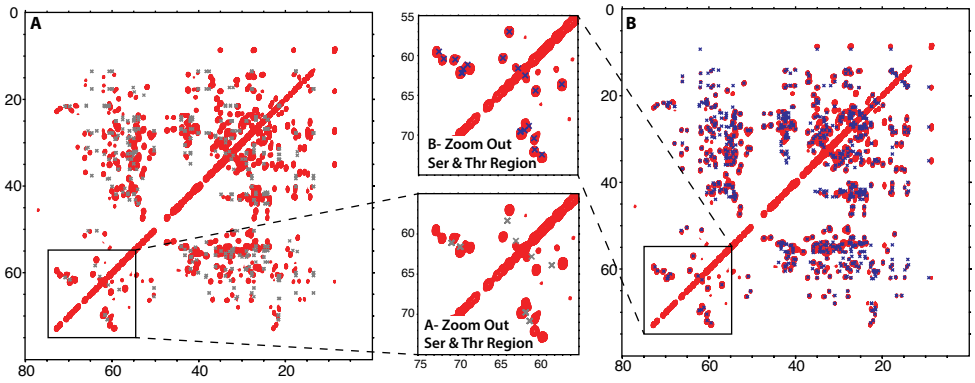


Figure 12: FANDAS Predictions of CC (PARIS) experiment[25] when only secondary structures are supplied (Panel A with the grey peaks) and in an information rich regime where chemical shifts of all atoms were predicted from the crystal structure using SHIFTX2 (Panel B with blue peaks). In the zoom in of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

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

Case Study- Choosing amino acid selective labelling 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 labelling 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 labelling schemes. To illustrate the output generated for the different labelling schemes, two 2D experiments have been selected:

1. N-Ca intra-residue correlations which probe protein residues that are both ^{13}C and ^{15}N labelled.

2. N-Co inter-residue correlations which probe the polarization transfer between the backbone nitrogen atom of the $(i)^{th}$ residue to the carboxyl carbon atom of the previous $(i-1)^{th}$ residue. This requires the $(i)^{th}$ residue to be at least ^{15}N labelled and the $(i-1)^{th}$ residue to be at least ^{13}C labelled.

Case 1- Fully Labelled (Default): The default option (Figure 13 A), assumes that all the residues are fully labelled. The output for such a labelling scheme would contain all peaks that could possibly occur in each experiment (Figure 13 B)

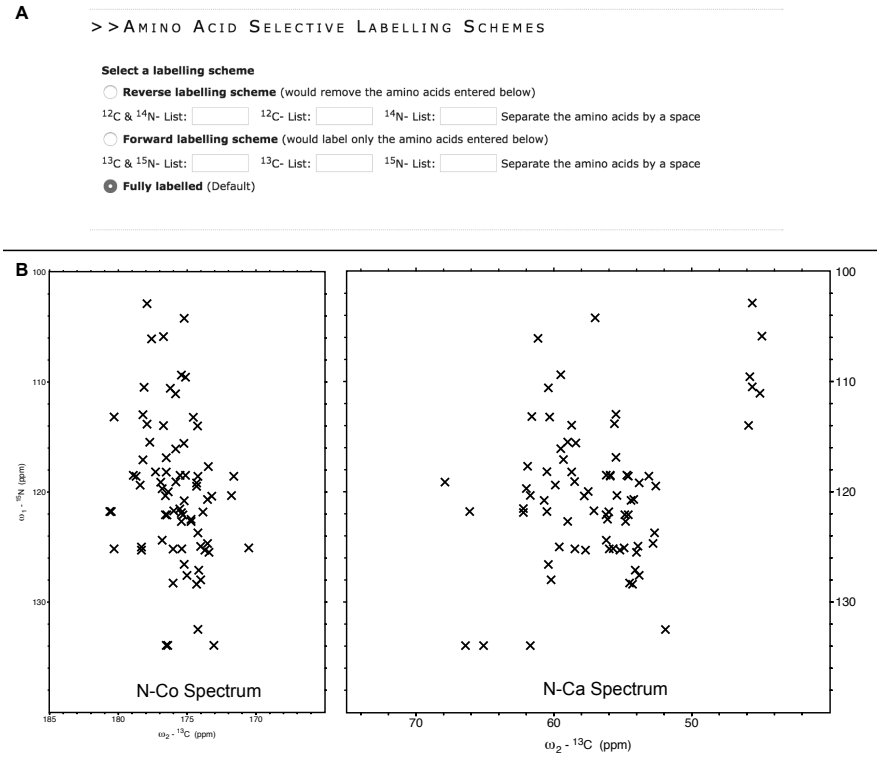


Figure 13: The default fully labelled option and the resulting predictions for peaks in the N-Co and N-Ca spectra of Ubiquitin.

Case 2: ^{13}C & ^{15}N labelled: If specific residue types are ^{13}C & ^{15}N labelled, FANDAS retains the ^{13}C & ^{15}N (forward labelled Figure 14 A) chemical shifts or removes them and retains the remaining amino acids (reverse labelled Figure 14 B). When a forward labelling 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 Figure 14 C.

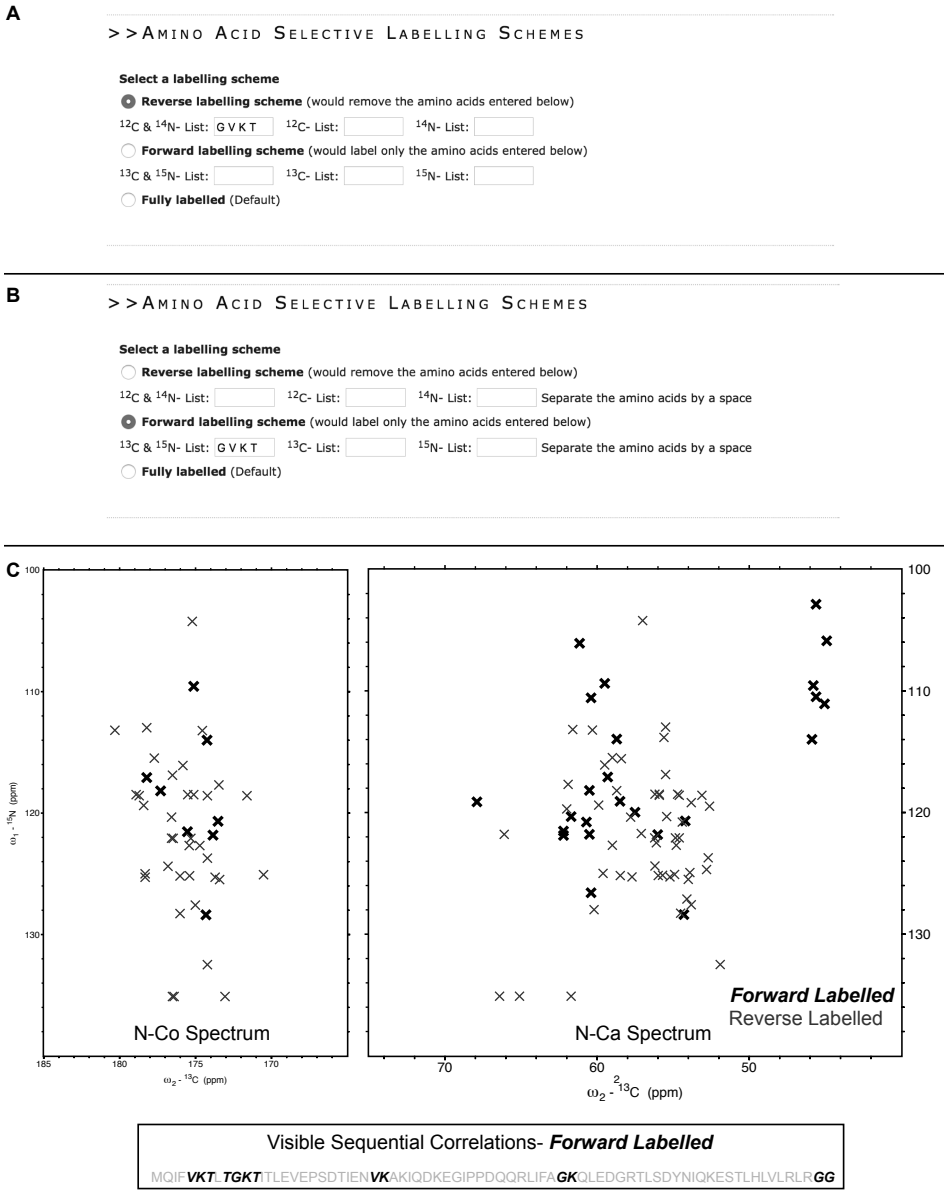


Figure 14: Incorporating amino acid selective ^{13}C + ^{15}N labelling scheme in FANDAS for glycine, valine, lysine and threonine residues. As it can be seen in the peak prediction, the two labelling schemes complement each other.

Case 3: Using ^{13}C only labelled amino acids and ^{15}N only labelled amino acids: There exist labelling schemes where either ^{13}C or ^{15}N amino acids are labelled in combination to probe site selective sequential correlations on proteins (see,

references[3,4]). FANDAS treats inputs for this labelling scheme in a similar fashion to the previous case. If residues are ^{13}C or ^{15}N labelled, FANDAS either retains only the ^{13}C or ^{15}N (forward labelled Figure 15 A) chemical shifts or removes them and retains the remaining amino acids (reverse labelled Figure 15 B) as shown in Figure 15 C.

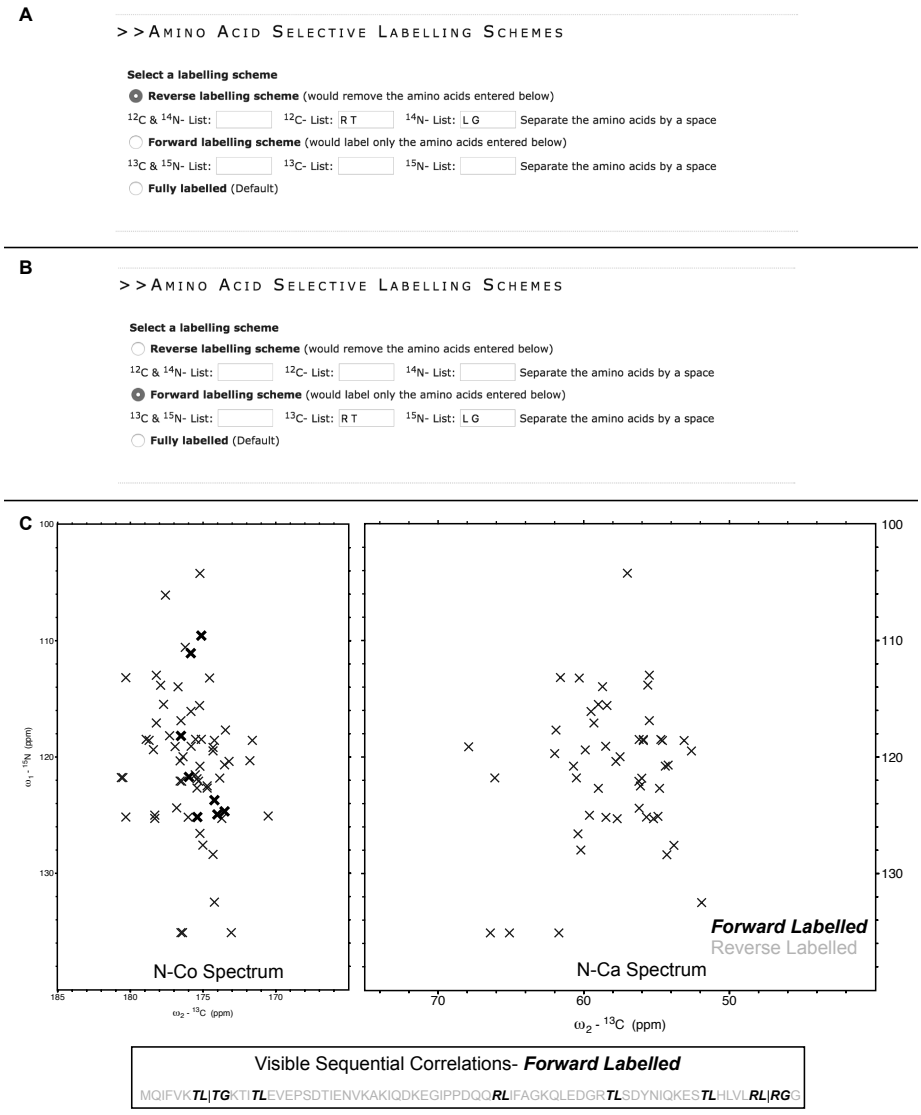


Figure 15: Incorporating ^{13}C labelled arginine, threonine and 15N labelled leucine, glycine amino acid selective labelling scheme in FANDAS. As it can be seen in the peak prediction, sequential correlations can be observed for specific parts of the protein.

References

1. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
2. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4863–8 (2012).
3. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12**, 5–9 (2015).
4. Kaplan, M. *et al.* EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **167**, 1241-1251.e11 (2016).
5. Kaplan, M., Pinto, C., Houben, K. & Baldus, M. Nuclear magnetic resonance (NMR) applied to membrane–protein complexes. *Q. Rev. Biophys.* **49**, e15 (2016).
6. Gradmann, S. *et al.* Rapid prediction of multi-dimensional NMR data sets. *J. Biomol. NMR* **54**, 377–87 (2012).
7. Sinnige, T. *et al.* Solid-State NMR Studies of Full-Length BamA in Lipid Bilayers Suggest Limited Overall POTRA Mobility. *J. Mol. Biol.* **426**, 2009–2021 (2014).
8. Sinnige, T. *et al.* Insight into the conformational stability of membrane-embedded BamA using a combined solution and solid-state NMR approach. *J. Biomol. NMR* **61**, 321–332 (2015).
9. Baker, L. A., Daniëls, M., van der Cruijsen, E. A. W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62**, 199–208 (2015).
10. Renault, M., Cukkemane, A. & Baldus, M. Solid-state NMR spectroscopy on complex biomolecules. *Angew. Chem. Int. Ed. Engl.* **49**, 8346–57 (2010).
11. Pauli, J., Baldus, M., van Rossum, B., de Groot, H. & Oschkinat, H. Backbone and side-chain 13C and 15N signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 Tesla. *Chembiochem* **2**, 272–81 (2001).
12. Sinnige, T., Daniëls, M., Baldus, M. & Weingarth, M. Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J. Am. Chem. Soc.* **136**, 4452–5 (2014).
13. Mance, D. *et al.* An Efficient Labelling Approach to Harness Backbone and Side-Chain Protons in (1) H-Detected Solid-State NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **54**, 15799–15803 (2015).
14. T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco.
15. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11**, 852–61 (2002).
16. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411-9 (2011).
17. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–79 (1995).
18. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389-94 (2015).
19. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
20. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
21. David M. LeMaster, A. & Kushlan, D. M. Dynamical Mapping of E. coli Thioredoxin via 13C NMR Relaxation Analysis. *J. Am. Chem. Soc.* **118**, 9255–9264 (1996).

22. Hong, M. & Jakes, K. Selective and extensive 13C labeling of a membrane protein for solid-state NMR investigations. *J. Biomol. NMR* **14**, 71–4 (1999).
23. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98–102 (2002).
24. Nand, D., Cukkemane, A., Becker, S. & Baldus, M. Fractional deuteration applied to biomolecular solid-state NMR spectroscopy. *J. Biomol. NMR* **52**, 91–101 (2012).
25. Weingarth, M., Demco, D. E., Bodenhausen, G. & Tekely, P. Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem. Phys. Lett.* **469**, 342–348 (2008).

Acknowledgement

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.

Chapter 7

General Discussion and outlook: The present and future of in-cell DNP-ssNMR

Parts of this chapter has been adapted and used for the following review article:

"*When small becomes too big: Expanding the use of in-cell solid-state NMR spectroscopy*"

Siddarth Narasimhan, Gert E. Folkers, Marc Baldus

DOI: 10.1002/cplu.202000167 ChemPlusChem. *in press*

Summary

In the recent years, there have been no unique protein structure topologies deposited in the PDB, which may reflect the finiteness of unique protein structures that exist in nature. Moreover, traditional structural studies have neglected the cellular context of the biomolecules, as there were no techniques that enabled atomistic studies in a highly heterogeneous cellular environment. Developments in biomolecular NMR and cryo-EM techniques have fueled the conception and the growth of cellular structural biology as a field. While cryo-EM focusses on very large molecular machines and protein assemblies, NMR has tackled small and heterogenous proteins. The two techniques thus complement each other. An analogous complementarity also exists within NMR, where cellular studies using solution-state NMR has been limited to small and soluble proteins while solid-state NMR has enabled studies on rather large insoluble proteins. The motivation behind pursuing the research presented in this thesis is to design and implement solid-state NMR approaches which would enable the study of soluble and promiscuous proteins within the cellular milieu. Despite their soluble nature, promiscuous proteins have remained out of the reach for solution-state NMR as they interact with their binding partners leading to reduced tumbling in crowded cellular environments. We have developed and tested an innovative DNP-ssNMR approach which is compatible with both mammalian and bacterial cells.

Chapter 2 describes a scheme to prepare DNP-ssNMR compatible samples of wild-type Ubiquitin (Ub) in HeLa cells. We first delivered isotope labelled Ub into unlabelled HeLa cells and then introduced DNP agents into the cells by means of rapid diffusion, immediately prior to our ssNMR experiments. We then performed functional assays to ensure that the exogenously delivered protein is functional in the cells and confirmed that the cells take up the radicals. Solution NMR experiments were recorded to check for protein degradation. Finally, we confirmed that the protein is folded properly in cells by recording and analyzing 2D and 3D DNP-ssNMR spectra.

Despite the small size, Ub is known to be involved in many regulatory functions in the cells. One of the major roles of Ub, is to promote proteasomal degradation of a variety of proteins. Proteasome inhibition has been previously shown to exert immense stress on the cell and leads to the migration of nuclear Ub into the cytoplasm to conjugate to proteins that need to be degraded. In **chapter 3**, we explored the effect of proteasome inhibition on various Ub residues in cells by analyzing the 2D DNP-ssNMR spectra. We first studied the uniformly ^{13}C-, ^{15}N-labelled Ub, where a large-scale disappearance of signals from residues throughout

the protein was observed upon proteasomal inhibition. Next, we focused on Lysine (K) residues, particularly K48, by using residue specific-labelled Ub in cells. We observed a small increase in the isopeptide linked K48 resonances along with the appearance of new, albeit weak, NMR signals supporting the idea of a thus far uncharacterized post-translational modification (PTM). Finally, we obtained evidence for conformational heterogeneity of the K48 backbone in proteasome inhibited cells. Taken together, we conclude that Ub molecules become much more heterogeneous in the cells both at the structural as well as chemical level upon proteasomal inhibition. Further NMR experiments and biochemical tests need to be performed to confirm our interpretation of the spectra.

In **Chapter 4**, we extended our in-cell DNP supported ssNMR approach to soluble bacterial proteins. As a test case, we resorted to labelling Ub in the *E. coli* BL21 strains: R2 and Lemo. Since we could not deliver exogenous proteins into the cells, we achieved targeted labelling of Ub in the cells using the rifampicin approach which was previously demonstrated on membrane proteins. We were confronted with a higher degree of background labelling in cells, in comparison to membrane protein preparations where cell envelopes were used. We mitigated the background labelling problem and achieved complete (>90%) deuteration of cellular background to obtain maximal DNP enhancements on Ub. In the process, we discovered the incompatibility of R2 strains for our DNP experiments, as we obtained very poor DNP enhancements when we tried to obliterate the background signals by using algal amino acid mixture to label Ub specifically.

Artificial metalloenzymes catalyze many new-to-nature reactions. In the recent years, performing catalysis in cells without requiring any purification, is touted to have several advantages including lower operational costs. LmrR, a *Lactococcus lactis* derived protein can assemble into a metalloenzyme by accommodating a Cu^{II} phenanthroline group in its dimeric interface by means of π-stacking interactions with Tryptophan side-chain indoles. In **chapter 5**, we selectively labelled LmrR in *E. coli* and confirmed the proper fold of the protein using 2D and 3D ssNMR experiments. Next, we observed specific (paramagnetic) quenching of signals from Trp aromatics in LmrR and a concomitant decrease in the T_1 relaxation times, confirming the assembly of the metalloenzyme. Solution-state NMR studies on the lysate further revealed the effect of the cellular components on the dynamics of LmrR. This information may be useful in the future to redesign LmrR with a superior catalytic activity in cells.

In **chapter 6**, we describe FANDAS 2.0, an in-house program & webserver that we developed on a Python platform. The application can be used to predict multidimensional NMR spectra of proteins for easy visualization and analysis.

We provide step-by-step instructions to use the program and describe its several features. It has potential applications in choosing different amino acid labelling schemes to obtain resolved spectra in complex cases such as in-cell NMR which is generally prone to low resolution spectra.

General Discussions

Since molecules must tumble sufficiently fast in cells, the application of solution-state NMR has been restricted to small and relatively inert proteins. Exploiting the rotational motion in the methyl side-chains through Methyl- selective Transverse Relaxation Optimized Spectroscopy (TROSY) has allowed studies on large proteins in cells[1]. However, this would limit our analysis to methyl containing amino acids that engage in similar types of interactions with intracellular components. The size limitation does not apply to solid-state NMR as extensively demonstrated in the study of membrane proteins in native lipids[2–6]. In the work presented in the preceding chapters, we show that extending solid-state NMR to study soluble proteins in cells can expand the scope of in-cell NMR techniques. As we have shown on HeLa and *E. coli* cells, our DNP-ssNMR approach is compatible with both mammalian and bacterial systems.

In chapter 3, these conceptual advances enabled us to explore and characterize the residue-specific changes in Ub that entail proteasomal inhibition. Using advanced labelling strategies, we could zoom into the Lysine residues and K48. We obtained experimental evidence of an increase in poly-Ub chain formation, preliminary insight into so far uncharacterized PTM and finally backbone conformational changes in K48 upon proteasomal inhibition.

In the case of bacterial cells, we could confirm the correct assembly of an artificial metalloenzyme that can catalyze Friedel-crafts alkylation of indoles. Here, the combination of solution-state and solid-state NMR studies enabled us to understand the implications of the cellular environment on the dynamics of the protein (LmrR) that assembles into the metalloenzyme. These insights open novel possibilities for a rational redesign of the protein by directed evolution to optimize catalytic activity in cells. This study also highlights the truly complementary nature of solution and solid-state NMR for studying difficult proteins in complex cellular environments.

To perform the aforementioned studies, we had to overcame two principal hurdles, i.e., to obtain molecule specific isotope labelling in cells[3,7,8] and to determine ideal parameters for implementing in-cell DNP. Since DNP inherently improves the sensitivity, it was important to ensure that the extent of background labelling was as minimal as possible. The parameters that are important to obtain molecule specific

labelling are discussed extensively in the previous chapters. Despite using optimal conditions for sample preparations, we still face challenges to further improve signal enhancement by DNP enhancements as well as the spectral resolution of our NMR data. In the following sections we discuss some considerations that are relevant in this context. We examine the DNP buffer composition, outline future advancements in DNP radical developments and we finally shed some thoughts on improving the spectral resolution of DNP-ssNMR spectra in future cellular studies.

7.1 The role of ^{13}C- depleted deuterated glycerol

We conducted our initial studies in HeLa cells, where we achieved molecule specific isotope labelling in the cells by adapting a pre-established approach of delivering isotope labelled proteins into unlabelled cells. The biggest challenge thereafter was to introduce DNP biradicals that retain their capability to enhance NMR signals in the presence of a reducing environment inside cells. Confocal microscopy on HeLa cells (Figure 2, Chapter 2) revealed that the radicals diffuse sufficiently fast into cells within the timeframe in which the rotors are filled and frozen (<15 mins). However, we learnt that the bulk of the DNP-active radicals are still located on the outside of the cells. This notion was followed from our observation that despite using fully deuterated ^{2}H glycerol, natural abundance ^{13}C signals of glycerol were more strongly enhanced in ^{1}H-^{13}C CP experiments than the protein of interest that is located inside the cells (Figure 1A).

The high enhancements on glycerol can in principle be reduced by washing the cells with DNP-juice (that contains glycerol, D_2O and H_2O in 60:20:10 ratio) devoid of the radicals, to ensure that the radicals on the outside of the cells are gradually removed. However, such a procedure would increase the time frame in which radicals in the cells could be quenched[9,10] and perhaps diffuse out of the cells. Therefore, we attempted to prepare our DNP samples in a buffer devoid of 60% glycerol (see Chapter 3, Section 1). The result was that the DNP enhancements were reduced by a factor 2 in the cellular samples which were devoid of glycerol (Figure 1B). The role of 60% glycerol is to enable a glass transition at low temperatures which is required to obtain high and consistent DNP enhancements[11,12]. Since the low enhancements in glycerol free samples prevented us from performing 3D experiments, we finally resorted to using a ^{2}H enriched and ^{13}C depleted variant of glycerol. This led to maximal enhancements on the proteins without any interfering background signals stemming from glycerol (Figure 1C).

In summary, the optimal DNP juice for cellular applications in our hands contains ^{2}H enriched and ^{13}C depleted glycerol to achieve maximum DNP enhancements on the protein of interest and minimizes unwanted background

signals from the buffer. Additionally, the DNP buffer must also contain appropriate salts (See Materials and Methods- Chapters 2-5) to prevent osmotic stress on the cells that could cause cell rupture during the sample filling stage.

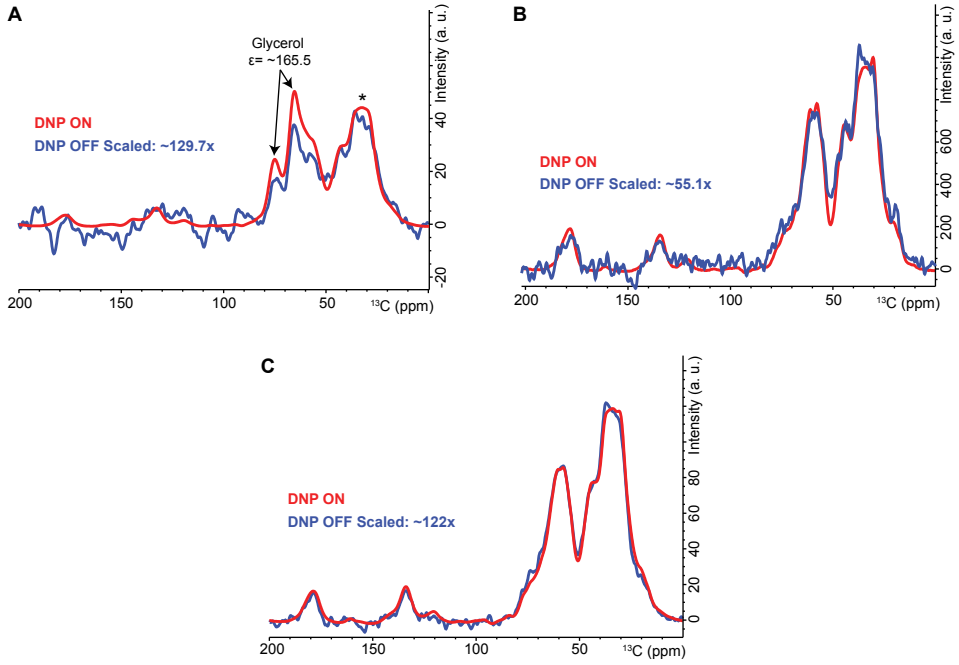


Figure 1: DNP enhancements seen on three different DNP buffers with HeLa cells containing uniformly ^{13}C-, ^{15}N- labelled Ub. A) Deuterated DNP juice (60% ^{2}H- glycerol) where the glycerol peaks are very pronounced and are enhanced to a larger degree than the proteins. B) DNP buffer devoid of glycerol, where enhancement is reduced greatly. C) Optimal DNP juice prepared using deuterated and ^{13}C deprived glycerol.

7.2 Sterically shielded and taggable DNP radicals

The advent of a new generation of water soluble DNP biradicals such as AMUPol and PyPol[13] has enabled a variety of biological applications for DNP-ssNMR[4,14–18]. However, these radicals were primarily developed and tested in the context of in-vitro biological applications which was of critical importance to expand DNP beyond studying materials soluble in organic solvents. In this thesis, we have shown the potential of DNP-ssNMR to enable molecular studies in much more complex cellular environments using protein concentrations that may approach endogenous levels. In spite of our achievements in optimizing the DNP performance by ascertaining the right buffer conditions and by using deuteration to reduce NMR signals of the cell background, there is clearly room for further improvements of the

current protocols.

The short-lived nature of the radicals in cells at room temperature is of a great hindrance to improving DNP enhancements. Ideally for in-cell DNP-NMR applications, it would be useful to prevent the reduction of the radicals by shielding them from reducing agents. Such radicals have been developed for use in in-cell electron paramagnetic resonance[19–21] (EPR) and they are in a preliminary stage of development for in-cell NMR applications[17]. Development of such sterically shielded/caged radicals could potentially allow us to incubate the cells in the DNP juice for longer times resulting in a higher concentration of radicals inside cells.

Additionally, such novel DNP radicals would widen the scope of applications for molecule specific enhancements by means of tagging. DNP radical tagged isotope labelled proteins could be introduced into the cells without the risk of reduction, if caged radicals are available. In a similar vein, taggable radicals that would irreversibly conjugate to the proteins of interest harboring an inert unnatural amino acid have recently been developed[22]. In complex and heterogeneous biological environments, it has already been shown that tagging the DNP agents directly to the protein of interest is advantageous in obtaining molecule-specific DNP enhancements[23,24].

In conclusion, the advancements in radical development could soon enable ssNMR experiments at much lower intracellular concentrations of proteins than is currently possible.

7.3 Improving resolution in DNP spectra

Due to the low temperatures and the strong couplings between electrons and nuclei, DNP-ssNMR spectra suffer from an inherent lack of resolution. As in the case of conventional NMR, adding spectral dimension(s) is a straightforward means to increase spectral resolution. We have demonstrated this strategy in our studies in Chapters 2, 4 & 6 by performing 3D experiments to resolve the spectra. Besides increasing the dimensionality, one could opt for higher strength of magnetic fields (B_0) to obtain better resolution as it has been shown before on DNP-ssNMR studies in-vitro[25].

Currently used water-soluble biradicals such as AMUPol and PyPol however fail to properly perform at higher magnetic fields[26]. For example, moving from 400 to 800 MHz DNP conditions in cells, leads to a reduction in the DNP enhancements in cells by a factor ~3.7 (see Chapter 2, Supporting Information, Figure S5). This is similar to values seen for in-vitro DNP applications to complex biomolecules[5]. Such a strong reduction in signal intensity currently still compromises the practical use of high-field DNP where spectral resolution is increased. Instead and as demonstrated in this thesis, performing 3D experiments at a lower field where DNP performance is

maximized is still often preferred. As mentioned above, the low DNP enhancements are intrinsic to the nature of the radicals currently used that contain two nitroxide groups. However, current chemical synthesis efforts are underway to design DNP radicals such as biradicals that contain trityl-nitroxide moieties which show superior performance at high magnetic fields[27].

Another, more indirect approach to reduce spectral congestion in DNP experiments has been demonstrated in Chapter 3, where the number of labelled residues in the protein are reduced by amino acid or residue specific isotope labelling. Although it does not inherently improve the resolution, this strategy reduces spectral crowding and thereby signal ambiguity. To this end, more complex sequential labelling schemes have been used in the past to reduce signal ambiguity in cellular preparations for DNP-ssNMR by zooming in on specific regions of interest in the protein[4,14]. If complete assignments are available, one could easily predict the spectra for a given ssNMR application to tailor the labelling scheme that leads to minimal ambiguity in the resulting 2D and 3D spectra[28]. As shown in Chapter 6, FANDAS 2.0 could play a powerful role in such studies.

References

1. Clark, L. *et al.* Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris. *J. Biomol. NMR* **62**, 239–45 (2015).
2. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51**, 2998–3001 (2012).
3. Baker, L. A., Daniëls, M., van der Cruijsen, E. A. W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62**, 199–208 (2015).
4. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat Meth* **12**, 5–9 (2015).
5. Kaplan, M. *et al.* Nuclear magnetic resonance (NMR) applied to membrane–protein complexes. *Q. Rev. Biophys.* **49**, e15 (2016).
6. Medeiros-Silva, J. *et al.* 1 H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angew. Chem. Int. Ed. Engl.* **55**, 13606–13610 (2016).
7. Almeida, F. C. L. *et al.* Selectively labeling the heterologous protein in Escherichia coli for NMR studies: A strategy to speed up NMR spectroscopy. *J. Magn. Reson.* **148**, 142–146 (2001).
8. Theillet, F.-X. X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).
9. Krstić, I. *et al.* Long-range distance measurements on nucleic acids in cells by pulsed EPR spectroscopy. *Angew. Chemie - Int. Ed.* **50**, 5070–5074 (2011).
10. McCoy, K., Rogawski, R., Stovicek, O. & McDermott, A. Stability of Nitroxide Biradical TOTAPOL in Biological Samples. *bioRxiv* 592766 (2019). doi:10.1101/592766
11. Lilly Thankamony, A. S., Wittmann, J. J., Kaushik, M. & Corzilius, B. Dynamic nuclear polarization for sensitivity enhancement in modern solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **102–103**, 120–195 (2017).

12. Leavesley, A., Wilson, C. B., Sherwin, M. & Han, S. Effect of water/glycerol polymorphism on dynamic nuclear polarization. *Phys. Chem. Chem. Phys.* **20**, 9897–9903 (2018).
13. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chemie Int. Ed.* **52**, 10858–10861 (2013).
14. Kaplan, M. *et al.* EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* **167**, 1241-1251.e11 (2016).
15. Sergeyev, I. V, Itin, B., Rogawski, R., Day, L. A. & McDermott, A. E. Efficient assignment and NMR analysis of an intact virus using sequential side-chain correlations and DNP sensitization. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5171–5176 (2017).
16. van der Cruijsen, E. A. W. *et al.* Biomolecular DNP-Supported NMR Spectroscopy using Site-Directed Spin Labeling. *Chem. - A Eur. J.* **21**, 12971–12977 (2015).
17. Albert, B. J. *et al.* Dynamic Nuclear Polarization Nuclear Magnetic Resonance in Human Cells Using Fluorescent Polarizing Agents. *Biochemistry* **57**, 4741–4746 (2018).
18. Narasimhan, S. *et al.* DNP-Supported Solid-State NMR Spectroscopy of Proteins Inside Mammalian Cells. *Angew. Chemie - Int. Ed.* **58**, 12969–12973 (2019).
19. Jagtap, A. P. *et al.* Sterically shielded spin labels for in-cell EPR spectroscopy: Analysis of stability in reducing environment. *Free Radic. Res.* **49**, 78–85 (2015).
20. Yang, Y. *et al.* A Reactive, Rigid Gd ^{III} Labeling Tag for In-Cell EPR Distance Measurements in Proteins. *Angew. Chemie Int. Ed.* **56**, 2914–2918 (2017).
21. Yang, Y., Yang, F., Li, X.-Y., Su, X.-C. & Goldfarb, D. In-Cell EPR Distance Measurements on Ubiquitin Labeled with a Rigid PyMTA-Gd(III) Tag. *J. Phys. Chem. B* **123**, 1050–1059 (2019).
22. Lim, B. J., Ackermann, B. E. & Debelouchina, G. T. Targetable Tetrazine-Based Dynamic Nuclear Polarization Agents for Biological Systems. *ChemBioChem* **21**, 1–6 (2020).
23. Rogawski, R. *et al.* Dynamic Nuclear Polarization Signal Enhancement with High-Affinity Biradical Tags. *J. Phys. Chem. B* **121**, 1169–1175 (2017).
24. Viennet, T. *et al.* Selective Protein Hyperpolarization in Cell Lysates Using Targeted Dynamic Nuclear Polarization. *Angew. Chemie Int. Ed.* **55**, 10746–10750 (2016).
25. Fricke, P. *et al.* High resolution observed in 800 MHz DNP spectra of extremely rigid type III secretion needles. *J. Biomol. NMR* **65**, 121–126 (2016).
26. Gast, P. *et al.* A tailored multi-frequency EPR approach to accurately determine the magnetic resonance parameters of dynamic nuclear polarization agents: application to AMUPol. *Phys. Chem. Chem. Phys.* **19**, 3777–3781 (2017).
27. Mathies, G. *et al.* Efficient Dynamic Nuclear Polarization at 800 MHz/527 GHz with Trityl-Nitroxide Biradicals. *Angew. Chemie Int. Ed.* **54**, 11770–11774 (2015).
28. Narasimhan, S. *et al.* Rapid prediction of multi-dimensional NMR data sets using FANDAS. in *Methods in Molecular Biology* **1688**, 111–132 (2018).

Samenvatting in het Nederlands

In recente jaren is er geen enkele unieke eiwitstructuur aan de PDB toegevoegd. Dit kan verklaard worden door de eindigheid van het aantal unieke eiwitstructuren dat in de natuur bestaat. Bovendien hebben traditionele structurele studies ongewild de cellulaire context van biomoleculen verwaarloosd, omdat er geen technieken bestonden die atomistische structurele studies in een heterogene cellulaire omgeving mogelijk maakten. Vooruitgang op het gebied van biomoleculaire NMR en cryo-EM technieken hebben de vorming en groei van cellulaire structuurbiologie gestimuleerd. Terwijl cryo-EM zich concentreert op zeer grote moleculaire machines en eiwitassemblages, heeft NMR kleine en grote heterogene eiwittenbestudeerd. De twee technieken complementeren elkaar dus. Een vergelijkbare complementariteit bestaat ook binnen NMR, waar studies die gebruik maken van vloeistof NMR gelimiteerd zijn tot kleine oplosbare eiwitten, terwijl vastestof NMR (ssNMR) studies van relatief grote, onoplosbare eiwitten mogelijk maakt. De motivatie achter het onderzoek gepresenteerd in dit proefschrift is vastestof NMR-methodes te ontwikkelen en toe te passen die ernaar streven oplosbare en promiscue eiwitten binnen de cellulaire omgeving te bestuderen. Ondanks hun oplosbaarheid blijven promiscue eiwitten buiten het bereik van vloeistof NMR, omdat ze in een compact cellulair milieu vertraagde rotatie en translatie ondervinden door de aanwezigheid van hun interactiepartners. We hebben een gevoeligheid-versterkte (DNP) ssNMR methode ontwikkeld en getest die met zowel zoogdier- als bacteriële cellen werkt.

Hoofdstuk 2 beschrijft een protocol om wild-type Ubiquitine (Ub) monsters in HeLa cellen voor DNP-ssNMR voor te bereiden. Eerst introduceerden we isotoop-gelabelde Ub in ongelabelde HeLa cellen die vervolgensook verrijkt werden met DNP-radicalen, door middel van passieve diffusie, vlak voor het ssNMR experiment. Vervolgens voerden we functionele experimenten uit om zeker te zijn dat het geïntroduceerde eiwit functioneerde in de cellen. Daarnaast hebben we gecontroleerd of de cellen de radicalen opnamen via diffusie. Vloeistof NMR-experimenten zijn uitgevoerd om te controleren op eiwitafbraak. Uiteindelijk konden we vaststellen dat het eiwit correct gevouwen is in de cellen door 2D- en 3D DNP-ssNMR spectra op te nemen en te analyseren.

Het is bekend dat, ondanks zijn kleine formaat, Ub bij vele regulerende functies in de cel betrokken is. Een van de belangrijkste rollen van Ub is het promoten van de afbraak van meerdere eiwitten door het proteasoom. Het is eerder aangetoond dat remmingvan het proteasoom een enorme stress voor de cel veroorzaakt en leidt tot de migratie van nucleaire Ub naar het cytoplasma om zich aan eiwitten te hechten die afgebroken moeten worden. In **hoofdstuk 3** bestuderen we het effect

van proteasoom-remming op verscheidene Ub residuen in de cel middels onze 2D DNP-ssNMR spectra. We onderzochten eerst uniform (dubbele) ^{13}C-,^{15}N-gelabelde Ub, waar we signalen van residuen van over het gehele eiwit op grote schaal zagen verdwijnen na proteasoom-inhibitie. Vervolgens concentreerden we ons op lysine (K) residuen, met name K48, met gebruik van residu-specifiek gelabelde Ub in cellen. We zagen een minimale toename in isopeptide-gelinkte K48 resonanties, naast een zwak signaal dat het idee van ongekarakteriseerde posttranslationele modificaties (PTM) ondersteunt. Tot slot vonden we bewijs voor conformationele heterogeniteit van de K48 *backbone* in proteasoom-geremde cellen. Uiteindelijk concluderen we dat Ub moleculen veel heterogener worden in de cel na proteasoom remming. Verdere biochemische experimenten zullen moeten worden uitgevoerd om onze interpretatie van de spectra te bevestigen.

In **hoofdstuk 4** breiden we onze in-cel DNP ssNMR methode uit naar oplosbare bacteriële eiwitten. Als testcase labelden we Ub in de *E. coli* BL21 stammen R2 en Lemo. Omdat we de eiwitten niet in de cel konden introduceren middels het protocol beschreven in hoofdstuk 2 en 3, gebruikten we de "rifampicin methode", die eerder was gedemonstreerd met membraanproteïnen, om specifiek nieuw gesynthethiseerd Ub te labelen. We werden met een hogere mate van achtergrondlabeling in de cellen geconfronteerd, vergeleken met de membraanproteïnetoepassingen waar celenveloppen waren gebruikt. We overkwamen het achtergrondlabelen tot op zekere hoogte en bereikten complete (90%) deuteratie van de achtergrond om maximale DNP-versterking voor Ub te krijgen. Tijdens dit proces kwamen we erachter dat de R2 stam ongeschikt is voor onze DNP-experimenten. We bereikten zeer slechte DNP-versterkingen wanneer we probeerden de achtergrondsignalen te verwijderen met gebruik van een algen aminozuurmengsel om Ub specifiek te labelen.

Kunstmatige metallo-enzymen katalyseren vele *new-to-nature* reacties. In de afgelopen jaren is het uitvoeren van katalyse in cellen zonder purificatie aangeprezen als methode die meerdere voordelen heeft, waaronder lagere kosten. LmrR, een eiwit gevonden in *Lactococcus lactis*, gedraagd zich als metallo-enzym wanneer het een Cu^{II} –fenantroline groep bindin het dimeer-interactieoppervlak door middel van een π-stapeling interactie met tryptofaan-indolen. In **hoofdstuk 5** labelden we selectief LmrR in *E. coli* en bevestigden we de correcte vouwing van het eiwit door middel van 2D- en 3D ssNMR-experimenten. Vervolgens observeerden we specifieke (paramagnetische) uitdoving van signalen van Trp aromatische ringen in LmrR en een gelijktijdige verlaging van T_1 relaxatiestijden, wat de vorming van het metallo-enzym bevestigt. Vloeistof NMR-studies van het lysaat onthulden het effect van de cellulaire componenten op de dynamica van LmrR. We verwachten deze informatie te gebruiken om LmrR te herontwerpen zodat het katalyse efficiënt

kan uitvoeren in de cel.

In **hoofdstuk 6** beschrijven we FANDAS 2.0, een *in-house* programma en webserver die we ontworpen op een Python platform. De applicatie kan gebruikt worden om multidimensionale NMR-spectra van eiwitten te voorspellen voor makkelijke visualisatie en analyse. We geven stap-voor-stap instructies voor het gebruik van het programma en beschrijven verschillende functies. Het heeft potentiële toepassingen voor het kiezen van verschillende aminozuur-labelschema's om spectra met goede spectrale resolutie te verkrijgen in ingewikkelde gevallen zoals in-cel NMR, dat over het algemeen neigt naar lage-resolutie spectra.

Acknowledgements

Before anything: my belated apologies to the ones who got "Good luck! -Sid" from me on their farewell cards. It's not personal. I just suck at saying goodbye appropriately.

Foremost, I thank **Prof. Marc Baldus** for giving me this opportunity. I am grateful to you for all the guidance, especially so during times when our research did not proceed as planned. I have admired your passion for NMR and your ability to come up with impressive out-of-the-box research ideas. What is also impressive is how you have played a huge role in putting together a solid-state NMR facility of such a magnificent scale in Utrecht. Over the time I spent in the group, I have seen you hire a bunch of wonderful colleagues. Many of them have become my good friends and have at some point been my mentors and/ or sources of inspiration, sometimes unwittingly. Thank you for that too! I also thank **Dr. Gert Folkers**, my co-supervisor who pretty much happens to be the only biochemistry teacher I have ever had. I have learnt a great deal from you in the form of small, distilled packets of information. You never cease to amaze me with the sea of biochemical knowledge you possess. I always got solutions to the problems I came running to you with. And as a bonus, you always have provided a comprehensive list of things to note, that I most likely overlooked. Thanks a ton! You and I are alike in how well we organize our desks and wet-lab benches, much to the other lab users' dismay. Haters may hate Gert, but we know that there is a very delicate order within our… intricate workspaces.

Speaking of such work spaces, **Mark Daniëls**, thank you for all the wonderful scientific discussions over beers and coffee. You are so multi-faceted that within a span of 10 mins, you can switch from being extremely silly & rowdy, to talk about the most sophisticated biochemistry. Over the years, you have contributed so much to diverse projects that it should be possible for you to claim multiple PhDs by now if you are willing to sit and write. We got to enjoy the benefits of having you on for way too long. Now it's time some other group gets that privilege. While we lost that privilege, we are privileged to have **Raymond Schellevis** with us. You are proving to be a great addition to the group! Thank you for organizing the lab very well by solving all the issues as they came. Thanks also to **Geeske Badart** for all her support and help in dealing with bureaucratic issues. Geeske, besides organizing our group, you have managed to make our lunch area and kitchen look so clean. I never believed I would see a day in the middle of the week, when the kitchen sink was not full of dirty dishes. There should be some award or multiple awards for all your work. I also thank **Barbara Hendricx** for her support during her time in the group and

for organizing matters relating to scheduling my defense.

I thank **Mohammed Kaplan** for his support and supervision during my masters. My pleasant experience in the group then, led me to stay on to pursue this PhD. I addition to Mohammed, I also thank **Klaartje**, for all the support and feedback during our group meetings. Your incredible degree of competence in many diverse fields and your eye for the tiniest details are truly mind blowing! When I moved here at the age of 21, having been raised in an incredibly divided and sexist society, I could not for a long time empathize with the issues you cared about. I eventually did. Your being vocal made me curious about things outside of work, and inspired me to learn about deep-rooted (sometimes subtle) discrimination that exists all around us. These were valuable life lessons. Thank you for that too!

Stephan, thanks a ton for synthesizing the many amazing compounds and proteins mentioned in Chapters 2 & 3. I look forward to finishing up the last bit of the proteasome inhibition story. It was also a pleasure to work with **Shreyans**, whom I have known for almost a decade now. I never expected us to collaborate at some point, that too, after an unexpected encounter at CHAINS. I also thank the people involved in the NMARRS program, especially **Julia Krug** and **Rubin Dasgupta** whom I have gotten to know over the last years.

Marie, before you joined the group, I had heard wonderful things about you from **Marc** and **Mohammed**, among others. You began the cellular studies in our first stint in the group, and we continue to stand on your shoulders. Thank you for all the great discussions and offering very timely help with refining my manuscript(s). I thank **Meg Ward** for maple whiskey, banana bread and for making the group meetings very enjoyable. I received nearly all of my initial training at the spectrometer from **Deni**, who still continues to teach me a thing or two about NMR when we happen to talk. The best part about learning concepts from you, Deni, is that you know how to deliver the exact amount of information needed for an average spectrometer user like me to help me figure things out on my own. This is why many of us fellow PhD students, have bothered you during our struggles with the magnets. Thank you so much for this! Along with Ivan, you made the office very pleasant during the first two years of my PhD. Which brings me to **Ivan**. I have seen you evolve from being a terrible at presenting, to be the best presenter in the whole group. The remarkable twists in your excellently delivered presentations were all delights. Theatre really brought out the best in you, I hope you keep doing it. Thank you for so many valuable discussions! You along with **Panos** have taught me the importance of being silent.

I also had the joy of sharing my office for a short while with **Cecilia**. Besides being a brilliant scientist, you are incredibly sympathetic and caring. I admire how efficiently you manage your work and your personal life, while also commuting for at least two hours everyday. While, I must add, you probably worked with the most complex, and challenging system at the time in our lab. I also now have the unexpected joy of working with you on a paper, which is turning out great thanks to your efforts and feedback. Finally, I thank you for all the great advice over our fresh-air breaks. I wish you great success in your career and hope you get to have lots of cats!

Until very recently, I also had the joy of sharing the office with **Agnes**. As much as it was annoying to have Čorba and you speak German to each other, I began missing it after he left. I was just about beginning to be able to eavesdrop well enough to understand what you two were on about. It has been just a year since you joined the group and you have already managed to procure very promising data. I am sure you will have a very productive time here. Which brings me to the other new people in the group, **Anamika** and **David**. David, welcome to the group and to the best office in the bunker. I hope you like saunas, because our office is going to be that every summer, sans steam. I wish you good luck with the project you will be taking over from me. I look forward to seeing some nice papers! **Anamika**, you are proving to be a great addition to the group. It has been fun to get to know you. Seeing how diligent and hardworking you are, I am sure you would do a wonderful job in the coming years.

For whatever the reasons may be, the computational structural biology group is a part of our building. I was lucky to be a part of it for a very small amount of time. I thank the very efficient **Prof. Alexandre Bonvin** for giving me a chance to work on a very interesting project. I remember waltzing in assuming that I knew programming because I had had half-baked C and C++ courses in college. Within weeks, I learnt how little useful computational skills I had. Thanks to the then group, especially **João Rodriguez**, **Anna**, **Gydo** and **Dr. Geng**, I managed to pick up quite a few computational skills which I continue to use till date. Also, **Dr. Geng**, thank you for your guided food experiences at Chinese restaurants. I thank **Mikael** and **Adrien** for all the fun discussions during lunch. The two of you are natural and prolific narrators. I do not remember a dull lunchtime when either of you was around. I thank **Zuzana**, **Siri**, **Jorge**, **Rodrigo** and **Francesco** for all the good times during borrels. Zuzana and Rodrigo, you both are so radiantly happy that it is impossible to be grumpy around you. **Siri**, I hope you make the board games a regular thing again after the lockdown ends. Finally, Dr. Borbonne, I can forgive

you for your horrible taste in coffee because you have provided me with enough entertainment through your Italian cuisine supremacy theories and by attempting to speak Dutch. It was fun to go to the trashy gym with you for free, and too bad we cannot go now because of the lockdown. I wish you good luck in Switzerland and hope you get to visit Rome more often to drive your Fiats around.

I thank **Hugo van Ingen** and his group for all the help with solution NMR. **Hugo**, I have admired your ability to give clear and perfectly structured presentations in your natural speaking style. Thanks also for the few NMR lectures you gave me during my master's. I wanted to learn crystallography during my internships in UU. Your lecture in the MCLS introduction course was the reason I decided to go off the course and try NMR for a change. Thank you also for bringing in some great people. **Ulric**, your lectures on everything ranging from NMR to hyena genitalia have been highly insightful and/or entertaining. **Heyi**, thank you for being a true OG amongst us nerds. Thank you also for pitching the idea of buying a blublublu. **Clara**, you and **Vincenzo** are the most levelheaded people I have ever known. Vincenzo, hang in there! You have a mammoth protein, which might have been hard to work with so far. But I have seen and admired your incredible dedication to your work which I am sure in due course would lead to wonderful publications. **Velten**, it was fun to talk to bullshit with you about everyting ranging from history to dumbells.

Reinier and **Jon**, whom I have known since my master's days, the 4ish years would not have been so enjoyable in the group without you two. Reinier, only you in our whole group can truly comprehend how dear FACS was to me. We have had many deep discussions on science and protein shakes. Besides being my cell biology consultant, you have taught me how to truly savor Kwark and Kaki fruit. I have seen you achieve so much by managing to work efficiently between sensible hours during the day. I wish I had the organizational skills that you have (except for when it comes for spoons). Jonnyboi, the last part except the spoon bit goes to you too. I have seen your inner cartoon character emerge during the 4 years. It has been fun to go around the Netherlands savoring Ramen like we are some connoisseurs. Hope you enjoy your new job and we get to resume our Ramen tour soon. I thank my mate **Leo**, for showing me around her wonderful town and all the fun times in the group. Having you as a colleague and a neighbor was every bit memorable. **Barend**, je bent incredibly grappig and multi-talenteerd. Someday, ik will take up your uitnodiging and Vastelaoend visiten. Succes met your work en muziek.

I thank **Markus Weingarth** for his valuable discussions in the groupmeetings and for bringing wonderful colleagues to the group. **Maik**, welcome

to the group! **Rhythm**, your chaperone Azo tells me that you are incredibly sharp. Having spent just a year in the group, you have managed to publish your first paper. Congrats and good luck on more wonderful things to come!

The two people in the group who have helped me the most with my work on a daily basis are my Paranymphs, **Johan** and **Alessandra**. Johan, you are an invaluable asset to our department. Your incredible ability to solve any problem that the infrastructure would throw at us is remarkable. There can be books written about all the ways you have solved our issues and made lives so easy. I have thoroughly enjoyed having long conversations with you while filling LN_2 on everything from fusion reactors to your love for crappy self-driven cars. You have gone extra miles for all of us and then you shrug off compliments, saying that it is after all your job. A big thanks for everything! **Luci**, I don't know if you can recall, in 2016, I was never keen on doing DNP. Look how the thesis turned out now! I can't count the number of times I would go back home at wee hours of the morning to then wake up to a reassuring text message from you with a spectrum showing me things worked. It helped me get enough sleep! Without your contribution to the in-cell work, things would have taken several times longer than they did. I could never get tired of your optimism and have hence always turned to you whenever I either felt sulky about a paper taking too long to publish or to whine about failed experiments. Lastly, I have learnt a lot of DNP and NMR from you. Thanks a ton!

Miranda, since you are officially moving out of Gouda, I won't call you that. Having seen how you supervise masters students has made me want to do masters again. I have not seen anyone who is so willing to go out of their way to help people. We both traverse through several worlds and cultures each day. Therefore, I have felt at home with you being around, especially when everyone goes on about wrapping christmas presents or skiing and sailing. You know what I mean. Toch? Congrats on your house and I am glad that if I am ever homeless, you have offered to adopt me in return for Maqluba. And oh, I wish you all the best with your future cat cafe in Greece.

Charlotte, firstly, thank you so much for the Dutch translation and the suggestions with the cover. Also thank you for all the support you have provided me despite me having annoyed you almost as much as Azo. I admire your near-insane ability to have a reasonably deep knowledge on very random things. You may consider yourself a pancake person like me, but I think you are more like a pizza (Deep dish, of course) with incredible toppings on them. Like cheese and more cheese and more cheese. And more cheese. Unlike me, you have an incredible ability

to be organize both your work and spaces around you. I am learning this from you, while also learning to appreciate normal food. Please please start making Nachos for borrels again.

Which lastly brings me to **Panos** and **João** with whom I spent the most amount of time at and outside of work in the last 4ish years. You both know all the ways you have helped and supported me, I won't go into it all here. Thank you both for all that. Azo, we learnt recently that Čorba and I may have messed up in 2015. Sorry for that! We could have been flat mates. Although I am sure you would have killed me and/or killed yourself. Whatever happened, happened for the best. Azo, you know how to make everyone around you happy. You go about life seeing the best in people and situations, which is a rare trait. I have always admired your dedication to your work, even when things do not always turn out to be perfect. Any group you are going to end up in will be lucky to have you on board. I hope to do more trips with you, we have to beat Karl's record. *O'right*? Koukos, I cannot call you the other thing I call you normally, which is one typo away from Malala. Behind the expressionless, stoic veneer, you are so full of life. Anyone who has gotten past the small energy barrier to get to know you would testify this. I have loved geeking out on so many things. Thank you for all that and teaching me how to BBQ. You know, not having grown up in this part of the world, I didn't know all that stuff. Your ability to be very self-aware and read the room appropriately has always amazed me. This of course comes also with an incredible attention to detail which is what makes you a great scientist that you are. I wish you good luck for whatever you choose to do next. Finally, thanks to you and your graphic designer for help with my cover.

Good luck! -Sid

Curriculum Vitae

Siddarth Narasimhan was born on the 27th of August 1992 in Hyderabad, India. He obtained his Bachelor of Technology (B.Tech.) in 2013, specializing in Industrial Biotechnology from SASTRA University, India. He then obtained a Master of Science (M.Sc.) in Biology (Molecular and Cellular Life Sciences) from Utrecht University, The Netherlands in 2015. During his master's studies, he carried out a major research project, where he used solid-state NMR to study large membrane proteins and complexes in their native environments, under the supervision of Prof. Marc Baldus and co-supervision of Dr. Paul van Bergen en Henegouwen. For his minor research project, he studied protein-peptide interactions using MD simulations and computational docking under the supervison of Prof. Alexandre M.J.J. Bonvin. In 2015, he joined the research group of Prof. Marc Baldus to study soluble and promiscuous proteins in the cellular context using solid-state NMR approaches.

List of Publications

1. **S. Narasimhan**, C. Pinto, A. Lucini Paioni, G. E. Folkers and Marc Baldus. A solid-state NMR protocol to characterize proteins in a native bacterial environment. In preparation for an invited contribution to Nature Protocols
2. S. Chordia, **S. Narasimhan**, A. Lucini Paioni, M. Baldus, G. Roelfes. An Artificial metalloenzyme catalyzes a new to nature reaction in the cytoplasm of bacteria. In preparation
3. **S. Narasimhan**, G. E. Folkers, M. Baldus. When small becomes too big: Expanding the use of in-cell solid-state NMR spectroscopy. ChemPlusChem, in-press DOI: 10.1002/cplu.202000167
4. R. Damman, **S. Narasimhan**, M. Weingarth, M. Baldus. Cellular solid-state NMR spectroscopy. New Dev. NMR 2020-January, 131–151.
5. **S. Narasimhan***, S. Scherpe*, A. Lucini Paioni, J. van der Zwan, G. E. Folkers, H. Ovaa, M. Baldus. DNP-Supported Solid-State NMR Spectroscopy of Proteins Inside Mammalian Cells. Angew. Chemie - Int. Ed. 2019, 58, 12969–12973.
6. **S. Narasimhan**, D. Mance, C. Pinto, M. Weingarth, A. M. J. J. Bonvin, M. Baldus. Rapid Prediction of Multi-dimensional NMR Data Sets Using FANDAS. Methods in Molecular Biology 2018, pp. 111–132.
7. C. Geng*, **S. Narasimhan***, J. P. G. L. M. Rodrigues, A. M. J. J. Bonvin. Information-Driven, Ensemble Flexible Peptide Docking Using HADDOCK. Methods in Molecular Biology 2017, pp. 109–138.
8. A. Vangone, J. P. G. L. M. Rodrigues, L. C. Xue, G. C. P. van Zundert, C. Geng, Z. Kurkcuoglu, M. Nellen, **S. Narasimhan**, E. Karaca, M. van Dijk, A. S. J. Melquiond, K. M. Visscher, M. Trellet, P. L. Kastritis, A. M. J. J. Bonvin. Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. Proteins 2017, 85, 417–423.
9. M. Kaplan, **S. Narasimhan**, C. de Heus, D. Mance, S. van Doorn, K. Houben, D. Popov-Čeleketić, R. Damman, E. A. Katrukha, P. Jain, W. J. C. Geerts, A. J. R. Heck, G. E. Folkers, L. C. Kapitein, S. Lemeer, P. M. P. van Bergen en Henegouwen, M. Baldus. EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. Cell 2016, 167, 1241–1251.e11.
10. D. Mance, T. Sinnige, M. Kaplan, **S. Narasimhan**, M. Daniëls, K. Houben, M. Baldus, M. Weingarth. An Efficient Labelling Approach to Harness Backbone and Side-Chain Protons in (1) H-Detected Solid-State NMR Spectroscopy. Angew. Chem. Int. Ed. Engl. 2015, 54, 15799–15803.
11. M. Kaplan, A. Cukkemane, G. C. P. van Zundert, **S. Narasimhan**, M. Daniels, D. Mance, G. Waksman, A. M. J. J. Bonvin, R. Fronzes, G. E. Folkers, M. Baldus. Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. Nature Methods 2015, 12, 5–9

* Equal author contributions