DISSOLUTION DYNAMIC NUCLEAR POLARIZATION OF POLYPEPTIDES

A Dissertation

by

MUKUNDAN RAGAVAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank Raushel
Co-Chair of Committee,	Christian Hilty
Committee Members,	James Sacchettini
	Paul Straight
Head of Department,	Gregory Reinhart

August 2014

Major Subject: Biochemistry

ABSTRACT

Nuclear Magnetic Resonance (NMR) spectroscopy provides remarkable site resolution, but often requires signal averaging because of low sensitivity. Dissolution dynamic nuclear polarization (DNP), which offers large signal enhancements, has been used to follow reactions involving small molecules that typically have long spin-lattice relaxation times. This thesis presents work in the development of dissolution DNP to directly hyperpolarize and observe polypeptides, which can subsequently be used for the study of a time dependent process, such as folding.

Dissolution DNP involves hyperpolarizing samples in the solid state, dissolving the samples with a stream of hot solvent and rapid transfer of the sample into an NMR tube for measurement in the solution state. Since protein samples are prone to foam under conditions for rapid sample injection, solvent systems were optimized. Solvents such as water/acetonitrile and water/methanol mixtures were utilized. An unlabeled peptide, bacitracin, was hyperpolarized on ^{1}H nuclei and enhancements of 30, 45 and 180 were obtained for amide, aliphatic and aromatic protons respectively. Although these enhancements are already significant, loss of hyperpolarization during sample injection was further alleviated by the use of isotopically enriched polypeptides. In [^{13}C, 50% ^{2}H] labeled samples of denatured L23, a 96 amino acid long ribosomal protein, signal enhancements of more than 500 times were obtained on the ^{13}C nuclei. This signal enhancement was then exploited to follow the protein folding process, using L23 as a model. Time resolved spectra of hyperpolarized L23 were measured after a pH jump and protein folding was monitored by observing changes in the carbonyl region of the spectra, which are indicative of the formation of secondary structures. Despite signal overlap in the protein spectra, using the statistical distribution of ^{13}C chemical shifts, the fractions of secondary structure elements were estimated for each scan of the DNP-NMR experiment. Additionally, individual resonances for methyl groups upfield of other protein resonances became resolved in the later transients. An option for the improvement of such site resolution by NMR experiments using coherence selection is discussed.

While DNP-NMR offers the capability to observe transient species, identification of such species is difficult in cases where not all chemical shifts are known. Here, a new strategy for the analysis of DNP-NMR data is proposed based on non-negative matrix factorization (NNMF). NNMF enables identification of various sources that contribute to an observed signal. This capability is demonstrated using a series of spectra measured from an enzymatic conversion reaction of oxaloacetic acid to malic acid. Simulations were carried out to evaluate the performance of NNMF under different experimental constraints, and the strengths and limitations of the method are discussed based on the simulations.

DEDICATION

கவையாகிக் கொம்பாகிக் காட்டகத்தே நிற்கும்
அவையல்ல நல்ல மரங்கள் - அவைநடுவே
நீட்டோலை வாசியா நின்றான் குறிப்பறிய
மாட்டாதவன் நன்மரம்
--ஔவையார்

A rough translation of this poem to English would read as ”Trees with branches, twigs and leaves that are present in the forest are not real trees. He who is not learned and is incapable of articulating his thoughts in an assembly of scholars is the one who is a tree”.

This 2000 year old poem was something that my grandmother used to recite to me almost every day during my formative years to stress on the importance of education. I have no doubt that she would be happiest person on the planet had she been alive the day I defend my thesis. Although it has been 14 years since she passed away, I fondly recall the time I spent in her company.

I dedicate this thesis to my beloved grandmother.

ACKNOWLEDGEMENTS

Wahrlich es ist nicht das Wissen, sondern das Lernen, nicht das Besitzen sondern das Erwerben, nicht das Da-Seyn, sondern das Hinkommen, was den grössten Genuss gewährt.

– Carl Friedrich Gauss

A rough translation of Gauss' words reads as "Truly, it is not knowledge, but learning, not possession, but acquisition, not being there but getting there, which gives the greatest enjoyment." As I look back on the time I have spent in research over the last few years, I realize I can relate to Gauss' words above very easily. I am obliged to thank a large number of people who have played a significant part in making my journey so far pleasurable.

Foremost, I thank my advisor, Christian, for his unwavering support through the entire time I have spent in the lab. I am, particularly, very grateful for his support and understanding during the time when my health was at its absolute worst. I am very thankful for the patience he showed as I was trying to get the protein experiments to work with the stopped flow setup. Aside from my own research, I am glad that he sought my participation in various outreach activities like the Chemistry Open House and the Youth Adventure Program (YAP). Participating in YAP, in particular, gave me an opportunity to interact with and teach small children and that experience has been gratifying. My involvement in assisting Christian with writing proposals which usually took the form of bouncing different ideas back and forth and diving into the literature to fine tune a specific idea were invigorating and I am extremely thankful for that opportunity. The numerous chats on topics ranging from "Are vegetarians likely to consider lab grown meat acceptable?" to "How to set up virtual machines and install different softwares on the lab server?" have provided me a lab environment that is fun, energetic and at the same time, focused.

I thank the members of my advisory committee, Dr. Raushel, Dr. Sacchettini and Dr. Straight for their consent to be members in my committee and for their constant support and guidance throughout the course of this work. The committee meetings and the preliminary examination were filled with intense scientific discussions which have played an important part in giving me clear directions on what aspects of my training need improvement.

I thank Dr. Mikael Oliveberg from Stockholm University for providing us with the plasmid containing L23 gene. I also thank our collaborator Dr. Richard Kriwacki for sending us isotopically labeled p27 samples.

I would be failing in my duty if I do not thank Hsueh-Ying Chen. His enthusiasm and constant desire to learn was infectious. It has been a great pleasure discussing science with him. His expertise with handling the injection hardware has been valuable in protein experiments on more than one occasion. I am glad to have had the opportunity to work with Hsueh-Ying.

The first person from Texas A&M whom I contacted before coming to College Station and the first person I met when I arrived in College Station for the first time in 2007 was Giridhar Sekar. My dinner on the first day I was in College Station was *aviyal*, a traditional South Indian delicacy. It would be no exaggeration to say that the first meal in College Station played no small part in my prolonged stay here! His immense desire to do correct science has always been inspirational. I have sought Giri's advice on more than one occasion and on every occasion, his advice has proven to be valuable. His encyclopedic knowledge on a large number of unrelated topics - particularly cooking and music - have made for a large number of interesting conversations.

I would not have been able to concentrate very much if all I did was work. The time I have spent with many of my friends have been filled with fun. My room-mate for 6 years Dr. S. P. Atul Narayan was a valuable companion till he went back to India last year. Looking back, I am amazed about the variety of situations we have seen and handled over the six years. Many thanks are due to Atul for his help with using Matlab in April 2009. I can confidently say that neither of us could have asked for better room-mates. Atul's brother, Keshav, has been constant source of fun and entertainment ever since he joined Atul and I two years ago. Keshav and I share a deep seated interest in understanding Hindu religious texts and our discussions have been educational on that front. Dr. Shriram Srinivasan and I have spent many hours watching cricket whenever the Indian Cricket team was playing. It has often been painful to watch the Indian team lose (quite badly at that) - but we have "gritted our teeth" through those losses. Of course, the happy times need no reminder - one such occassion was India winning the ODI world cup. Discussions with Jayavel Arumugam have almost always reminded me to not be a "frog in a well". The time spent with Balaji Kothandaraman has always been light hearted and thoroughly relaxing - be it watching cricket or movies or eating at Genghis Grill. Atul and Shriram - special thanks to you guys for introducing me to LaTeX. Atul, Keshav, Shriram, Jayavel and Balaji - Thanks for all the good times!

Last but not the least, I thank my father, mother and sister for their love and unwavering support. It was a time of financial hardship when I started applying to different PhD programs. Without my parents' support, I would not even have started a PhD program let alone completing one. I am very thankful for their trust in me when I decided to choose higher education instead of the well paying job for which I had an offer. My sister, Sowmya, has been a source of strength and a trusted confidante. Watching her complete undergraduate studies and do well in her career so far gives me great happiness. My family is my most cherished treasure in life and knowing that they will be happy now is more than anything I can ask for.

NOMENCLATURE

BSS	Blind Source Separation
CD	Circular Dichroism
CIDNP	Chemically Induced Dynamic Nuclear Polarization
DHFR	Dihydrofolate reductase
DNP	Dynamic Nuclear Polarization
GdnCl	Guanidium Hydrochloride
DMSO	Dimethyl Sulphoxide
DSS	Sodium 2,2-dimethyl-2-silapentane-5-sulphonate
ECOSY	Exclusive Correlation Spectroscopy
HMQC	Heteronuclear Multiple Quantum Coherence
HPLC	High Pressure Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
PHIP	ParaHydrogen Induced Polarization
MES	4-Morpholinoehthanesulphonic acid
NMR	Nuclear Magnetic Resonance
NNMF	Non-Negative Matrix Factorization
OX-63	Tris[8-carboxyl-2,2,6,6-tetra[2-(1-hydroxymethyl)]-benzo-(1,2-d:4,5-d')-bis(1,3)dithiole-4-yl]methyl sodium salt
SABRE	Signal Amplification By Reversible Exchange
TEMPOL	4-hydroxy-2,2,6,6-tetramethylpiperidine-1-oxyl
TOCSY	TOtal Correlation SpectroscopY
T_1	Spin–Lattice Relaxation Time
T_2	Spin–Spin Relaxation Time

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the go-to techniques in chemistry and biochemistry. NMR spectroscopy offers unparalleled site resolution for a wide variety of chemical species and enables study of analytes usually without the need for chemical modifications. However, this technique suffers from inherently low sensitivity, because of which high sample concentrations and/or extensive signal averaging is needed. However, hardware improvements including the development of cryogenically cooled probes have aided in obtaining higher signal to noise (S/N) ratio in the spectra to some extent.

Another approach to address the S/N ratio is to use various hyperpolarization techniques, one of which is dissolution dynamic nuclear polarization (DNP). Here, electron polarization is transferred to the nucleus of choice (^{1}H, ^{13}C, ^{19}F, etc.) by irradiating the sample with microwave radiation at appropriate frequency at liquid helium temperatures. The sample is subsequently dissolved using a stream of hot solvent, and NMR spectra are measured in the solution state at ambient temperatures. Dissolution DNP has so far been used in a range of applications extending from study of organic reactions to enzymatic reactions and protein ligand binding. Further, dissolution DNP has also been used extensively to investigate metabolic pathways both *in vitro* and *in vivo*.

This thesis describes experiments that extend the application of dissolution DNP to observe full length proteins. Chapter 3 describes the solvent systems used to enable transfer of protein samples into the NMR tube and subsequent measurement of NMR spectra. Obtained enhancements are discussed in the context of relaxation times of different sites.

Experiments demonstrating the study of protein folding in Chapter 4 provide evidence for the capability of performing time resolved experiments with full length proteins. Protein folding is monitored by following the NMR signal from backbone carbonyl resonances of a model protein, L23. Information about secondary structure content is obtained using chemical shift information from external databases. A discussion about solvent systems used in these experiments and potential alternative strategies are presented. A multiple quantum coherence based NMR experiment is described to spectroscopically gain site resolution in dissolution DNP experiments. The experiment is demonstrated using ^{13}C–labeled acetate in hyperpolarized experiments and with L23 in non-hyperpolarized experiments. An experiment where the transfer of protein samples using only aqueous solvents is demonstrated. An intrinsically disordered protein, p27, is hyperpolarized on ^{13}C nuclei, dissolved using aqueous buffer and mixed with cyclin dependent kinase/cylin-A complex. Secondary structure formation was observed by measuring a time-series of ^{13}C spectra.

Lastly, chapter 5 describes analysis of dissolution DNP data using a blind source separation algorithm. Although dissolution DNP offers the ability to record spectra with high S/N ratio, NMR signal from low concentration reaction intermediates may appear with low signal intensity. Since signal averaging is not possible with dissolution DNP experiments, alternative strategies to improve the S/N in DNP spectra would have to be explored. One such approach is to use blind source separation algorithms like Non-Negative Matrix Factorization (NNMF). NNMF is used to parse multiple sources that contribute to the observed DNP enhanced NMR signal. The method is demonstrated by using it to analyze the enzyme catalyzed conversion of oxaloacetic acid to malic acid. Simulations for evaluating the performance of NNMF algorithm under different conditions such as poor signal to noise ratio in the spectra and variations in the relaxation times of different resonances are discussed.

2. NMR SPECTROSCOPY AND DYNAMIC NUCLEAR POLARIZATION

2.1 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a technique that has found use in diverse areas ranging from organic chemistry [1] to determining protein structure [2, 3, 4] and understanding dynamics [5, 6, 7]. The wide array of experimental conditions that can be accessed using solution NMR spectroscopy and the available site resolution makes it an extremely valuable tool in research.

NMR spectroscopy relies on the fact that NMR active nuclei (^{1}H, ^{13}C, ^{15}N, etc.) occupy non-degenerate energy levels when placed in an external magnetic field as explained by the so called *Zeeman effect*. The difference in energy between the two levels for a spin-$\frac{1}{2}$ nucleus is given by the equation

$$\Delta E = h\gamma B_0 = h\nu \tag{2.1}$$

where ΔE indicates the energy difference between the two Zeeman levels, h the Planck's constant, γ the gyromagnetic ratio of a specific nucleus, B_0 the external magnetic field and ν the Larmor frequency of the nucleus.

The Larmor frequency denotes the precession of the nuclear spin in the presence of an external magnetic field and the frequency of precession is directly proportional to the strength of the external magnetic field and the gyromagnetic ratio of the specific nucleus. A list of nuclei commonly used in biomolecular NMR spectroscopy and their precession frequencies at 9.4 T are shown in Table 2.1.

Table 2.1: Table showing some NMR active nuclei with their gyromagnetic ratios and their Larmor frequencies at 9.4 T. Sign of the gyromagnetic ratio indicates the direction of precession with negative sign referring to anti-clockwise precession [8]

Nucleus	Gyromagnetic ratio (MHz/T)	Larmor Frequency (MHz)
^{1}H	42.58	400.2
^{2}H	6.536	61.44
^{13}C	10.71	100.7
^{15}N	(-)4.316	40.57
^{19}F	40.05	376.5
^{31}P	17.24	162.1

The fraction of spins occupying each Zeeman energy level is dictated by the Boltzmann distribution, which can be expressed as

$$\frac{N_\alpha}{N_\beta} = e^{\frac{-\Delta E}{k_B T}} \tag{2.2}$$

where N_α and N_β denote the population of "spin up" and "spin down" states respectively, k_B the Boltzman constant and T the temperature. Since NMR signal arises only from the excess spins in one state over the other, a more meaningful measure to evaluate is the polarization, P, which is the ratio of difference of populations between the two spin states to the total population. Mathematically, polarization is expressed as

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} \tag{2.3}$$

Combining equations 2.2 and 2.3, it can be shown that

$$P = tanh\left(\frac{h\gamma B_0}{2kT}\right) \tag{2.4}$$

At a temperature of 298 K and magnetic field of 9.4 T, the value of polarization is of the order of 10^{-5}. The lack of sensitivity of NMR stems from the fact that the observed signal arises from a weakly polarized system. This lack of sensitivity has led to the development of various approaches towards signal enhancement by hyperpolarizing the spins prior to measurement of NMR signal. These approaches are discussed later in this chapter.

2.1.1 Protein NMR Spectroscopy

Solution NMR spectroscopy can provide information about both structure and dynamics of proteins. This makes it a very useful technique to study proteins in isolation, and also to monitor protein interactions with its binding partners (ligands or other proteins). The advent of recombinant protein expression and rapid development of NMR techniques has greatly expanded the avenues of applications over the last three to four decades. NMR spectroscopy, along with X-ray crystallography have become the *de facto* standard for structure determination of macromolecules, often providing complementary information.

2.1.1.1 Protein Structure Determination

From the early days when Wüthrich and coworkers first determined the structure of bull seminal proteinase inhibitor [9], NMR spectroscopy has been applied to thousands of proteins.

The first step towards determining structures of proteins is to obtain chemical shift assignments. The assignment of chemical shifts refers to the process of uniquely identifying all the resonances present in NMR spectra. In the case of smaller proteins, homonuclear (^{1}H) experiments may suffice. Two dimensional (2-D) experiments such as Correlation Spectroscopy (COSY) and Total Correlation Spectroscopy (TOCSY) provide through bond correlations, while a Nuclear Overhauser Effect Spectroscopy (NOESY) experiment provides through space correlation. The data obtained from these experiments together can provide complete chemical shift assignments. However, since the signal overlap becomes insurmountable in larger systems, heteronuclear (^{1}H, ^{13}C, ^{15}N) experiments are preferred. Samples with uniform ^{13}C and ^{15}N labeling can be used to measure 2-D experiments including [^{1}H-^{15}N] heteronuclear single quantum coherence (HSQC) and [^{1}H-^{13}C] HSQC experiments for observing ^{1}H-^{15}N and ^{1}H-^{13}C correlations. The former is useful as an index to assess the sample quality and state of the protein (folded vs unfolded). Chemical shift assignments for the protein backbone (and C^{β}) can be obtained from triple resonance experiments such as HNCO, HNCA, HNCACO and HNCACB [10]. Further, complete side chain assignments can be extracted using other triple resonance experiments including HCCCONH and HCCH-TOCSY. Once the chemical shifts of different resonances are known, further information on secondary structures can be obtained by calculating the dispersion of the chemical shift from the random coil values. Typically, the variation of chemical shifts of H^{α}, C^{α}, C^{β}, amide protons (NH) and N resonances of a specific protein from the random coil chemical shifts are calculated. Using the proposed cut-off values for these differences [11, 12] secondary structures in different regions of the proteins can be predicted.

Although chemical shift assignments can yield information on secondary structures in different regions, to determine the actual tertiary structure of a protein, structural constraints must be obtained. The primary constraints in NMR based structure determination are ^{1}H-^{1}H NOEs. Experiments such as homonuclear NOESY for short peptides or using ^{13}C (or ^{15}N) resolved NOESY (NOESY-HSQC) experiments can provide the distance constraints [13]. Distances can be estimated from the intensity of NOE cross-peaks since NOE intensity is inversely proportional to the sixth power of the distance between the two protons. Further, based on the identities of NOEs, it would be possible to determine the secondary structure. For example, NOE cross peaks between residues *i* and *i+4* are diagnostic of α–helices, while long range NOEs are diagnostic of β–sheets.

In addition to distance constraints, torsion angles can be obtained from scalar coupling (J–coupling) values determined experimentally using the Karplus equation [14]. Scalar coupling constants can be measured by utilizing experiments of the Exclusive Correlation Spectroscopy (ECOSY) type [15, 16] or from heteronuclear experiments mentioned earlier. Residual Dipolar Couplings (RDC) can provide orientation information of different parts of protein structure [17].

To determine the structure of a protein, an initial model is constructed using the experimentally obtained distance constraints, torsion angles and RDC values (if available). This model is computationally refined to arrive at an energy minimized ensemble of structures which conform to all or most of the constraints provided initially [18, 19]. Regions of the protein which are mobile (e.g. loops) may have large RMSD values since NOE cross peaks from those "flexible" regions are typically poorly defined. However, the "flexibility" of these regions can be quantified using NMR experiments that are tailored to extract information on protein dynamics and time scale of motion can be experimentally determined.

One of the biggest limitations of the NMR spectroscopy is the size of the proteins that can be studied. As the proteins become larger, the correlation time becomes longer which results in rapid decay of NMR signal. This situation is further complicated by heavy overlap of signals in the NMR spectra. Several strategies have been explored to overcome these problems including development of novel pulse sequences, development of data sampling approaches to improve spectral resolution and selective labeling approaches.

An elegant spectroscopic approach was the development of Transverse relaxation optimized spectroscopy (TROSY) experiment [20]. Interplay between different relaxation mechanisms like dipole-dipole relaxation and chemical shift anisotropy contribute to overall relaxation of the signal in an NMR experiment. In the TROSY experiment, destructive interference between different mechanism is used to select a single well defined peak for each site on the protein. When the TROSY scheme is implemented in a [^{1}H,^{15}N]-HSQC experiment, it yields a spectrum with sharp peaks specifically for large proteins [20, 21, 22].

A novel spectroscopic approach towards obtaining better spectral resolution and enabling faster acquisition is non-uniform sampling of data from 3-D or 4-D NMR experiments. Non-uniform sampling (NUS) is a deviation from the current paradigm of collecting time domain data in linear increments [23]. One of the first proposals was to collect time domain data with exponentially decreasing separation of data points. This approach has been applied in conjunction with various triple resonance experiments. In a work reported by Wagner and coworkers, acquisition time was reduced from 6 days for complete set of triple resonance experiments (HNCACB, HNCA, HNCO, HN(CO)CA and HN(CO)CACB) to 32 hours for a 11 kDa protein [24]. The spectra obtained from non-uniformly sampled experiments must be reconstructed using one of the many different approaches, the most popular one being maximum entropy reconstruction [25]. NUS based experiments have also been used to resolve peak overlaps in a 53 kDa monomeric protein [23].

A biochemical approach to addressing the issue of signal overlap is development of strategies to carry out site specific labeling of proteins. In the strategy pioneered by Kay and coworkers, isotopically labeled amino acid precursors are used in the cell culture medium to selectively label the methyl groups of Ile, Leu and Val.[26] The samples prepared in this strategy can be used to measure methyl-TROSY based experiments which almost completely eliminates the signal overlap problem. This approach is very useful for two primary reasons - (i) they are often found in hydrophobic core of the proteins and (ii) methyl groups are very favorable for use with TROSY experiments and hence large systems [27]. Another approach has been to carry out amino acid type labeling instead of methyl groups. This approach, pioneered by Prestegard and coworkers, has been demonstrated in mammalian expression systems and used in the study of a 36 kDa (without glycans) glycoprotein, ST6Gal1 [28].

2.1.1.2 Protein Dynamics

Despite the large number of structures available in the PDB, understanding the functions of the proteins at a mechanistic level necessitates understanding of protein dynamics. In enzyme catalysis, for example, binding of ligands often require conformational changes, which can be rate-limiting, as was shown in the case of dihydrofolate reductase (DHFR) [29]. Dynamics in proteins span different time scales depending on the type of motions. While vibrations happen in sub-picosecond time scale, backbone and side chain fluctuations are picosecond(ps)-nanosecond(ns) events. Comparatively, slower events like conformational rearrangements and global (un)folding events happen in millisecond and second time scale respectively [30]. NMR spectroscopy has been used extensively to probe the ps-ns relaxation events and conformational exchange.

Backbone and side chain fluctuations are typically probed by measuring spin lattice relaxation rate (R_1), spin spin relaxation rate (R_2) and heteronuclear NOE [31]. Since reorientation of the molecules causes fluctuations of local magnetic fields, NMR observable coherences dephase. The spectral density function which is directly related to R_1, R_2 and NOE can be analyzed using the Model-Free formalism proposed by Lipari and Szabo [32, 33].

$$J(\omega) = \frac{S^2\tau_m}{1 + \omega^2{\tau_m}^2} + \frac{(1 - S^2)\tau}{1 + \omega^2\tau^2} \tag{2.5}$$

Here, τ_m is the correlation time of the protein and S is the order parameter, which is a measure of fluctuations of the measured bond vector. A value of 1.0 for S^2 indicates a rigid structure and a value of 0 indicates unrestricted internal motion. Further,

$$\frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e} \tag{2.6}$$

where τ_e is the timescale for the motion of the internal bond vector. Backbone dynamics are obtained by measuring the [^{1}H-^{15}N] correlation experiment to obtain the ^{15}N relaxation rates while the side chain dynamics are measured using experiments that show [^{1}H-^{13}C] correlation to extract ^{13}C relaxation rates.

Conformational exchange, which occurs at the μs-ms time scale cannot be completely described by R_2 alone, since the protein samples different conformational states and hence significantly different local environments. R_{ex}, relaxation due to conformational exchange, must also be included in the analysis. Experimentally, R_2 relaxation dispersion and $R_{1\rho}$ experiments [34] can be used to obtain information on dynamics in this time scale. When more than two conformations are sampled, multiple quantum coherence experiments have been proposed to probe these conformations [35]. Lineshape analysis is another approach to obtain information about events that happen in the millisecond time scale [36]. Here, a series of spectra are measured with variations in a single experimental parameter such as temperature, pH, etc. The effect of these variations on the NMR observables are then fitted to appropriate models and information about the dynamics of the can be extracted [37].

2.2 Hyperpolarization

Hyperpolarization is the process of generating non-equilibrium population of different spin states. It can be seen from equation 2.4 that a decrease in operational temperature can lead to increased polarization of the nuclear spins. However, the required low temperatures are not readily achieved and so a number of alternate approaches have been taken to generate highly polarized spin systems including optical pumping, parahydrogen induced polarization (PHIP), chemically induced polarization and dynamic nuclear polarization.

2.2.1 Optical Pumping

Spin exchange optical pumping has been used to hyperpolarize nuclei of noble gases. In this technique, circularly polarized light is used to excite electronic transitions of vapors of alkali-metal atoms in a chamber also containing an excess of noble gas atoms. When collisions occur between the alkali metal and noble gas atoms, high electron spin polarization created by the optical pumping is transferred from alkali metal atoms to the nuclei of noble gas atoms. Nitrogen gas is also used in the chamber to quench the excited atoms [38]. Systems capable of producing polarization levels (^{3}He and ^{129}Xe) as high as 64% have been developed [39].

Hyperpolarized ^{3}He and ^{129}Xe have been proposed in magnetic resonance imaging of the lung [40, 39]. Another interesting application in this area has been the use of hyperpolarized xenon as a sensor to detect protein-protein interactions [41, 42]. Biosensors that bind to hyperpolarized xenon have been developed and used to detect protein-ligand interactions using human carbonic anyhydrase as a model [43].

2.2.2 Parahydrogen Induced Polarization

Hydrogen exists as two different spin isomeric forms - ortho and para - where the two proton spins are aligned parallel and anti-parallel to each other respectively. These spin isomers differ in their nuclear spin configurations, electrical conductivities and heat capacities. Under standard conditions, the ratio between is orthohydrogen and parahydrogen is 3:1. However, 1:1 ratio of the two isomers and 100% parahydrogen can be generated at liquid nitrogen and liquid helium temperatures respectively, in the presence of paramagnetic catalysts.

The generated parahydrogen is not suited for direct NMR measurement since it has a nuclear spin of zero. However, a hydrogenation reaction using parahydrogen has been shown to provide signal enhancement of several orders of magnitude [44, 45, 46]. Another approach in using parahydrogen is transferring the polarization from hydrogen after the reaction to another nucleus (e.g. ^{13}C, ^{19}F) [47, 48] using coherence transfer NMR pulse sequences such as INEPT [47] and measure a ^{13}C spectrum. This has the added advantage that the polarization is preserved for a longer duration since ^{13}C spin lattice relaxation times are, in general, longer than those of ^{1}H.

In a seminal work, Signal Amplification By Reversible Exchange (SABRE) was demonstrated as a technique to provide signal enhancement using para-hydrogen species even in cases where a hydrogenation reaction is not possible [49, 50]. It was shown in this work that, transient association of para-hydrogen to a substrate of interest via a transition metal complex can transfer the polarization to the substrate at low magnetic fields. Enhancements of greater than 500-fold were reported for ^{1}H, ^{13}C and ^{15}N [49].

Parahydrogen induced polarization (PHIP) has been used, for example, to identity intermediates of various hydrogenation reactions in organic synthesis [51, 52], to carry out ^{13}C imaging of succinate in brain cancer [53] and using SABRE to detect trace amount of drugs used to treat diseases like Parkinson's disease [54].

2.2.3 Chemically Induced Dynamic Nuclear Polarization

Chemically Induced Dynamic Nuclear Polarization, abbreviated as CIDNP, refers to the hyperpolarization technique where non-Boltzmann nuclear spin populations are generated using a chemical reaction. Most modern experimental setup involves a pulsed laser to trigger the photochemical reaction that generates the radical pair and after a short delay radiofrquency pulse is applied to measure NMR signal.

The phenomenon of CIDNP can be explained by the radical pair mechanism. CIDNP effect is commonly generated using a cyclic reaction of photo-excited dye with an analyte under study. Here, the dye is photoexcited to a single state which then undergoes intersystem crossing to a triplet state. Molecules in triplet state accept an electron from a donor and a radical pair is generated. This radical pair then undergoes intersystem crossing to a singlet state induced by the hyperfine interactions with the nuclear spins. Since the rate of intersystem crossing depends on the nuclear spin state, this step, in effect, acts as a spin sorting mechanism. As a result, the recombination products show enhancement of a particular nuclear spin state while escape products show enhancement of the other nuclear spin state [55]. This setup offers the capability to carry out time resolved NMR measurements.

CIDNP was employed to following protein folding of lysozyme using a setup that involved flavin mononucleotide (FMN) as the photosensitizer. Protein folding was monitored by observing the evolution of tyrosine and trytophan resonances in the spectra [56]. CIDNP has also been used to differentiate surface exposed and buried residues [57] and in structure determination in combination with nuclear Overhauser effect (NOE) experiments [58].

2.2.4 Dynamic Nuclear Polarization

Arguably, the most generic of hyperpolarization techniques is dynamic nuclear polarization (DNP), which does not rely on a specific reaction or molecular characteristics and can polarize many different nuclei [59]. This hyperpolarization phenomenon was first described theoretically by Albert Overhauser [60] and was experimentally confirmed by Charles Slichter and coworkers [61, 62]. DNP has found renewed interest as a signal enhancement technique for NMR over the last two decades predominantly due to the improvements in hardware [63].

2.2.4.1 Overhauser Effect

The only mechanism for DNP in solution is the Overhauser effect [60]. The Overhauser effect involves microwave irradiation of samples containing the analyte of interest and a source of unpaired electrons. This irradiation saturates the electron spin and transfers the polarization to nuclear spins because of time-dependent dipolar coupling between nuclear spins and unpaired electrons [64]. The first experimental validation of this effect was accomplished by using samples of metallic Li, metallic Na and solution of Na in anhydrous liquid ammonia at room temperature [61, 62]. It was subsequently shown that the Overhauser effect was not limited to metals but will also exist in free radicals [65]. Current experiments take advantage of this property and use stable radicals added to the samples which provide the unpaired electrons for the DNP process and NMR measurements are carried out at ambient temperatures [66].

Overhauser DNP has been used to monitor protein aggregation by hyperpolarizing water and using water protein interactions to probe specific sites (sites that have nitroxide labels) on the protein [67]. This helps to distinguish random aggregation from ordered packing. However, the drawback in using Overhauser DNP is heating up of the sample since solvents such as water absorb microwave radiation. Since saturating electron resonances at high magnetic fields require high power microwave radiation, Overhauser DNP experiments are often limited to low magnetic fields. Utilizing the Overhauser effect at higher magnetic fields is an active area of research and a detailed study is available in the literature [66].

2.2.4.2 Solid Effect

The Solid effect is a DNP mechanism where the electron-nuclear interactions are time-independent. As with Overhauser effect, this mechanism also involves irradiating the samples with microwave at frequencies ($\omega_{\mu w}$) corresponding to sum or difference of electron (ω_{e^-}) and nuclear (ω_{nuc}) Larmor frequencies

$$\omega_{\mu w} = \omega_{e^-} \pm \omega_{nuc} \tag{2.7}$$

This irradiation induces transitions, where the both the nuclear and electron spins flip due to hyperfine interactions between them. Although the probability of these forbidden transitions (zero- or double quantum transitions) is small in comparison with those of allowed transition, the presence of these transitions results in the solid effect. This process while slower than the spin lattice relaxation of electron is much faster than spin lattice relaxation of the nuclear spins, resulting in higher than equilibrium nuclear polarization. It has also been shown that enhancement due to the solid effect scales as B_0^{-2} and so at higher fields becomes less efficient [68].

2.2.4.3 Cross Effect

Another effect observed in solid state DNP is the cross effect, which occurs only in the presence of two unpaired electrons [69]. The interactions between the nuclear spins and electron spins that exhibit dipolar coupling between them causes the cross effect. For most efficient polarization transfer from electrons to nuclear spins, the condition shown in equation 2.8 must be met.

$$|\omega_{0e1} - \omega_{0e2}| = \omega_{0ni} \tag{2.8}$$

Here, ω_{0e1} and ω_{0e2} are the Larmor frequencies of the coupled electrons and ω_{0ni} Larmor frequency of the nucleus. For the electron-electron dipolar coupling to be efficient, optimal distance between the electrons should be maintained. Although DNP via cross effect using common radicals such as TEMPO has been reported [70], only a fraction of the radicals in the sample would be expected to satisfy the condition in equation 2.8 since the distance between the radicals and their relative orientation would be random. To achieve higher enhancement using cross effect, biradicals such as bis-TEMPO-n-ethylene glycol [71] and TOTAPOL have been developed [72]. Enhancements of 100–300 have been reported using these biradicals [73, 74].

2.2.4.4 Thermal Mixing

Thermal mixing differs from solid and cross effect in that the interaction between a system of unpaired electrons and nuclear spins gives rise to the polarization of nuclear spins. Strong electron-electron interactions generate a spin ensemble which first undergoes flip-flop transitions. These transitions can be further enhanced by irradiating near the electron Larmor frequency. This then gives rise higher nuclear spin polarization because of the interactions between the “system of electrons” and nuclear spins. Since, in thermal mixing, only allowed transitions happen, the transitions have a greater probability of happening and hence likely to be very efficient [75, 76].

2.2.4.5 Case Studies Involving DNP Enhanced Solid State NMR

Many groups have used DNP enhanced solid state NMR to study proteins and other biological preparations. Griffin and coworkers demonstrated that signal enhancements of 50 were obtained in experiments where ^{15}N-Alanine labeled T4 lysozyme samples were hyperpolarized and ^{15}N solid state NMR spectra were measured at 40 K [77]. In another application, spectra of nanocrystals of myloidogenic peptide, GNNQQNY, were measured using DNP enhanced Cross Polarization/Magic Angle Spinning (CP/MAS) experiments [78]. Here, the samples of nanocrystals were hyperpolarized on ^{1}H nuclei and the polarization was transferred to ^{13}C and ^{15}N for solid state NMR measurement using cross polarization. Enhancements of more than 100 times compared to non-DNP experiments were reported in these experiments. It is of particular interest that with the help of the enhancement obtained from the DNP process even two dimensional Dipolar Assisted Rotational Resonance/RF assisted diffusion (DARR/RAD) experiments were measured, which would otherwise not have been possible.

Signal enhancements provided by DNP, experiments have enabled studies using pelleted ribosome complexes (from *Thermus thermophilus*) by Bodenhausen and coworkers

[79]. ^{13}C–^{13}C and ^{13}C–^{15}N correlation experiments following a CP step to transfer ^{1}H to ^{13}C and ^{15}N have been demonstrated to yield sufficient signal to carry out characterization of the sample and study protein-protein interactions even in a 800 kDa ribosome complex.

Furthermore, Reif and coworkers have demonstrated the use of DNP enhanced solid state NMR spectroscopy to characterize over-expressed membrane proteins by isolating membrane fractions containing the proteins from the cell lysates [80]. In a similar strategy, Baldus and coworkers have demonstrated measurements of solid state NMR spectra of whole cells (*E. coli*) where the proteins of interest are membrane targeted and DNP enhanced two dimensional heteronuclear correlation experiments were carried out to obtain chemical shift assignments of selected resonances. [81].

2.2.4.6 Dissolution Dynamic Nuclear Polarization

It can be seen from a representative compilation above that DNP improves the range of applications that are accessible to NMR spectroscopy. Solution NMR spectroscopy offers greater spectral resolution than solid state NMR since the anisotropic interactions that contribute to line broadening are averaged out in solution due to molecular tumbling. Therefore, combining a generic signal enhancement technique such as DNP with solution NMR is likely to be beneficial. Pioneered by Ardenkjær-Larsen and coworkers [82], dissolution DNP combines signal enhancement obtained from the DNP process in the solid state with NMR measurement from samples in the solution state. A schematic representing a typical dissolution DNP setup is in Figure 2.1. In the dissolution DNP setup, hypeprolaroization of the samples take place at ~1 K. A plot showing comparison of electron, ^{1}H and ^{13}C polarization in the temperature range of 0.1 mK to 100 K is shown in Figure 2.2.

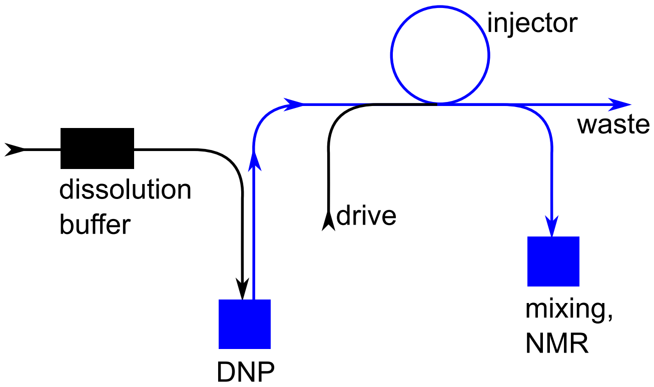


Figure 2.1: Schematic representation of a dissolution DNP setup. Dissolution buffer is a superheated solvent and is used to dissolve the frozen sample in the polarizer which is loaded on the injector loop. The sample is transferred to the NMR tube using a gas drive followed by NMR measurement (triggered automatically).

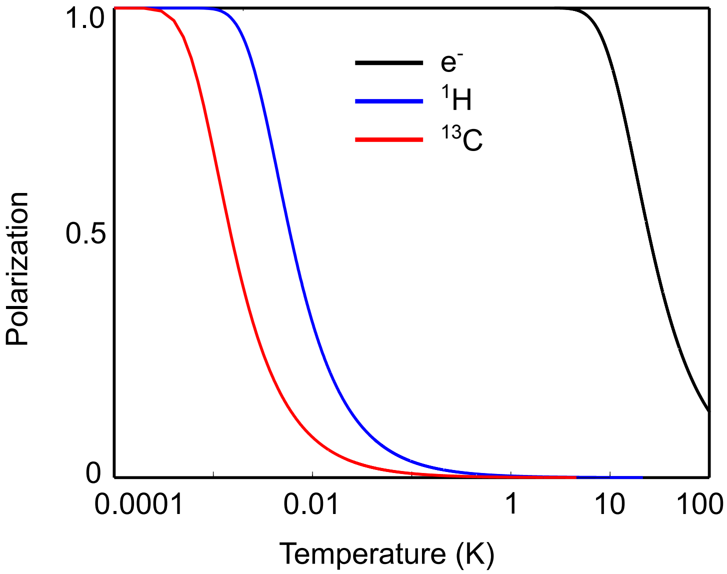


Figure 2.2: Plot showing polarization levels of electron, ^{1}H and ^{13}C at a temperature range of 0.1 mK to 100 K. External magnetic field used in the calculation is 3.35 T. Plotted after [82].

The main drawback in polarizing samples *ex situ* is that there is a significant loss of spin polarization during the sample transfer step. Strategies that have been pursued to minimize losses during the sample transfer include the development of rapid injection devices in our group at Texas A&M [83] and development of dual isocenter magnets by Köckenberger and coworkers [84]. In the latter approach, spin hyperpolarization is carried out at a lower magnetic field (3.35 T) in an upper compartment, while NMR measurement is carried out at 9.4 T, the lower compartment. This setup greatly reduces the time taken for sample transfer, thereby presenting greater polarization of spins at the start of NMR measurement.

Perhaps the greatest advantage in using dissolution DNP is that a wide range of transient processes, including chemical reactions, can be studied. While a large majority of NMR experiments have been carried out under steady state conditions, adding a time dimension becomes practical by combining dissolution DNP with NMR spectroscopy, which is discussed in the next section.

2.3 Time-resolved NMR Spectroscopy

Although NMR spectroscopy has been generally used for steady state measurements, studies have been performed that have provided kinetic information. Approaches to record time resolved NMR spectra including interfacing stopped-flow apparatus to inject samples into a sample tube placed in an NMR spectrometer, design of pulse sequences for rapid data acquisition and use of hyperpolarization techniques like DNP are discussed below.

2.3.1 Stopped Flow NMR

A number of studies have utilized stopped flow instrumentation, akin to that used in optical spectroscopy techniques, in conjunction with NMR. One of the first stopped flow NMR experiments was the investigation of kinetic deuterium isotope effect in the hydrolysis of methyl formate by Ernst and coworkers [85]. In the work by Hoeltzli and Frieden, 6-

^{19}F-Tryptophan labeled dihydrofolate reductase (DHFR) was recombinantly expressed and unfolding of the protein (0.9 mM final concentration) was monitored as a function of time after mixing using ^{19}F NMR [86]. Similar experimental setups have been used to study reactions such as chymotrypsin catalyzed hydrolysis of tertiary butly-L-phenylalanine [87] and polymerization of 1-hexene [88].

Another approach to gain temporal resolution has been to use NMR-laser coupled setup. In a work by Schwalbe and coworkers, RNA folding was monitored by NMR by triggering a photoreaction, which removes the photolabile protecting group around the RNA. Removal of this group facilitated ligand binding to RNA and initiated folding, which was then monitored using NMR [89]. In an alternative approach to stopped flow setup, Sweedler and coworkers reported a flow based system based on a microfluidic chip to measure time resolved NMR spectra of ubiquitin corresponding to different conformational states [90].

2.3.2 Rapid Acquisition of Multidimensional NMR Spectra

Multidimensional NMR experiments show correlation between different spin types and provide sufficient resolution to identify individual sites in a macromolecule. These experiments typically employ increments of evolution delay for gaining resolution in the indirect (second) dimension. This approach results in long experiment times, which makes it useful for studies at equilibrium but not transient processes. Rapid measurements of equivalent data would aid in observing non-equilibrium processes.

Several approaches towards acquiring multidimensional NMR spectra within a few seconds have been demonstrated. Single scan acquisition of 2D NMR spectra was demonstrated with ultrafast TOCSY [91] and HSQC [92] experiments by Frydman and coworkers. An alternative strategy to acquire multidimensional spectra was presented in the form of band-Selective Optimized Flip-Angle Short-Transient heternuclear multiple quantum coherence (SOFAST-HMQC) [93] experiment, which has been used to measure protein [^{1}H-^{15}N] correlation spectra in a matter of seconds. Hydrogen/Deuterium exchange experiments with temporal resolution were possible with this scheme [94]. Additionally, Hadamard frequency encoding coupled with SOFAST-HMQC further improves the time resolution [95].

2.3.3 Dissolution DNP Enhanced NMR

Most of the stopped flow NMR experiments measure ^{1}H nuclei which, while sensitive, yield crowded spectra. Alternatively, labeling schemes involving introduction of NMR active nuclei, such as ^{19}F might have to be undertaken. In the case of rapid acquisition schemes, very high sample concentrations are needed. These restrictions, unfortunately, may not be readily overcome in all cases. Dissolution DNP, on the other hand, can be used to polarize different nuclei and measure NMR spectra. The dissolution DNP enhanced NMR spectrum of [U - ^{2}H] leucine is shown in Figure 2.3 and the enhancements obtained in comparison with non-hyperpolarized experiments are presented in Table 2.2.

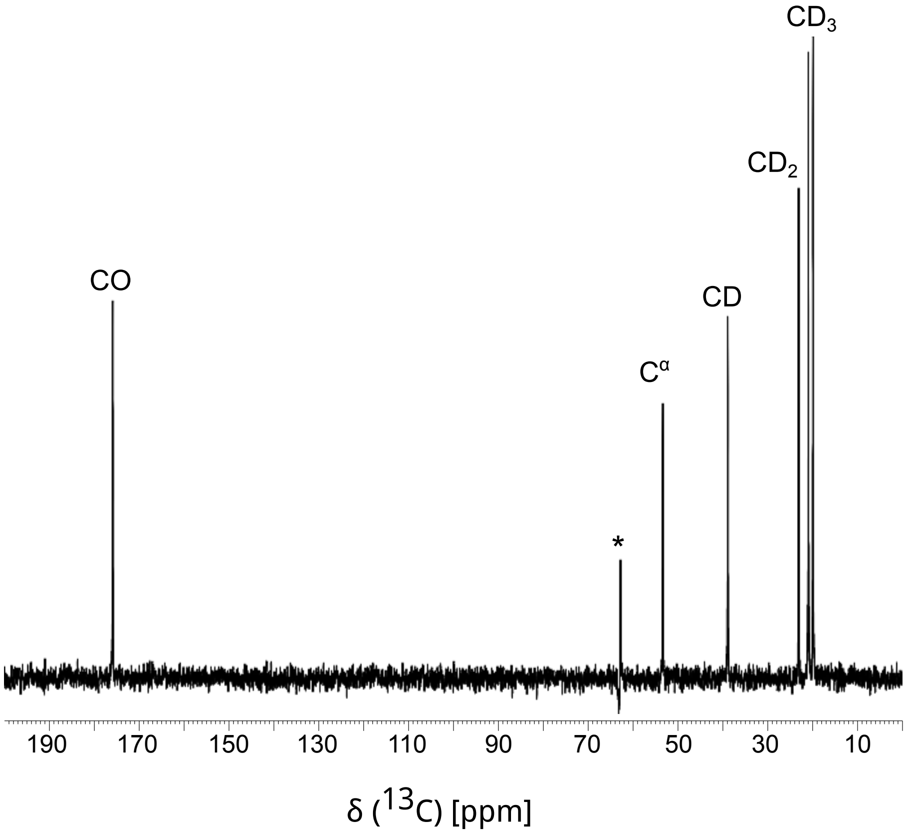


Figure 2.3: Spectrum of ^{2}H–Leucine hyperpolarized on ^{13}C nuclei by irradiating with with 60 mW microwaves (93.974 GHz) at 1.4 K for 3 hours. The hyperpolarized sample was dissolved using a hot stream of 50 mM potassium phosphate, pH 6.8 and NMR spectrum was measured with ^{2}H decoupling. Enhancements for individual sites on the amino acid are shown in Table 2.2. The peak indicated by * corresponds to ethylene glycol, which is suppressed.

Table 2.2: Signal enhancement of different sites on Leucine obtained from a dissolution DNP experiment as compared to a non-hyperpolarized measurement

Group	Enhancement
CO	1645
C^α	920
CD_2	1750
CD	2000
CD_3	2500
CD_3	2400

Dissolution DNP enhanced NMR has been used to visualize metabolism in live yeast [96], develop assays for drug metabolism [97], observe metabolic patterns in cancer cells [98] and perfused liver [99] and in *in vivo* assays of transaminase activity [100]. In our group, dissolution DNP has been used to monitor Diels-Alder reaction and estimate the reaction rate, [101], to identify biosynthetic pathways, [102], to investigate ligand binding, [103, 104] and detect intermediates in polymerization reactions [105].

Although DNP enhanced NMR of proteins has been performed in the solid state at low temperatures, [106, 78, 77] such experiments requiring the NMR measurement of signals from proteins in the solution state have been challenging. The work in this thesis reports on the application of dissolution DNP for direct observation of full length proteins in solution state.

3. SOLUTION NMR OF POLYPEPTIDES HYPERPOLARIZED BY DYNAMIC NUCLEAR POLARIZATION[*]

3.1 Introduction

Since proteins have limited solubility, measurement of NMR spectra often requires extended signal averaging resulting in long experiment times. Although this requirement does not impede measurements under equilibrium conditions, such as of multidimensional spectra for protein structure determination, processes of transient nature are more difficult to observe. Application of dissolution dynamic nuclear polarization to larger molecules has so far proven more challenging, largely because the intensity of the observable signal depends on the spin–lattice relaxation during transfer of the hyperpolarized sample. Current investigations involving full length proteins using this method have involved indirect observation of proteins using hyperpolarized small molecules that act as ligands or substrates to proteins [107, 97, 99, 103, 104]. Here, a report on high-resolution dissolution DNP-NMR, where hyperpolarized polypeptides are directly observed, is presented. The experimental conditions and signal enhancements obtained, as well as the applicability of this method for the study of various biochemical processes are discussed.

3.2 Materials and Methods

3.2.1 Protein Expression and Purification

The samples of L23 protein were prepared by expression in *Escherichia coli* BL21(DE3) cells harboring a pET11c vector, which contains the gene for L23 from *Thermus thermophilus*. The complete composition of minimal medium (adapted from [108, 109]) is

---

[*] Reprinted (adapted) with permission from “Solution NMR of Polypeptides Hyperpolarized by Dynamic Nuclear Polarization, Mukundan Ragavan, Hsueh-Ying Chen, Giridhar Sekar and Christian Hilty, Analytical Chemistry, 83, 6054–6059, 2011”. Copyright (2011) Americal Chemical Society.

given in Tables 3.1 & 3.2. Fresh transformation of bacteria with the plasmid was always used since no protein expression was observed in cultures started from glycerol stocks stored at -80°C. A 30–100 mL cell culture was started from a LB agar plate in LB growth medium (37°C; carbenicillin at a final concentration of 100 μg/mL) and allowed to grow to mid-log phase (OD_{600} = 0.6). This culture was centrifuged at 4000×g and 4°C for 10 minutes. The supernatant was discarded and the cell pellet was resuspended in minimal medium (with isotope labeling as needed; carbenicillin to a final concentration of 100 μg/mL) and grown at 37°C. When the OD_{600} of cell culture reaches 0.6, protein expression was triggered by adding isopropyl β-D-1-thiogalactopyranoside (IPTG), to a final concentration of 1 mM. Cells were grown for 20 hours after induction in minimal medium, or for <12 hours in LB medium. Cells were harvested by centrifugation at 4000×g and 4°C for 20 minutes and stored at -80°C until protein purification.

The purification protocol was adapted from the procedures reported [110]. Harvested cell pellets were resuspended in 50 mM Tris, pH 8.0 (buffer A) and cell lysis was accomplished by sonication. Sonication pulses were applied at 60% duty cycle for 2 minutes followed by 10 minutes of cooling period. This cycle was repeated for a total of 5 times (10 minutes of sonication). The cell lysate was spun down in a centrifuge in 50 mL conical tubes at 10,000×g for 45 minutes at 4°C. Supernatant from the tube was decanted and was incubated at 50°C for 1 hour. This incubation step precipitates most of the proteins in the cell lysate, while L23 stays in solution. The supernatant should turn cloudy during the course of incubation. The tubes containing the incubated cell lysate were then centrifuged as described earlier and the supernatant was loaded on a cation exchange column, HiTrap SP XL (GE Healthcare Life Sciences, Pittsburgh, PA). A gradient of 0 – 100% buffer B was used to elute the protein where buffer B is 50 mM Tris, pH 8.0 containing 1 M NaCl. L23 elutes out between 65–80% buffer B over a total volume of ~15 mL. The purity of the sample was assessed using SDS–PAGE and subsequently dialyzed against water at 4°C for 12–16 hours using a 1000 MWCO semi-permeable membrane. This sample was then lyophilized to get L23 in powder form. Typical protein yield from 1 liter cell culture was 10 –15 mg (for both LB and minimal medium).

Table 3.1: Composition of Minimal medium used in this study [108, 109]

Component	Amount/Volume
$Na_2HPO_4 \cdot 7H_2O$	12.8 g
KH_2PO_4	3.0 g
NaCl	0.5 g
^{13}C-D-Glucose	3.0 g
NH_4Cl	1.0 g
Solution Q	1 mL
1 M $MgSO_4$	2 mL
Vitamin Mix	10 mL
H_2O/D_2O	Volume to 1000 mL

Table 3.2: Composition of Solution Q and Vitamin Mix used in the preparation of minimal medium

Solution Q		Vitamin Mix	
Component	Amount/Volume	Component	Amount/Volume
HCl	8 mL	Thiamine	50 mg
$FeCl_2 \cdot 2H_2O$	5.0 g	D–Biotin	10 mg
$CaCl_2 \cdot 2H_2O$	184 mg	Choline chloride	10 mg
H_3BO_3	64 mg	Folic acid	10 mg
$CoCl_2 \cdot 6H_2O$	18 mg	Niacinamide	10 mg
$CuCl_2 \cdot 2H_2O$	4 mg	D–Pantothenic acid	10 mg
$ZnCl_2$	340 mg	Pyridoxal	10 mg
$Na_2MoO_4 \cdot 2H_2O$	605 mg	Riboflavin	1 mg
$MnCl_2 \cdot 4H_2O$	40 mg		
H_2O	Volume to 1000 mL	H_2O	100 mL

3.2.2 Sample Preparation

Samples for DNP hyperpolarization were prepared by dissolving the protein to a concentration of 2.1 mM in a mixture of 60% (v/v) ethylene glycol/D_2O containing 7 M urea. Urea was omitted from the mixture when the protein was polarized in native conditions. The protein sample was then mixed with 15 mM tris[8-carboxyl-2,2,6,6-tetra[2-(1-hydroxymethyl)]-benzo(1,2-d:4,5-d')bis(1,3)dithiole-4-yl]methyl sodium salt(OX63; Oxford Instruments, U.K; Figure 3.1 (a)). Samples for conventional NMR spectroscopy were prepared by dissolving L23 in 50 mM potassium phosphate buffer at pH 6.8 to a concentration of 0.4 mM. Ten percent (v/v) D_2O was added as the lock solvent.

Bacitracin A (Sigma-Aldrich, St. Louis, MO) was purified using a desalting column

(GE Life Sciences, Piscataway, NJ) and treated with Chelex-100 resin (Sigma-Aldrich, St. Louis, MO) to remove metal ions. The sample for polarization was prepared by dissolving 7.5 mM purified peptide in 60% (v/v) ethylene glycol/D_2O. The peptide sample was then mixed with 4-hydroxy-2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPOL) free radical (Sigma-Aldrich, St. Louis, MO; Figure 3.1 (b)) to a final concentration of 15 mM.

3.2.3 Dynamic Nuclear Polarization and NMR Spectroscopy

L23 samples were hyperpolarized in a Hypersense DNP polarizer (Oxford Instruments, Tubney Woods, U.K.) by irradiation with 60 mW of microwaves at a frequency of 93.974 GHz for 4 h at 1.4 K for ^{13}C hyperpolarization. The samples were polarized for 7 h in the experiment performed to measure the spin–lattice relaxation. Subsequently, samples were dissolved by a stream of hot solvent consisting of 60% (v/v) acetonitrile/50 mM potassium phosphate (pH 3.1) (or 40% (v/v) methanol/50 mM potassium phosphate (pH 3.1)). The resulting liquid was automatically injected into a 400 MHz NMR spectrometer (Bruker, Billerica, MA) using the rapid sample injection device described [83]. Total time elapsed between dissolution and start of the NMR experiment was 1620 ms. NMR spectra were acquired using a pulse sequence $[Ps–Gx–Ps–Gz]_3$–P90–Acq. Here, Ps stands for a selective 90 pulse (EBURP shape, 15 ms duration) applied at the solvent chemical shift of 63.7 ppm for ethylene glycol or 163.1 ppm for urea (where necessary). Gx, Gy, and Gz are pulsed field gradients of 50 G/cm and 1 ms duration. P90 is a hard pulse, and Acq designates signal acquisition at 298 K (128K complex points, spectral width of 251 ppm). During acquisition, ^{1}H and ^{2}H decoupling was applied using WALTZ-16 at field strengths of 2.2 and 0.3 kHz, respectively.

Bacitracin samples were hyperpolarized in the same system by irradiating with 100 mW of microwaves at a frequency of 94.005 GHz for 30 min at 1.4 K for ^{1}H polarization. The sample was then dissolved using methanol and injected as described above. NMR spectra were acquired using the pulse sequence [Ps–Gx–Ps–Gy–Ps–Gz]$_3$–P90–Acq at 298 K (32K complex points, spectral width of 20 ppm; Ps: EBURP shape, 25 ms). Methanol, ethylene glycol, and water were suppressed at 3.4, 3.15, and 4.76 ppm. Chemical shifts in the samples of L23 were referenced indirectly against standard samples containing 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) in the same solvent, via the chemical shifts of solvent resonances. Chemical shifts in the bacitracin A samples were referenced indirectly against tetramethylsilane (TMS) in the same solvent.

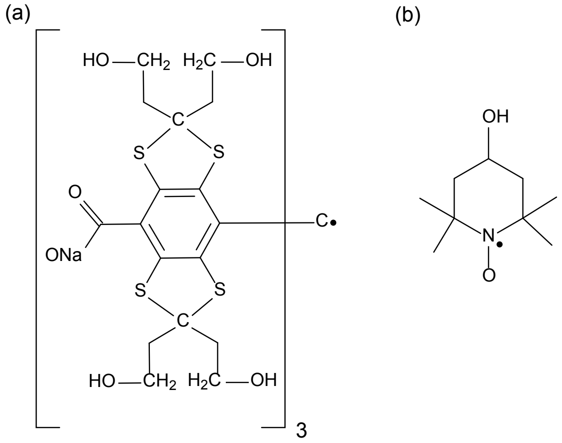


Figure 3.1: Chemical structures of radicals (a) OX–63 and (b) 4-hydroxy-TEMPO (TEMPOL) used to hyperpolarize ^{13}C and ^{1}H nuclei, respectively.

3.2.4 Calculation of Signal Enhancement

The ^{13}C signal enhancement reported is the ratio of the integral of the specified peak or spectral region in the polarized spectrum to the integral of the corresponding peak or region in a conventional NMR spectrum acquired with a repetition delay larger than $5 \times T_1$. Integrals were normalized to concentration and to number of scans.

3.2.5 Measurement of ^{13}C Spin-Lattice Relaxation Time of the Hyperpolarized Protein

In order to measure the spin relaxation times of the hyperpolarized samples, ^{13}C spectra of the protein sample were measured using variable flip angle according to the scheme described with 32K complex points and spectral width of 251 ppm [111]. The experiment was carried out using ^{1}H decoupling to enable suppression of solvent signals. Conventional ^{13}C NMR spectra were acquired using a 500 MHz NMR spectrometer (Bruker, Billerica, MA).

3.2.6 Data Analysis

A ^{13}C spectrum of L23 was computed from the reported random coil chemical shifts, [112] by superimposing Lorentzian functions with 20 Hz line width corresponding to the amino acid content of L23 using MATLAB (MathWorks, Natick, MA). Spin relaxation parameters were determined by fit to single exponential.

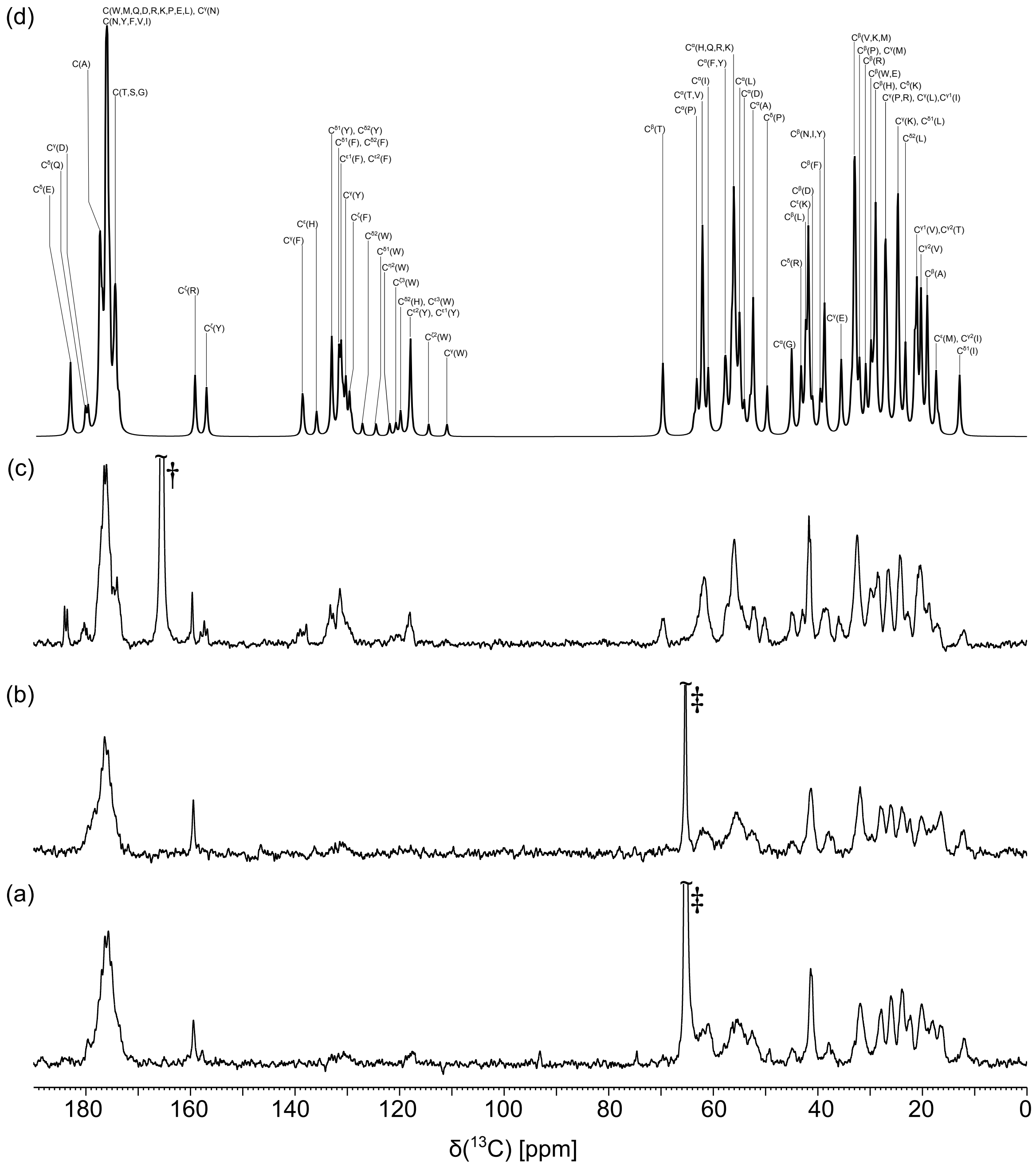
3.3 Results and Discussion

3.3.1 ^{13}C Hyperpolarization of Polypeptides

An important consideration for measuring a dissolution dynamic nuclear polarization spectrum is the sample conditions to be used. In general, two solvent systems are needed: one for DNP in the solid state, and a second for dissolution of the sample and transfer to the NMR instrument. A mixture of 60% (v/v) ethylene glycol/D_2O was the glass-forming mixture used in all experiments (see the Materials and Methods). Despite the high concentration of urea (where used), this mixture forms a glass upon freezing, which is required for sufficiently close contact of the radical with the sample species, allowing for DNP [113]. The typical solubility of polypeptides in this mixture was found to be on the order of few millimoles per liter. The solvent used for dissolution determines whether it is possible to transfer the hyperpolarized sample while preserving the spin polarization. Since the present sample transfer system involves the use of pressurized gases, protein samples have a propensity to foam. In order to overcome these problems, various dissolution solvents were explored. Specifically mixtures of organic solvents with water were considered since they have a lower surface tension than water [114, 115]. Protein samples can be suitably dissolved without foaming in a mixture of acetonitrile, or methanol, and water with a transfer time of 1.6 s.

[U-^{13}C, 50%-^{2}H]-labeled protein samples were polarized in native and in denaturing conditions (7 M urea). In both cases, samples were dissolved in denaturing conditions, such that the NMR spectrum is of the denatured polypeptide. The spectrum of the sample originally polarized in native conditions is shown in Figure 3.2 (a), and the spectrum of the sample polarized in denaturing conditions is shown in Figure 3.2 (b). The expected ^{13}C spectrum of the protein was computed (see the Materials and Methods) in order to identify resonances in the hyperpolarized spectrum. The computed and experimental spectra agree quite well, and resonances can be identified clearly. When ^{13}C spectra obtained using hyperpolarized L23 samples are compared to thermally polarized samples of higher concentration (Figure 3.2 (a)–(c)), it is apparent that the hyperpolarized polypeptide is indeed denatured in the dissolution solvents used. The denatured state of the protein was also confirmed by using [^{1}H, ^{15}N] heteronuclear single-quantum correlation (HSQC) spectra under the same conditions, albeit with samples at higher concentration (Figure 3.3).

Figure 3.2: (a) Spectrum of L23 hyperpolarized on ^{13}C nuclei. (b) Spectrum of L23 hyperpolarized on ^{13}C nuclei in the presence of 7 M urea. (c) Conventional ^{13}C NMR spectrum of 0.4 mM samples of L23 unfolded in 8 M urea measured with 1200 scans. The protein samples were dissolved in 60% (v/v) acetonitrile/50 mM potassium phosphate pH 3.1, and final protein concentrations were determined to be 15 and 14 μM for the spectra in panels a and b, respectively. (d) Theoretically computed ^{13}C spectrum of L23. Functional groups that contribute to the individually resolved peak groups are indicated. Resonances stemming from solvent, ethylene glycol (‡), and urea (†) are indicated.


(d)
C(W,M,Q,D,R,K,P,E,L), C^γ(N)
C(N,Y,F,V,I)
C(A)
C(T,S,G)
C^γ(D)
C^δ(Q)
C^δ(E)
C^ζ(R)
C^ζ(Y)
C^δ1(Y), C^δ2(Y)
C^δ1(F), C^δ2(F)
C^ε1(F), C^ε2(F)
C^γ(Y)
C^ε(H)
C^γ(F)
C^ζ(F)
C^δ2(W)
C^δ1(W)
C^η2(W)
C^ζ3(W)
C^δ2(H), C^ε3(W)
C^ε2(Y), C^ε1(Y)
C^ζ2(W)
C^γ(W)
C^β(T)
C^α(H,Q,R,K)
C^α(F,Y)
C^α(I)
C^α(T,V)
C^α(P)
C^α(L)
C^α(D)
C^α(A)
C^δ(P)
C^β(N,I,Y)
C^β(F)
C^β(D)
C^ε(K)
C^β(L)
C^δ(R)
C^α(G)
C^γ(E)
C^β(V,K,M)
C^β(P), C^γ(M)
C^β(R)
C^β(W,E)
C^β(H), C^δ(K)
C^γ(P,R), C^γ(L),C^γ1(I)
C^γ(K), C^δ1(L)
C^δ2(L)
C^γ1(V),C^γ2(T)
C^γ2(V)
C^β(A)
C^ε(M), C^γ2(I)
C^δ1(I)
(c)
†
(b)
‡
(a)
‡
180
160
140
120
100
80
60
40
20
0
δ(13C) [ppm]

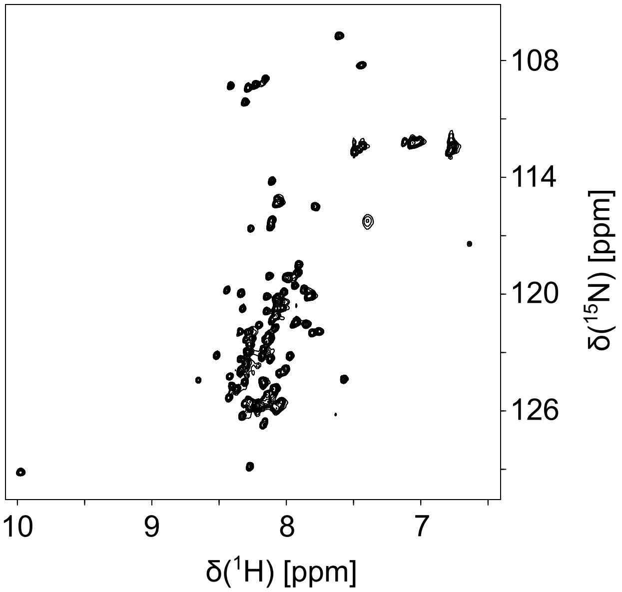


Figure 3.3: [^{15}N, ^{1}H]–HSQC spectrum of the sample [u-^{15}N, ^{1}H]–L23 under conditions similar to DNP experiments (0.4 mM L23 in 60% acetonitrile/50 mM potassium phosphate, pH 3.1, 50 mM urea). The spectrum was measured with ^{15}N decoupling on a 500 MHz NMR spectrometer equipped with a TCI cryoprobe at 298 K.

The most important spectral characteristics in the liquid-state NMR experiment are the attainable resolution and the signal enhancement. Typical line widths obtained in similar experiments with small molecules have been reported to be as low as 2 Hz, enabling the resolution of peaks with only minor chemical shift differences [83]. In the case of protein samples, the line widths are expected to be somewhat broader. In the DNP experiments, line widths the order of 15–20 Hz were obtained, compared to typical line widths obtained in the conventional experiment of 10 Hz. On the basis of these parameters, the spectra are of sufficient quality to resolve the peaks from the various functional groups in the spectrum of the denatured polypeptide.

In order to reduce loss of polarization during the transfer, the protein samples were fractionally deuterated [116]. The signal enhancement in the spectra of Figure 3.2 can be estimated for the different functional groups present in the polypeptide. In the present experiments, a typical enhancement of about 700 is observed for the carbonyl groups when compared to thermal polarization in the 400 MHz NMR spectrometer (see Materials and Methods for the calculation procedure). Substantial enhancements were also observed for aliphatic carbon atoms, with typical values on the order of 400. It is of significance that the final concentrations of L23 in the hyperpolarized measurements were only 15 μM (Figure 3.2 (a), enhancement of 750 and 500 for carbonyl and aliphatic groups, respectively) and 14 μM (Figure 3.2 (b), enhancement of 600 and 300 carbonyl and aliphatic groups, respectively), and the spectra were obtained in a single scan. In comparison, the spectrum obtained using conventional NMR was acquired using a 0.4 mM sample with 1200 scans. Obtaining an enhancement in signal-to-noise ratio of the same magnitude as the carbonyl groups in the hyperpolarized spectrum, but by conventional NMR, would require us to average 5×10^5 transients. The resonances from aromatic amino acid side chains show stronger attenuation, which is consistent with their shorter relaxation time due to lower levels of deuteration [117]. It is also to be noted that the enhancements obtained from proteins hyperpolarized in native and denaturing conditions were similar over the entire spectrum. This shows that even residues present in the hydrophobic core of the proteins appear to be accessible for DNP at 1.4 K, as was reported to be the case at 90 K [106].

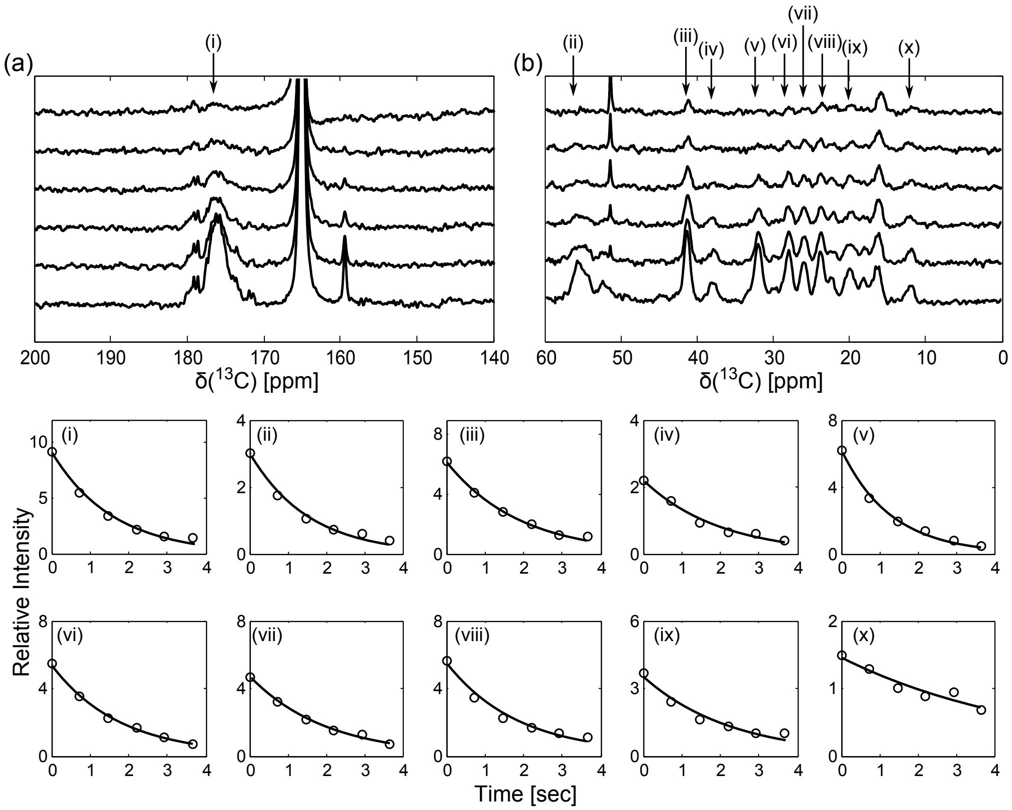


Figure 3.4: (a and b) Series of ^{13}C spectra from a single sample of hyperpolarized L23 dissolved in a mixture of 40% methanol/50 mM potassium phosphate (pH 3.1) acquired using variable flip angles (ref. [111]) with a time between scans of 733 ms. The two panels show different regions from the same spectra. (i–x) Signal decay of individually resolved resonances indicated in the spectra in panels a and b. The lines represent the functions obtained from fit of the equation for spin relaxation to the experimental data.

In the dissolution DNP experiment, the signal enhancements obtained depend both on the initial solid-state polarization and on the amount of polarization lost during the transfer of the liquid sample. Initial polarization may depend on parameters such as the type of functional group, as well as the radical and solvent system used for DNP. With the use of present instrumentation, these effects cannot be individually determined due to lack of chemical shift resolution in the solid state. The signal loss during the transfer of the sample after dissolution, on the other hand, is due to well- characterized spin-lattice relaxation processes in the liquid state. Even though the magnetic field encountered by the sample during the transfer is low, and spin–relaxation parameters are expected to differ from the high-field case [118], a comparison between the obtained enhancements and high-field spin relaxation parameters may still be of interest and is discussed below.

Spin relaxation parameters of the hyperpolarized samples were obtained from a second set of experiments. Here, a series of NMR spectra was acquired using variable flip-angle excitation from a hyperpolarized sample of L23 (Figure 3.4, parts (a) and (b)). The peaks indicated in parts (a) and (b) of Figure 3.4 were integrated. Despite a certain amount of signal overlap in the integrated regions, it was possible to fit the integrals to a single-exponential decay, yielding an apparent relaxation time τ. Values of τ, along with the observed enhancements for the different groups in the protein, are shown in Table 3.3. To calculate the enhancement of various groups, the integrals from the hyperpolarized and conventional spectra were normalized for number of transients, protein concentration in both samples, and the differences in the flip angle between the conventional spectrum and the first scan of the multiple scan experiment. From Table 3.3, it can be seen that enhancements and apparent relaxation times for most aliphatic positions are uniform, except for some methyl groups, which show longer apparent relaxation time concomitant with higher enhancement. For example, the resonance indicated as x in Figure 3.4, which arises from $C^{\delta1}$ of isoleucine, belongs to this group. The longer relaxation time is consistent with the fact that this position includes a high fraction of perdeuterated isotopomer in partially deuterated, recombinantly expressed proteins [119]. A further illustration of the effect of deuteration can be seen when resonance ix in Figure 3.4 (b) is considered. This peak, which contains contributions from $C^{\gamma1}$ of valine and $C^{\gamma2}$ of valine and threonine, has a shorter apparent relaxation time. In this case, the fraction of completely deuterated isotopomer for valine is almost zero, whereas for threonine, it would be expected to be similar to $C^{\delta1}$ of Ile.

Table 3.3: Comparison between enhancement values obtained from the experiment shown in Figure 3.4 for different groups with the high field spin relaxation parameters[a]

Peak no.	Chemical Shift Range(ppm)	Apparent Relaxation Time Constant, τ(s)[b]	Identity[c]	Enhancement
i	177.7–174.9	1.6	C	750
ii	58.2–53.9	1.6	C^{α}	550
iii	42.1–40.7	1.9	C^{δ}(R), C^{β}(L), C^{ϵ}(K)	800
iv	38.7–37.3	2.0	C^{β}(F,N,I,Y)	500
v	32.7–31.4	1.3	C^{β}(V,K,M),C^{γ}(Q)	600
vi	28.6–27.5	1.8	C^{β}(H,W,E)	800
vii	26.5–25.5	2.0	$C^{\gamma 1}$(I),C^{γ}(I),C^{γ}(P,R)	650
viii	24.3–23.3	1.9	$C^{\delta 1}$(L),C^{γ}(K)	700
ix	20.8–19.2	2.3	$C^{\gamma 2}$(V,T)$C^{\gamma 1}$(V)	500
x	13.0–11.1	5.1	$C^{\delta 1}$(I)	2000

[a] Peak numbers correspond to the labels in Figure 3.4. [b] The confidence intervals at the 95% confidence level of apparent relaxation time constants obtained from the fits range from (0.2 to (0.5 s for peak nos. i–ix. For peak no. x, the confidence interval obtained from the fit is (2 s due to the measurement time being shorter than the relaxation time.

[c] Based on comparison with the calculated spectrum.

A priori, it might be expected that the values for τ would be equal to values of spin-lattice relaxation time constants determined from conventional NMR experiments. This has been found to be true in most cases, except for those chemical shifts that include important contributions from multiple chemical species, such as $C^{\delta 1}$ of isoleucine, which contains significant fractions of CH_2D, CHD_2, and CD_3 isotopomers. The spin–lattice relaxation time computed from conventional NMR measurement for this position is shorter (1.5 s), because the protonated components carry more intensity in the conventional measurement.

3.3.2 Unlabeled Polypeptides

At present, typical protein structure determinations are made using triple-resonance experiments with recombinantly expressed, isotopically enriched samples. However, for smaller peptides, which often can most easily be prepared by solid-phase synthesis, or are extracted at low yield from natural source, homonuclear ^1H NMR is still a method of choice. Therefore, it may also be of interest to determine the potential of dissolution DNP for ^1H NMR of polypeptides. A spectrum of hyperpolarized bacitracin A peptide is shown in Figure 3.5 (a). When compared to the conventional NMR spectrum (Figure 3.5 (b)), it is apparent that most of the resonances are still present in the hyperpolarized spectrum. The enhancement of different protons was calculated by comparing to the conventional spectrum of the same sample. The enhancements in the regions of amide and aliphatic protons were 30 and 45, respectively. The aromatic region of the spectrum showed significantly higher enhancement of up to 180. This difference may be due to shorter relaxation times for aliphatic protons, as well as exchange of amide protons with the solvent. In part due to the shorter relaxation time of ^1H in bacitracin A, the enhancement is lower than that for ^{13}C-polarized L23.

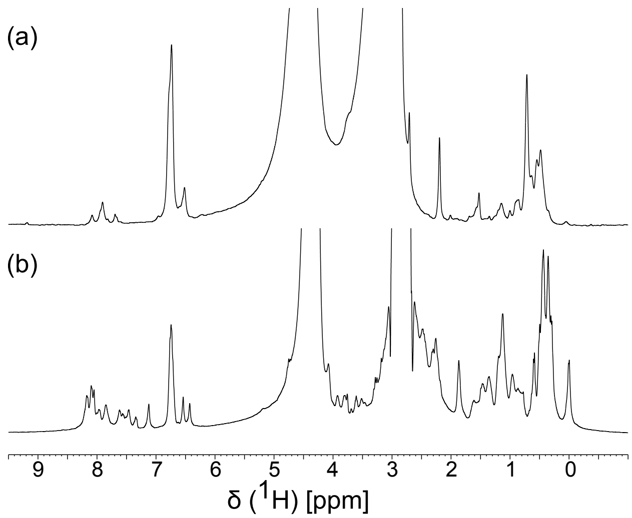


Figure 3.5: (a) Spectrum of bacitracin hyperpolarized on ^{1}H nuclei and dissolved in methanol. The final sample concentration was 37 μM. Suppressed solvent peaks from water and DMSO are observed at 4.7 ppm and 3.4 ppm respectively. (b) Conventional ^{1}H NMR spectrum of a 9 mM solution of bacitracin with 64 scans.

3.4 Conclusions

In summary, we have demonstrated that dissolution DNP is amenable toward obtaining NMR spectra of unfolded polypeptides, despite the short spinlattice relaxation times in proteins. Here, biosynthetic incorporation of ^{13}C and ^{2}H was used to increase signal and decrease spin relaxation rates. In ^{13}C spectra of the unfolded polypeptides, the various different chemical shifts of the functional groups in the molecule could readily be identified. On the basis of these results, DNP–NMR opens up new possibilities for studying transient processes, such as interactions involving antibiotic or other biologically active peptides and the process of protein folding with high time resolution and without the need for signal averaging. Some of these applications will already be amenable to the selection of solvent systems presented in this work, whereas others may require further developments enabling the use of pure water as dissolution solvent.

4. INVESTIGATION OF PROTEIN FOLDING BY DISSOLUTION DYNAMIC NUCLEAR POLARIZATION[*]

4.1 Introduction

If a protein were to sample all the possible conformations when transitioning from an unfolded to folded state, this process could easily take longer than the age of the universe. Yet, all proteins fold in a finite time ranging from few seconds to few days [120]. This suggests that protein folding is unlikely to be a "random walk" event. This complex process has been studied by researchers for many years both for understanding the mechanism of protein folding and to gain insight into processes involved protein misfolding diseases. Protein folding is governed by a large number of factors including the sequence of the protein themselves, nature of the solvent, pH and concentration of salts [121]. Protein folding in the cell occurs at many different time points. Some proteins fold when they are nascent polypeptides [122], while others fold in the cytoplasm or after translocation through membranes with the aid of molecular chaperones [123, 124]. The presence of chaperones is believed to prevent misfolding of the nascent polypeptide chains due to interactions with the crowded molecular environment in addition to improving the overall efficiency of the folding process.

To understand the mechanism of protein folding, two models that have been proposed are the diffusion-collision model and the nucleation-propagation model [125]. In the diffusion-collision model, a protein is viewed as composed of several micro-domains, each of which can sample a range of conformations that are sterically allowed. These micro-domains, which interact with each other based on collisions start forming higher order

---

[*] Reprinted (adapted) with permission from "Protein Folding Studied by Dissolution Dynamic Nuclear Polarization, Hsueh-Ying Chen, Mukundan Ragavan and Christian Hilty, Angewandte Chemie International Edition, 52, 9192–9195, 2013". Copyright (2013) Americal Chemical Society.

structures eventually leading to the formation of stable tertiary structures [126]. Folding of several proteins have been successfully described using the diffusion-collision model [127, 128]. In the random-search nucleation-propagation model, a segment of protein is considered small enough to be able to complete a random-search for attaining a stable conformation and then serve as a nucleation site for subsequent folding of the protein. This model was described in the study of helix–coil transitions of poly-glutamic acid [129] and poly-L-asparatic acid [130].

Experimentally, there has been a great interest in understanding how proteins fold, with most of the studies being carried out *in vitro*. These studies typically involve transferring the proteins from denaturing conditions to conditions that trigger folding. Approaches to accomplish a rapid change in solution conditions include pH jump, temperature jump and dilution of denaturants (urea, guanidium hydrochloride (GdnCl), etc.) following which signal measurement has been achieved using a wide range of techniques to follow the folding process of proteins [131].

Intrinsic tryptophan fluorescence has been used monitor the folding of proteins [132, 133]. It relies on the change in observed fluorescence when the tryptophan side-chains experience a change in the chemical environment. This change has also been exploited in experiments using hydrophilic fluorescence quenchers [134]. Stopped flow fluorescence experiments have been carried out to identify molten globule intermediate states in carbonic anhydrase, β–lactamase and α–lactalbumin [135]. Circular dichroism (CD) spectroscopy is another popular technique to study protein folding. Far UV CD (170–260 nm) has been to monitor changes in the secondary structures of proteins [136, 137]. One advantage of CD spectroscopy is the ability to monitor the backbone conformational changes instead of monitoring side-chains as in the case of tryptophan fluorescence measurements. CD spectroscopy can also be used to estimate the fractions of α-helix, β sheets and random coil elements. This technique has been applied to study several proteins including cytochrome oxidase [138], staphylococcal nuclease [139] and β–lactoglobulin [140]. Folding time constants for proteins have been obtained by monitoring time resolved changes at 222 nm which report on the helical content of the proteins [141].

With the advent of greater computational capabilities, *in silico* approaches such as the energy landscape theory and molecular dynamics (MD) simulations have offered insight into understanding protein folding. The energy landscape theory takes a statistical approach where the fundamental assumption is that protein sequences are biased towards the native state, which is at a energy minimum [142]. Using this theory, folding mechanisms for proteins such as adenylate kinase have been determined computationally [143]. Molecular dynamics provides remarkable detail in the case of folding short peptides. Due to computational requirements, explicit solvent models are largely impractical in MD simulations, using results from MD simulations, design of proteins that fold faster than their wild type sequences have been carried out successfully [144]. Strategies such as distributed computing are being pursued to gain more computational power to carry out detailed simulations of behavior of proteins [145, 146].

Employing more than one technique to study protein folding is likely to provide a better understanding of protein folding since each technique could address a different facet of the problem. For example, optical techniques offer excellent time resolution can be used to monitor events at sub-millisecond time scale. NMR spectroscopy, on the other hand, offers increased site resolution, and hence is capable of providing a detailed picture of complex biological processes such as protein folding albeit with much lower time resolution. Hyperpolarization using dissolution dynamic nuclear polarization enables time resolved measurements of one-dimensional protein NMR spectra in a single scan without the need for conventional signal averaging.

In this chapter, results from the development of methods to study protein folding using dissolution DNP are described. Ribosomal protein L23, has been experimentally determined to follow two-state model of protein folding [147] and is used as the model system in this work. To improve site resolution when using uniformly labeled samples, a multiple quantum coherence based experiment using selective excitation to improve site resolution is demonstrated in a hyperpolarized experiment with ^{13}C–acetate and in non-hyperpolarized experiments using L23. Finally, an intrinsically disordered protein, p27 is used to demonstrate that dissolution DNP experiments are possible even with aqueous solvents. Changes observed in the ^{13}C spectra upon mixing with cyclin dependent kinase (CDK) – cyclin-A complex are discussed. Sequences of all the proteins used in this chapter are shown in Figure 4.1.

L23

MKTAYDVILA PVLSEKAYAG FAEGKYTFWV HPKATKTEIK NAVETAFKVK VVKVNTLHVR
GKKKRLGRYL GKRPDRKKAI VQVAPGQKIE ALEGLI

p27

GSHMEHPKPS ACRNLFGPVD HEELTRDLEK HCRDMEEASQ RKWNFDFQNH KPLEGKYEWQ
EVEKGSLPEF YYRPPRPPKG ACKVPAQE

Figure 4.1: Sequence of two different proteins used in this work. L23 is a 96 amino acid ribosomal protein from *Thermus thermophilus* which is one of the constituent proteins making up the 50S subunit of the ribosome. p27 is a 88 amino acid intrinsically disordered protein that has been reported to bind to cyclin dependent kinase/cyclin A complex.

4.2 Materials and Methods

4.2.1 Expression and Purification of L23

Protein expression and purification was carried out as described in Section 3.2.1.

4.2.2 Dynamic Nuclear Polarization and NMR Spectroscopy

L23 samples for DNP experiments were prepared at a nominal concentration of 4 mM by dissolving lyophilized powder in 60%(v/v) ethylene glycol/water mixture containing 6 M urea. Hyperpolarized samples additionally contained 15 mM (final concentration) sodium salt of tris-8-carboxyl-2,2,6,6- tetra[2-(1-hydroxethyl)]-benzo(1,2-d:4,5-dS)bis(1,3)dithiole-4-ylmethyl free radical (OX63; Oxford Instruments, Tubney Woods, U.K.), and 1 mM diethylenetriamine pentaacetic acid gadolinium complex (Gd-DTPA; Sigma-Aldrich, St. Louis, MO). Protein samples were hyperpolarized on the ^{13}C nuclei in a HyperSense Polarizer (Oxford Instruments) by irradiating with microwaves corresponding to a frequency of 93.974 GHz at 1.4 K for 4 – 6 hours. This sample was then dissolved using a solvent consisting of 10%(v/v) methanol/5 mM potassium phosphate in water, pH 3.1 heated to a vapor pressure of 10 bar. The dissolved protein sample, which is denatured at pH 3.1, was injected into an NMR tube placed in the 400 MHz NMR spectrometer using a rapid sample injection system [83]. The injected solution was mixed at a ratio of 17:1 with 25 µL of 810 mM of 4-morpholineethanesulphonic acid (MES) buffer, pH 5.6–6.2, which was already present in the NMR tube, for triggering the pH jump. Samples, after mixing, contained 40 µM L23 (concentration as determined by HPLC after the NMR experiment), 100 mM ethylene glycol, 45 mM urea, 5 mM potassium phosphate, 10% methanol and 45 mM MES at pH 5.6–6.2. Experiments to observe the changes in side chain resonances of isoleucine residues were carried out by mixing hyperpolarized protein samples with 25 µL of 1 M potassium phosphate, pH 5.7 containing 6 M urea. The final sample composition in this experiment was 40 µM L23 (concentration as determined by HPLC after the NMR experiment), 100 mM ethylene glycol, 400 mM urea, 60 mM potassium phosphate, pH 5.7 and 10% methanol.

Acquisition of NMR spectra was triggered automatically after sample injection. The temperature at the time of NMR measurement was 301 K. A segment of 3×[Sh-G] with "Sh" referring to $\pi/2$ selective radio frequency pulse (Eburp, 15 ms) and "G" referring to pulsed field gradients (x,y,z axes, 50 G/cm, 1 ms) was utilized for solvent suppression applied at the ethylene glycol resonance in the DNP experiments. Multiple spectra were acquired by applying radio frequency pulses of variable flip angles [111] followed by data acquisition (12800 complex points with t_{max} = 253.5 ms) in order to monitor the folding process. During the data acquisition, ^{1}H and ^{2}H decoupling were applied at field strengths of γB_1 = 10.2 kHz and 0.7 kHz respectively. In the pH jump experiments, the exact pH of the solution was measured after the experiment. All ^{13}C spectra were calibrated indirectly to a standard of 4,4-dimethyl-4-silapentane-1-sulphonic acid (DSS) at 0 ppm.

4.2.3 Chemical Shift Assignments

Chemical shift assignments were obtained from published values [110] and adapted for L23 in solution conditions used in DNP–NMR experiments. Backbone assignments were obtained using triple resonance experiments including 3D HNCA (2048, 64 and 64 points on ^{1}H, ^{15}N and ^{13}C dimensions respectively), HNCO (2048, 64 and 64 points) and HNCACB (2048, 64 and 96 points). Side chain assignments were obtained using 3D HCCH-TOCSY (2048, 96 and 96 points on ^{1}H, ^{13}C and ^{13}C dimensions respectively) and (H)C(C-CO)NH (2048, 64 and 96 points on ^{1}H, ^{15}N and ^{13}C dimensions respectively) experiments specifically for mapping the ^{13}C resonances of side chain methyl groups to the corresponding isoleucine residues in the protein. Spectra were acquired on a 500 MHz spectrometer equipped with a TCI cryoprobe (Bruker Biospin, Billerica, MA).

4.3 Results and Discussion

Two different solvent systems are used in dissolution DNP experiments - one for hyperpolarization (glassing mixture) and one for sample transfer. In these experiments, binary mixture of water and methanol was used for sample transfer. Protein folding in these experiments are triggered by a pH jump experiment where the hyperpolarized sample was mixed with a second buffer. Since solvent systems used in the dissolution DNP experiments are binary mixtures, the conditions used in the experiment were validated by comparing the state of the protein in these conditions against protein samples in an aqueous buffer at near-neutral pH (referred to as "standard conditions"). The protein samples were characterized using steady state multi-dimensional NMR experiments in denaturing conditions and in conditions that promote protein folding.

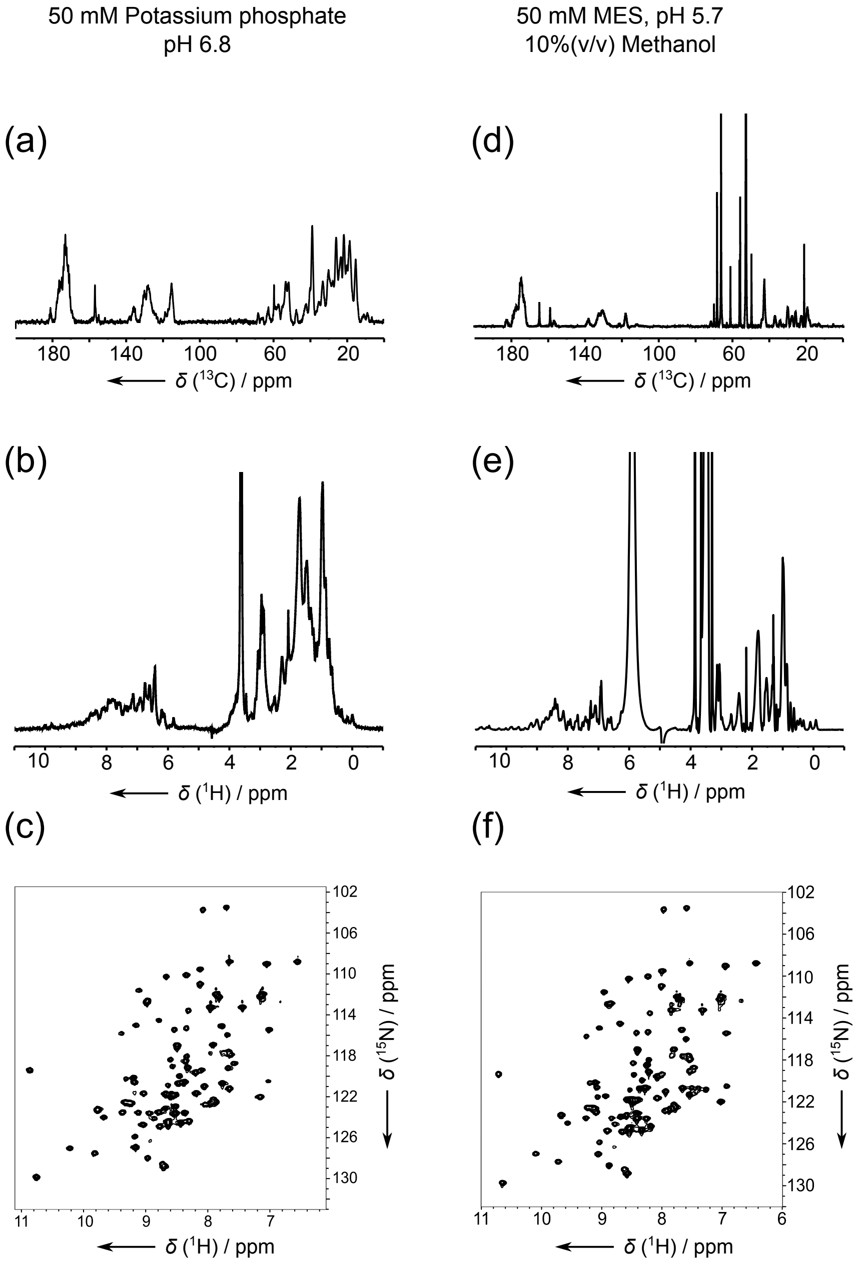
4.3.1 Protein Characterization

Protein sample quality was monitored for each preparation using ^{1}H and ^{13}C NMR spectroscopy. Spectra for sample characterization were measured in standard conditions - 50 mM potassium phosphate, pH 6.8 for folded protein standard and 50 mM potassium phosphate, pH 6.8 containing 8 M urea for unfolded protein standard. Control spectra of samples measured under conditions close to those in the DNP–NMR experiment were used to compare the state of the samples to the protein under standard conditions. Figure 4.2 shows spectra with protein in folded conformation. Figure 4.2 (a)–(c) show spectra measured at standard conditions and (d)–(f) show spectra measured at conditions used in the DNP experiment (50 mM MES, pH 5.7, 10% methanol, 50 mM urea). It can be seen that in Figure 4.2(f), the [^{1}H,^{15}N]–HSQC spectrum shows dispersed peaks with ^{1}H chemical shifts larger than 9 ppm and smaller than 8 ppm which are readily matched with those in Figure 4.2 suggesting a similar protein structure under both conditions. Further, a representative strip from one of the triple resonance experiments (3D-HNCA) used for chemical shift assignments is shown in Figure 4.3.

Similarly, Figure 4.4 shows protein spectra measured under standard denaturing conditions ((a)–(c)) and conditions that exist in the DNP experiment during sample transfer (50 mM potassium phosphate, pH 3.1, 10% methanol; panels (d)–(f)). It can be seen that protein spectra under these conditions are similar to each other and that the low pH solvent used at the start of the folding experiment is representative of the unfolded state of the protein. Spectra obtained from protein sample that was pH denatured were compared with spectra from urea denatured spectra (panels (g)–(i)). It can be seen that in all three cases, the fingerprint region of the HSQC spectra are similar suggesting similar a protein structure.

To evaluate the effect of rapid thermal cycling on the protein sample during the dissolution process, protein samples after use in the DNP experiment were characterized using reverse phase high performance liquid chromatography (HPLC, Shimadzu Scientific Instruments, Columbia, MD) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE). Figure 4.5 (a) & (b) show chromatograms obtained from pure OX-63 and [^{13}C, ^{2}H]–L23. The retention times of OX-63 and L23 were 7.7 and 13.8 minutes, respectively. Impurities in the OX-63 sample appear in Figure 4.5a at retention times 8.6 and 14.3 minutes. Figure 4.5c shows the chromatogram from a sample that underwent the DNP process. It can be seen the peaks match closely with those from pure samples, suggesting that the freeze thaw cycle did not measurably degrade the sample or alter its characteristics.

Figure 4.2: (a) ^{13}C NMR spectrum of [u-^{13}C, 50% ^{2}H]–L23 under native conditions (0.4 mM L23 in 50 mM potassium phosphate, pH 6.8). Spectrum was acquired with ^{1}H and ^{2}H decoupling. (b) ^{1}H NMR spectrum of [u-^{15}N]–L23 acquired with ^{15}N decoupling. Sample conditions are as in (a). (c) [^{15}N, ^{1}H]–HSQC spectrum of the sample from (b). (d) ^{13}C NMR spectrum of [u-^{13}C, 50%^{2}H]–L23 under folding conditions similar to DNP experiments (0.4 mM L23 in 10% methanol/50 mM MES, pH 5.7, 50 mM urea). Spectrum was acquired with ^{1}H and ^{2}H decoupling. (e) ^{1}H NMR spectrum of [u-^{15}N]–L23 acquired with ^{15}N decoupling. Sample conditions are as in (d). (f) [^{15}N, ^{1}H]–HSQC spectrum of the sample from (e). All spectra were measured on a 500 MHz NMR spectrometer equipped with a TCI cryoprobe at 301 K.


50 mM Potassium phosphate
pH 6.8
50 mM MES, pH 5.7
10%(v/v) Methanol
(a)
180 140 100 60 20
$\longleftarrow \delta\,(^{13}C)$ / ppm
(d)
180 140 100 60 20
$\longleftarrow \delta\,(^{13}C)$ / ppm
(b)
10 8 6 4 2 0
$\longleftarrow \delta\,(^{1}H)$ / ppm
(e)
10 8 6 4 2 0
$\longleftarrow \delta\,(^{1}H)$ / ppm
(c)
102 106 110 114 118 122 126 130
$\delta\,(^{15}N)$ / ppm $\longrightarrow$
11 10 9 8 7
$\longleftarrow \delta\,(^{1}H)$ / ppm
(f)
102 106 110 114 118 122 126 130
$\delta\,(^{15}N)$ / ppm $\longrightarrow$
11 10 9 8 7 6
$\longleftarrow \delta\,(^{1}H)$ / ppm

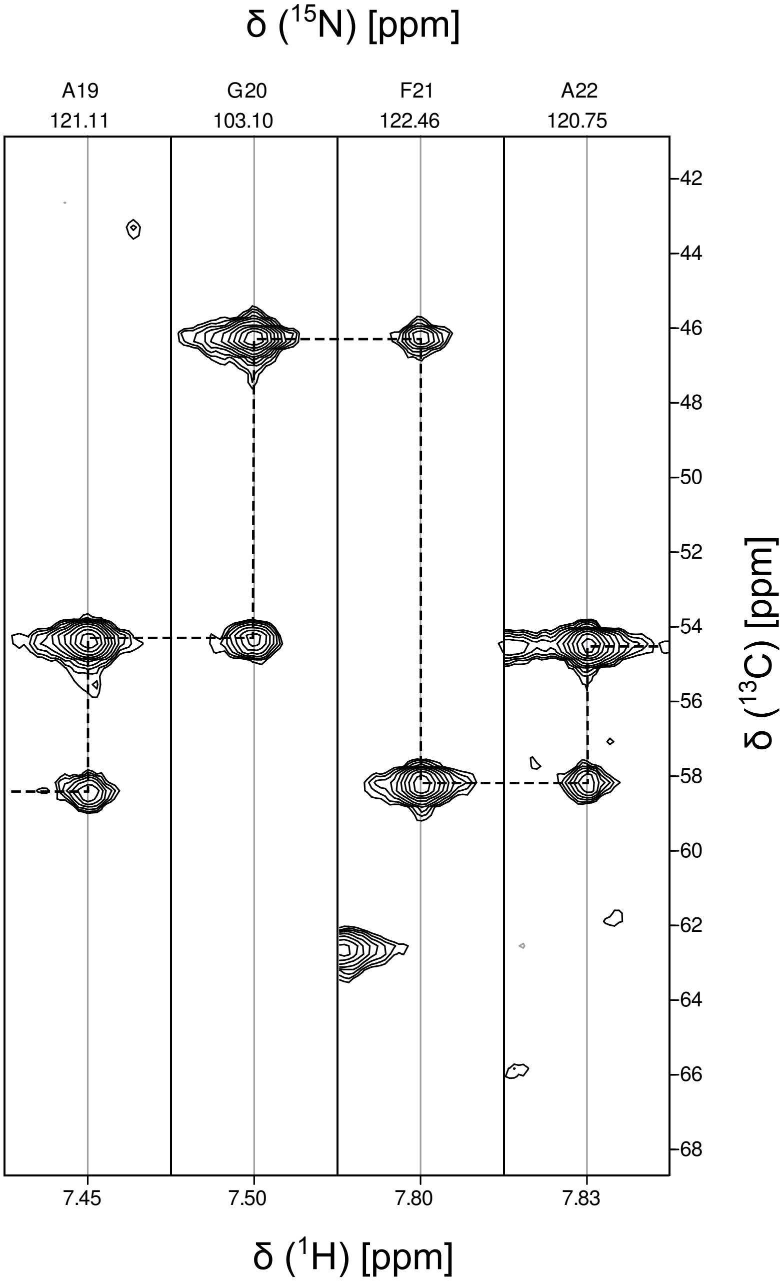
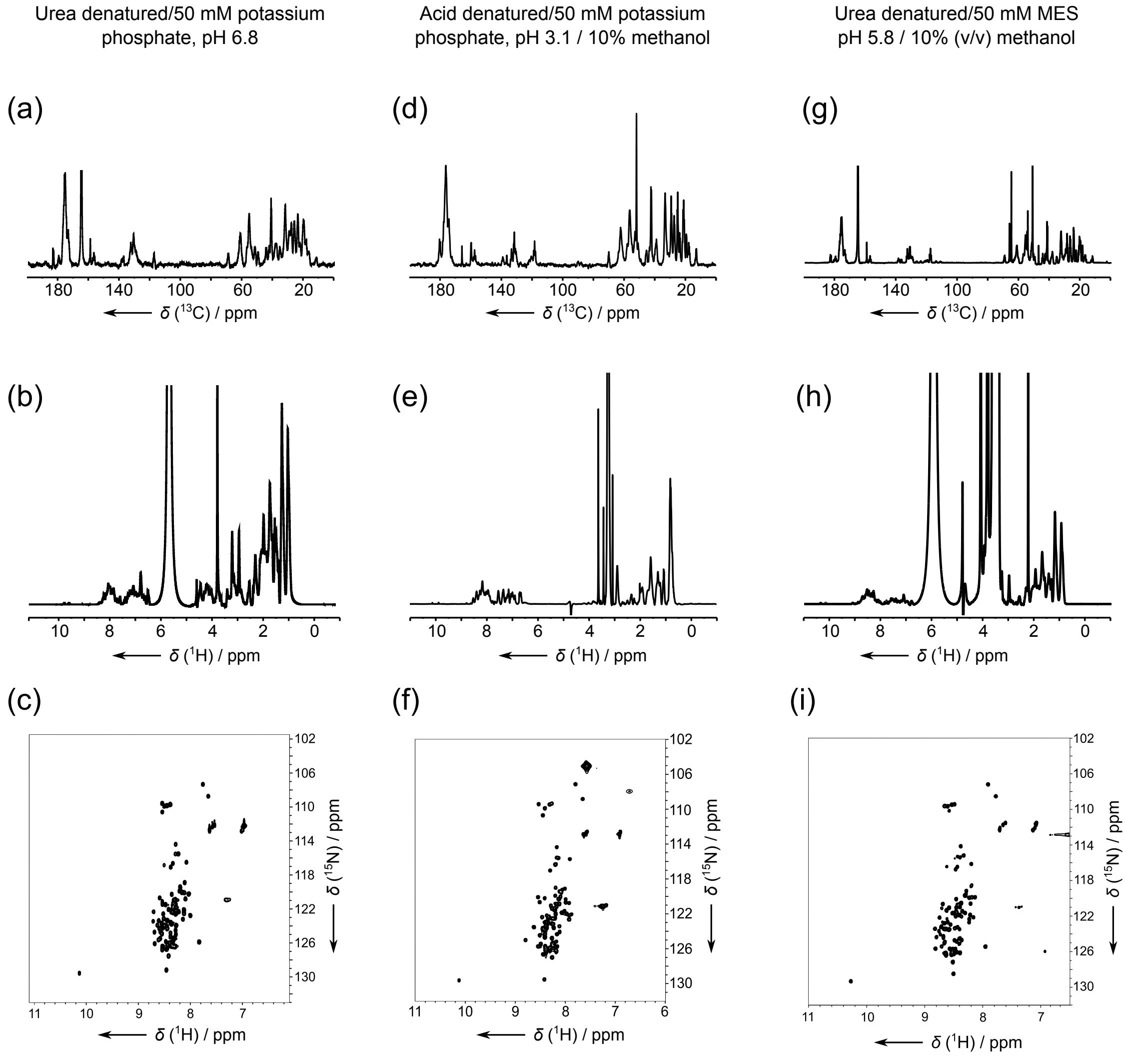


Figure 4.3: Strips from a 3D HNCA experiment showing sequential connectivity from residues 19 to 22. The spectrum was measured using a 1 mM [u-^{13}C, ^{15}N] L23 sample in 10%methanol/50 mM potassium phosphate, pH 5.7. The spectrum was acquired with 2048, 64 and 64 complex points in ^{1}H, ^{15}N and ^{13}C, respectively.

Figure 4.4: (a)^{13}C NMR spectrum of [u-^{13}C, 50%^{2}H]–L23 under standard conditions (0.4 mM L23 in 50 mM potassium phosphate,pH 6.8 containing 8 M urea). Spectrum was acquired with ^{1}H and ^{2}H decoupling. (b) ^{1}H NMR spectrum of [u-^{15}N]–L23 acquired with ^{15}N decoupling. Sample conditions are as in (a). (c) [^{15}N,^{1}H]–HSQC spectrum of the sample from (b). (d) ^{13}C NMR spectrum of [u-^{13}C, 50% ^{2}H]–L23 under denaturing conditions similar to DNP experiments (0.4 mM L23 in 10% methanol/50 mM potassium phosphate, pH 3.1, 50 mM urea). Spectrum was acquired with ^{1}H and ^{2}H decoupling. (e)^{1}H NMR spectrum of [u-^{15}N]–L23 acquired with ^{15}N decoupling. Sample conditions are as in (d). (f) [^{15}N, ^{1}H]–HSQC spectrum of the sample from (e). (g) ^{13}C NMR spectrum of [u-^{13}C, ^{15}N]–L23 under conditions similar to those existing in the DNP experiments (0.4 mM L23 in 10%(v/v) methanol/50 mM MES pH 5.8, 100 mM ethylene glycol containing 8 M urea). Spectrum was acquired with ^{1}H and ^{15}N decoupling. (h)^{1}H NMR spectrum of [u- ^{15}N,^{13}C]–L23 acquired with ^{15}N and ^{13}C decoupling. Sample conditions are as in (g). (i) [^{15}N, ^{1}H]–HSQC spectrum of the sample from (h). All [^{15}N, ^{1}H]–HSQC spectra were measured with 1024 and 256 complex points in ^{1}H and ^{15}N dimensions, respectively, on a 500 MHz NMR spectrometer equipped with a TCI cryoprobe at 301 K.


Urea denatured/50 mM potassium phosphate, pH 6.8
Acid denatured/50 mM potassium phosphate, pH 3.1 / 10% methanol
Urea denatured/50 mM MES pH 5.8 / 10% (v/v) methanol
(a)
(d)
(g)
180 140 100 60 20
$\longleftarrow \delta\,(^{13}C)$ / ppm
(b)
(e)
(h)
10 8 6 4 2 0
$\longleftarrow \delta\,(^{1}H)$ / ppm
(c)
(f)
(i)
102 106 110 114 118 122 126 130
$\delta\,(^{15}N)$ / ppm $\downarrow$
11 10 9 8 7 6
$\longleftarrow \delta\,(^{1}H)$ / ppm

Additionally, a sample of L23 that experienced the freeze-thaw cycle was compared with a newly purified L23 sample on a SDS-PAGE gel. The coomassie blue stained gel is shown in Figure 4.5d. No degradation bands can be observed in the gel and the signatures between the two samples are indistinguishable. Based on the HPLC and gel electrophoresis data, it is likely that freeze-thaw process does not have an adverse affect on the protein samples.

4.3.2 Observation of Protein Folding

The refolding of L23 was monitored using a pH jump experiment. The hyperpolarized sample was dissolved under denaturing conditions (pH 3.1) and injected into the NMR spectrometer, where it is mixed with a buffer at higher pH (pH 5.7). This pH jump triggers the folding of L23 which is then monitored over a period of few seconds by measuring a series of 1D ^{13}C NMR spectra using small flip-angle excitation pulses. It should be noted here that the samples contain 45 mM urea and 100 mM ethylene glycol (from the glassing mixture) in addition to L23 and buffer components (potassium phosphate and MES).

The spectra from this experiment are shown in Figure 4.6(a). Signal changes over different transients that occur due to the folding of the protein, as well as due to non-uniform nuclear spin relaxation, can be seen in Figure 4.6(b). Here, the differences of each scan to the first (cyan–blue) and last (yellow–red) scan in the time series are shown, after normalization of each spectrum to equal intensity integrated over carbonyl and aliphatic regions. The composite plot shown in (b) was computed using the equations 4.1 (cyan–blue) and 4.2 (yellow–red).

$$\Delta(D'_{k,n}, D'_{5,n}) = max(0, D'_{k,n} - D'_{5,n}) \tag{4.1}$$

$$\Delta(D'_{k,n}, D'_{1,n}) = max(0, D'_{k,n} - D'_{1,n}) \tag{4.2}$$

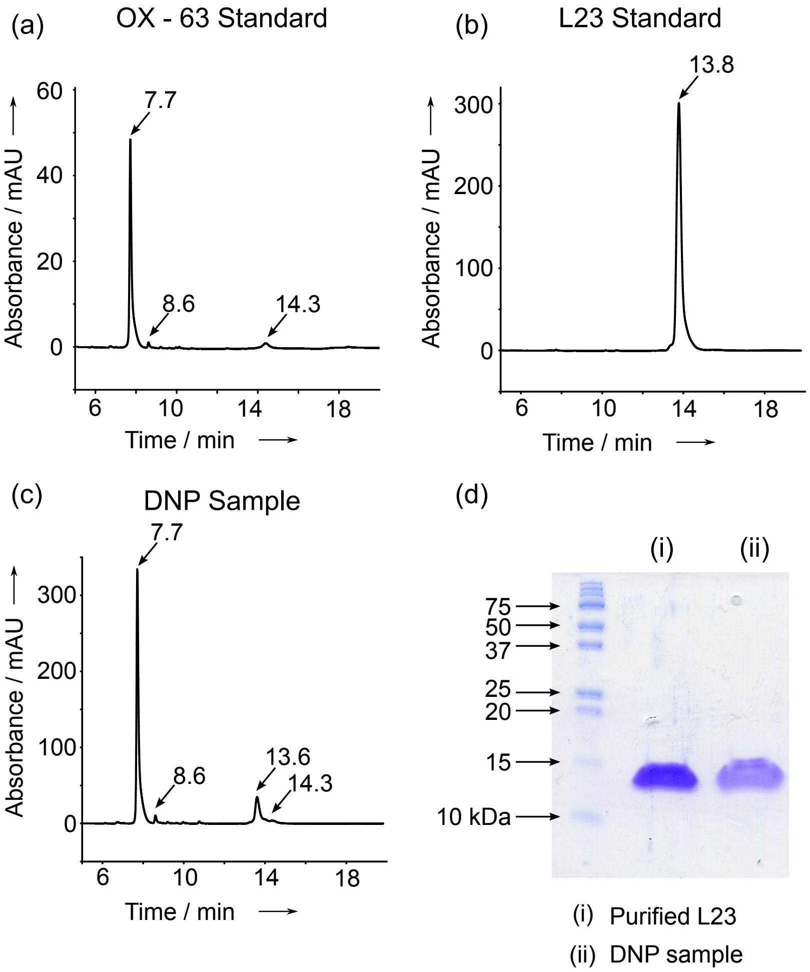


Figure 4.5: (a) HPLC chromatogram of OX–63 radical obtained using a C–18 column with a gradient of 10%–100% acetonitrile in water and 0.08% trifluoro acetic acid. The retention time is 7.7 minutes. (b) HPLC chromatogram of purified [^{13}C,^{2}H] L23 (retention time = 13.8 min). (c) HPLC chromatogram of [^{13}C,^{2}H] L23 after hyperpolarization and dissolution (retention time = 13.6 min). OX–63 in the DNP sample was eluted out of the column with a retention time of 7.7 min. (d) SDS–PAGE comparing a fresh L23 sample with a sample that had undergone a DNP–NMR experiment as obtained on a 15% polyacrylamide gel with coomassie blue staining.

where, $D'_{k,n}$ is obtained by rescaling each spectrum ($D_{k,n}$) to the same integration value between the backbone carbonyl (185.7 –170 ppm) and aliphatic (63–7 ppm) regions of the spectrum. k is the index of different spectra in the time series of spectra measured in the dissolution DNP experiment and n is the number of data points in each spectrum.

Since the signal intensity in these spectra decreases gradually because of both spin relaxation and the folding process, in order to observe the evolution of carbonyl resonances, all the spectra in the time series were re-scaled to unit maximum intensity at ~176 ppm. After re-scaling, increase in relative intensities can be observed in the backbone carbonyl chemical shift range (170–185 ppm) as shown in Figure 4.6 (c). Two "shoulders" (indicated by arrows in the figure) are clearly visible in the normalized traces of the corresponding region and are characteristic of the folded protein as can be seen from the standard spectra shown in Figure 4.6 (d). Spectra from a control experiment where no pH jump was employed is shown in Figure 4.6(e).

In these spectra, backbone carbonyl resonances are between 182–170 ppm whereas the resonance at ~183 ppm (and higher) correspond to C^{δ} of glutamic acid. L23, which has 6 Glu residues, show a strong peak at this chemical shift. However, the appearance of this peak depends on the pH of the sample during measurement. It can be seen in Figure 4.6(d) that the C^{δ} resonance appears when the sample has pH 5.7 and not in samples of pH 3.1. These sample conditions correspond to deprotonated and protonated states of the side-chain carbonyl of Glu residues.

4.3.3 Estimation of Secondary Structure Formation

NMR chemical shifts are strongly dependent on the nature of secondary structures in the protein. For example, in the case of ^1H chemical shifts, residues existing in α–helices experiences a downfield shift while residues in β–sheets exhibit an upfield shift when compared with random coil chemical shift values [11]. Using this feature in NMR spectra, a

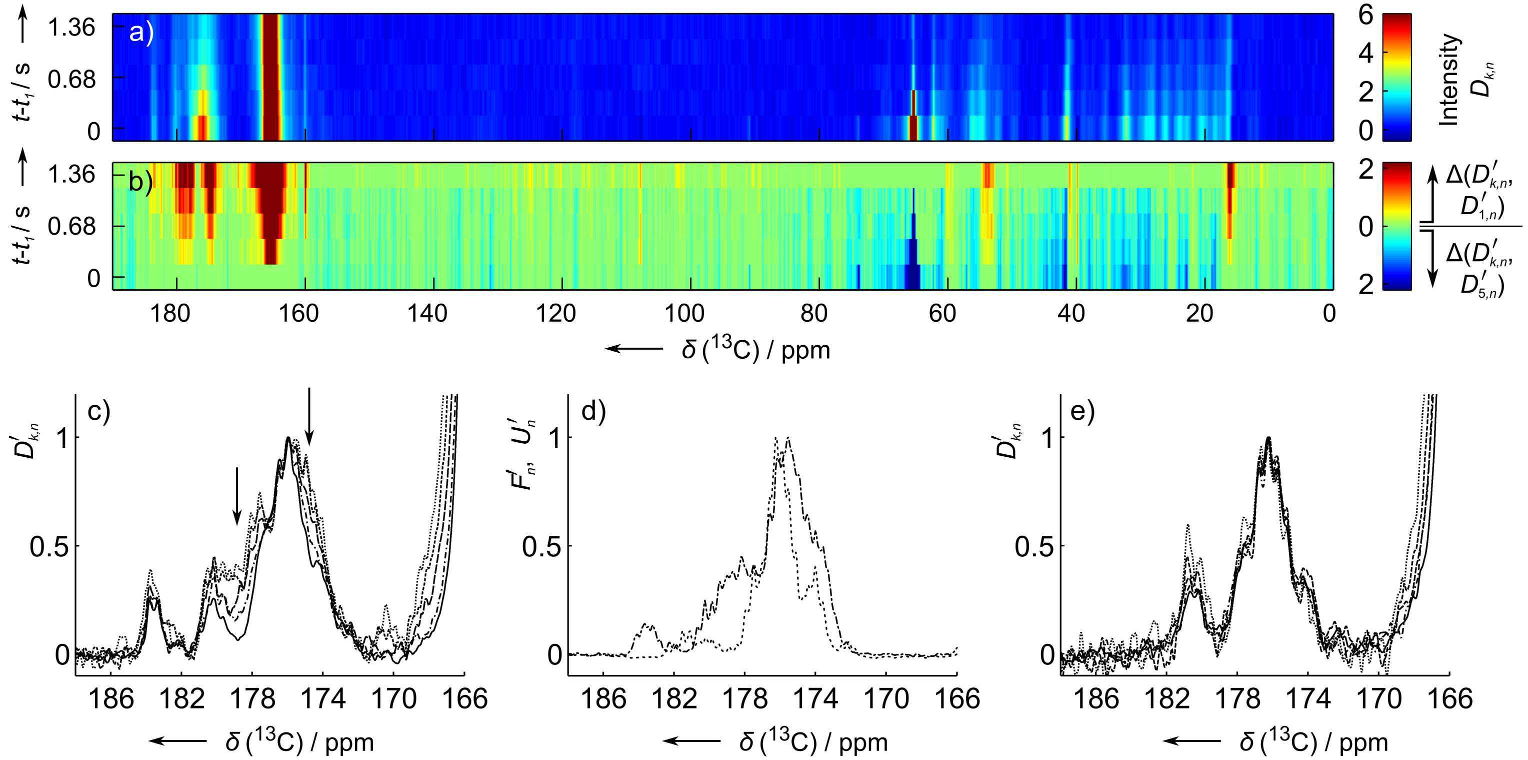


Figure 4.6: (a) Intensity plot of a time series of 5 spectra from a single hyperpolarized sample of [u-^{13}C, 50%-^{2}H]-L23, after pH jump to initiate protein folding (40 μM L23, 100 mM ethylene glycol, 45 mM urea, 10% (v/v) MeOH (methanol), 5 mM KH_2PO_4 and 45 mM 4-morpholineethanesulfonic acid (MES) buffer, pH 5.8, T=301 K, B=9.4 T). t_1 is the time of the first scan. b) Composite plot showing in cyan–blue the difference compared to the last spectrum, $\Delta(D'_{k,n}, D'_{5,n}) = \max(0, D'_{k,n} - D'_{5,n})$, and in yellow–red the difference compared to the first spectrum, $\Delta(D'_{k,n}, D'_{1,n}) = \max(0, D'_{k,n} - D'_{1,n})$ in the time series. Each spectrum $D'_{k,n}$ is obtained by rescaling $D_{k,n}$ to the same total integration value in the range 185.7 – 170 ppm and 63 – 7 ppm prior to taking the difference. (c) Time evolution of carbonyl resonances (increasingly shorter dashes for later time; spectra are re-scaled to unit maximum intensity near 176 ppm). (d) Reference spectra measured without hyperpolarization, of unfolded L23 (- - -; 40 μM L23, 100 mM ethylene glycol, 45 mM urea, 10% (v/v) MeOH, 5 mM KH_2PO_4, pH 3.1, 32,000 scans) and folded L23 (–·–·; using sample from (a), 32,000 scans), cryoprobe at B=11.7 T. (e) Series of spectra as in (c), but without pH jump.

chemical shift index approach has been developed to identify secondary structure elements in different regions of the spectrum using ^{1}H chemical shifts [148] and ^{13}C chemical shifts [12].

Such variations in the chemical shifts between unfolded and folded conformations of a protein sample can be seen in spectra obtained from DNP experiments as well. In Figure 4.6 (c) & (d), the two shoulders around ~178 ppm and ~174 ppm arise from the folded protein, while the resonances are collapsed in the case of denatured protein. These shoulders represent intensities from α–helix and β–sheet secondary structures, respectively. Using these data obtained from dissolution DNP experiments, it would be possible to estimate secondary structure content in the protein at each measured time point in the course of the reaction by expressing the DNP–NMR spectra as a linear combinations of statistical distributions corresponding to different secondary structure elements.

Chemical shifts corresponding to >2000 proteins were obtained from a database [149], and distributions of ^{13}C and ^{15}N shifts corresponding to α–helix, β–sheet and random coil structures were computed. These distributions are shown for ^{13}C of backbone carbonyl and ^{15}N of backbone amide groups in Figure 4.7 (a) & (b), respectively. It is of interest that in the case of backbone carbonyl ^{13}CO, chemical shifts pertaining to β–sheets and random coil show significant overlap, whereas in the case of backbone amides ^{15}N, it is the α–helix and random coil chemical shifts that overlap most.

Linear decomposition using the ^{13}C chemical shift distributions was applied (Figure 4.8(a)–(e), corresponding to scans #(1)–(5)). Percentages of the different types of structures were extracted and summarized in (f). As expected, the fraction of α-helix and β-sheet structure increases, while the proportion of random coil decreases during the folding process. Comparing to the percentage of residues present in secondary structure elements in the PDB structure, the amount of secondary structure appears underestimated (~45% estimated random coil from spectrum vs ~40% random coil as estimated from the pub-

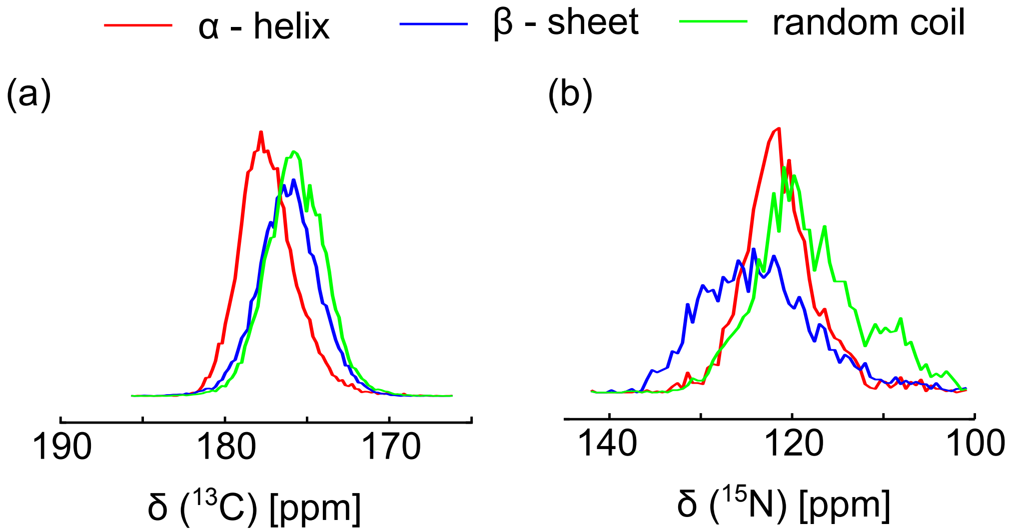


Figure 4.7: (a) Distribution of ^{13}C chemical shifts in the carbonyl region. (b) Distribution of ^{15}N chemical shifts. The histograms were computed using the data re-referenced chemical shifts found in the literature [149].

lished structure (PDB I.D. 1n88)). This may be due to the fact that the statistical chemical shift distributions do not match L23 exactly. Overlap of the chemical shift distributions for the three secondary structure types may affect the fit, particularly for differentiating β-sheet from random coil contribution, for which the maxima of chemical shift distributions lie closest together.

Although in the present case, only ^{13}C chemical shift distributions were used, this type of analysis would be applicable to data from ^{15}N polarized experiments as well. Although ^{15}N is a nucleus of low sensitivity in view of measuring NMR spectra, amide group ^{1}H and ^{15}N chemical shifts are typically used as the fingerprint of proteins, making analyses of ^{15}N chemical shifts valuable.

4.3.4 Resolving Individual Sites

In addition to fractions of secondary structures, information pertaining to specific sites can also be obtained from the DNP–NMR spectra. In the case of L23, Ile $C^{\delta 1}$ groups

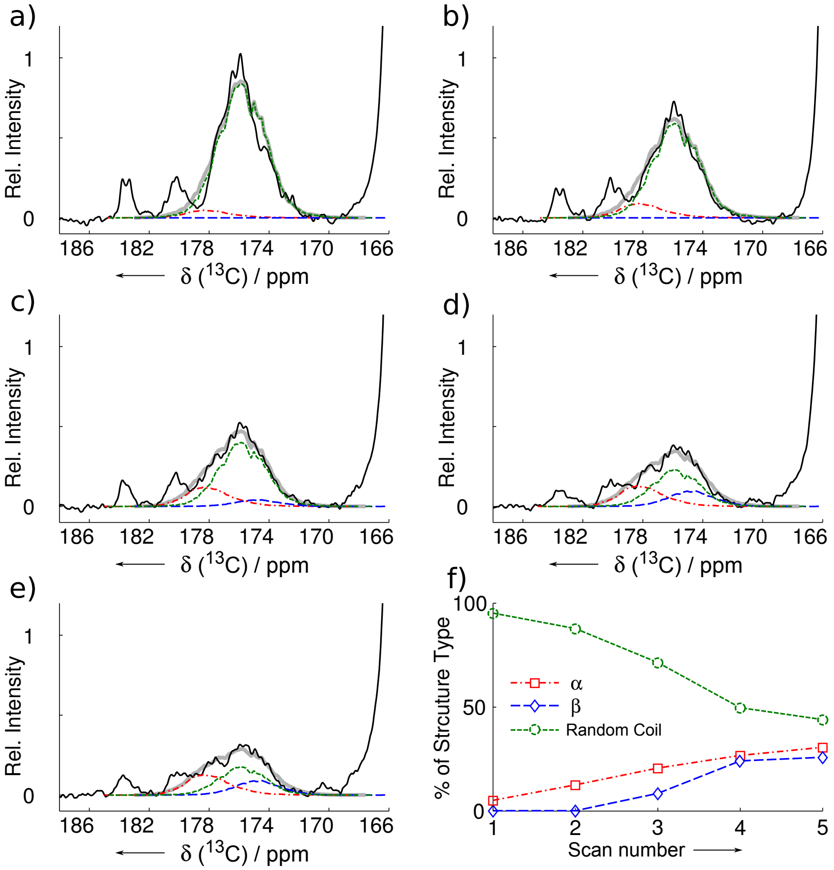


Figure 4.8: (a)–(e) Decomposition of DNP-NMR spectra (from the same dataset as in Figure 4.6) (black —) in scans #1–5 into three references generated from statistical distribution of chemical shifts in α-helix (red –·–·), β-sheet (blue – –) and random coil (green - - -) secondary structure. Chemical shift distributions were derived from published re-referenced chemical shifts [149]. The reconstructed spectrum from linear combination of references is shown as a thick grey line. (f) Time-course of obtained fractions of secondary structures; α-helix (□), β-sheet (◇) and random coil (○).

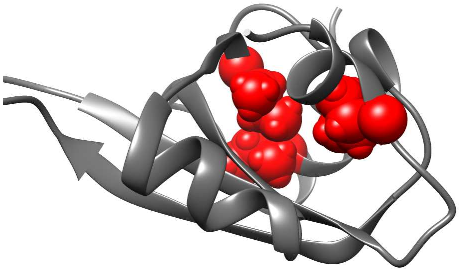


Figure 4.9: Representation of Ile residues that are present in hydrophobic core of L23. Ile residues 8, 39 and 89 are shown in red spheres space-filling representation with rest of the protein shown in a cartoon representation. The methyl groups in the hydrophobic core would be expected to undergo a significant change in the local chemical environment upon folding of the protein. This change in chemical environment would be reflected in NMR spectra as shown in Figure 4.10. Image was prepared by using the NMR structure deposited in PDB with I.D. 1n88.

are sufficiently separated for this purpose (Figure 4.10). The methyl groups are often located in the hydrophobic core of a protein (in L23, Ile 8, 39, and 89 are part of the hydrophobic core, Figure 4.9). These methyl groups experience a significant change in chemical environment upon folding and hence are useful indicators of protein folding.

Spectra taken at three time points reveal a large initial contribution from random-coil chemical shifts, followed at later time by resolution of individual peaks for the different methyl groups in the folded protein. The identities of chemical shifts of the individual residues were obtained from triple resonance experiments (Figures 4.11 and 4.12) and compared as shown in Figure 4.10. The comparison of DNP spectra and non-hyperpolarized spectra must take into account the isotope effect on chemical shifts. The signal for individual methyl groups in the DNP experiment will be dominated by the isotopomer with the longest T_1 ($CD_3 > CD_2H > CDH_2 > CH_3$) while the non-hyperpolarized spectra will have greater contribution from the isotopomers with shorter relaxation times. The relative

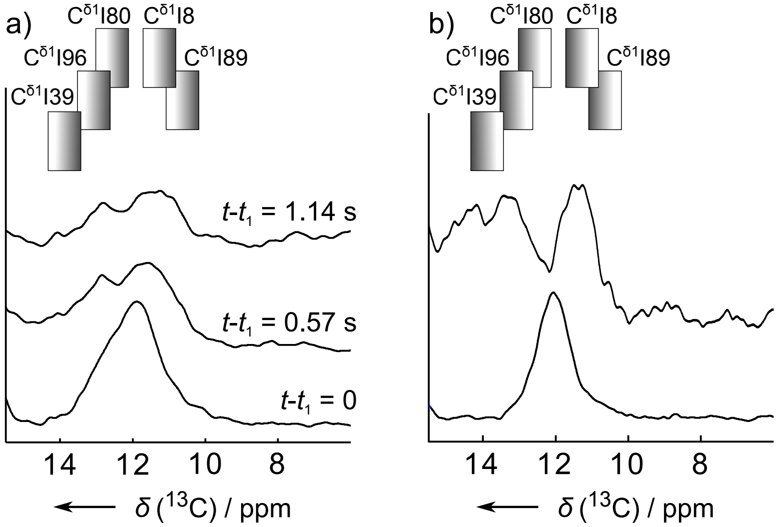


Figure 4.10: (a) DNP-NMR spectra showing the time evolution of Ile $C^{\delta1}$ chemical shifts after pH jump (40 μM L23, 60 mM KH_2PO_4, pH 5.7, 10% (v/v) MeOH, 400 mM urea, 100 mM ethylene glycol, T=301 K). (b) Non-hyperpolarized reference spectra of L23 unfolded (bottom; 0.4 mM L23, 10% (v/v) MeOH/50 mM KH_2PO_4, pH 3.1) and folded (top; sample as in (a), 50,000 scans on cryoprobe at 11.7 T). In (a) and (b), chemical shifts of folded L23 are indicated with ranges covering isotopomers CD_3, CD_2H, CDH_2 and CH_3 [116]. Dark shading indicates the position of highest relative detection sensitivity; relaxation losses in the DNP experiment are lowest for CD_3, and equilibrium polarization in the non-hyperpolarized experiment is largest for CH_3.

contributions of isotopomers to intensity in the spectra are indicated using shaded bars in Figure 4.10. The chemical shift changes due to deuteration has been discussed elsewhere in the literature [116]. These data illustrate the complimentary nature of dissolution DNP and high resolution NMR experiments. While the dissolution DNP experiments enable following transient processes in real time, high resolution experiments offer unparalleled resolution to uniquely identify different sites.

4.3.5 Selective Multiple Quantum Experiment for Improving Site Resolution

In the ^{13}C spectra of L23 measured in dissolution DNP experiments, only the methyl groups of Ile are clearly resolved. Since this is a drawback in measuring one-dimensional

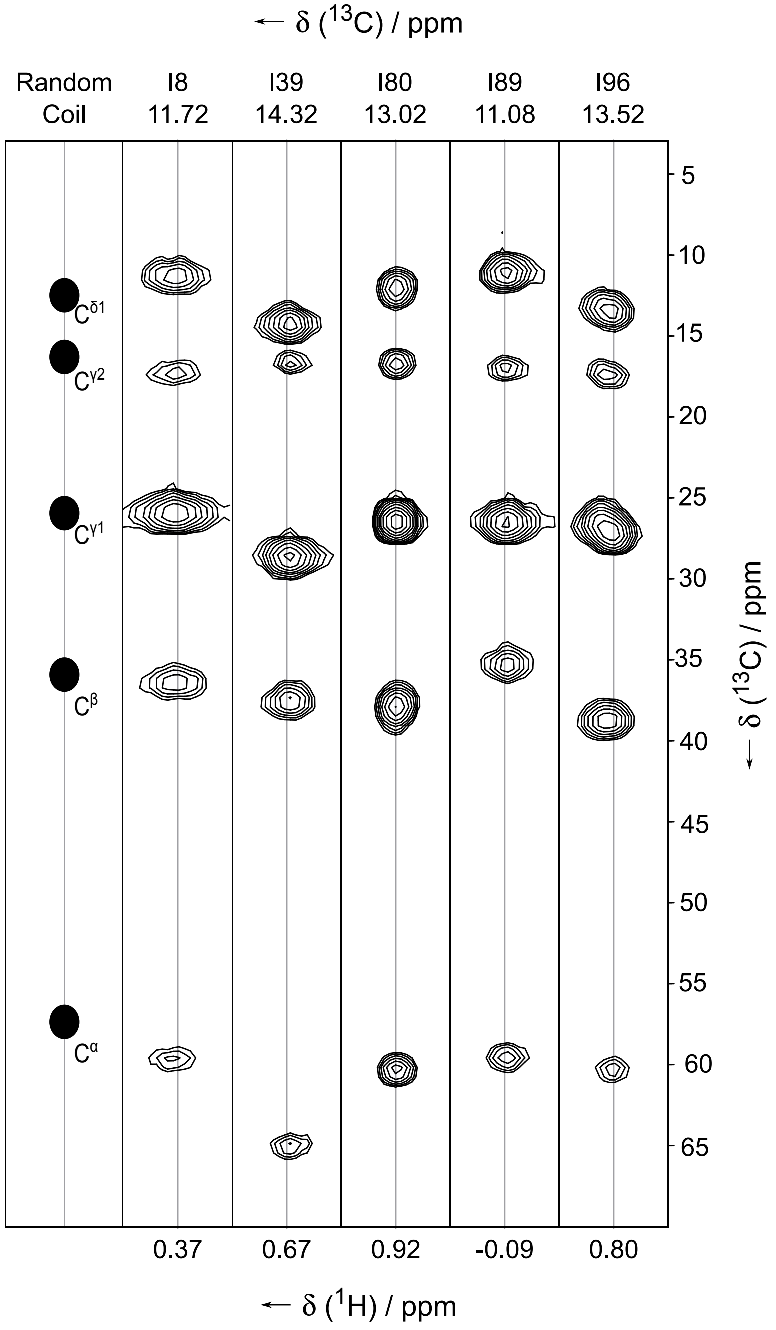


Figure 4.11: Strips from a 3D HCCH-TOCSY experiment showing side chain resonances of isoleucine residues. Strips are taken at chemical shifts of $C^{\delta 1}/H^{\delta 1}$. The spectrum was measured using a 1 mM [u-^{13}C, ^{15}N] L23 sample in 10%methanol/50 mM potassium phosphate, pH 5.7. The spectrum was acquired with 2048, 96 and 96 complex points in ^{1}H, ^{13}C and ^{13}C respectively.

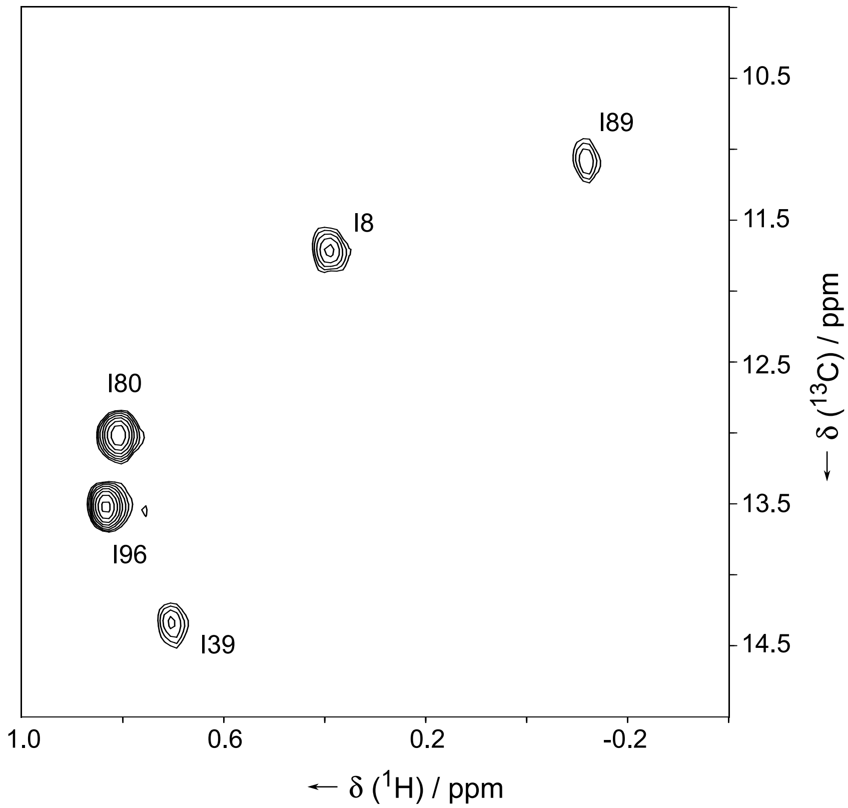


Figure 4.12: [^{13}C, ^{1}H]-HSQC spectrum showing the region with peaks from $C^{\delta 1}$ groups of isoleucine residues. This spectrum was measured using 0.4 mM L23 (no isotope enrichment) in 10% methanol/50 mM potassium phosphate, pH 5.7. 1024 and 700 complex points were acquired for ^{1}H and ^{13}C dimensions, respectively, on a 500 MHz NMR spectrometer equipped with a TCI cryoprobe at 301 K.

experiments, multidimensional experiments have been routinely used in conventional NMR to gain spectral resolution. However, due to non-recoverable nature of polarization in DNP experiments, using multidimensional experiments is not always straightforward. To improve the spectral resolution in the DNP experiments, sequentially acquired two-dimensional and ultrafast two-dimensional NMR experiments have been proposed [111, 92].

In conjunction with using two- and three-dimensional NMR experiments, use of site selective labeling or amino acid selective labeling have enabled improving resolution in the NMR spectra. This biochemical approach which is undoubtedly, an effective strategy may not always be feasible. When using uniformly labeled samples, amino acid type-selective experiments to resolve ambiguities in 2D and 3D experiments have been implemented [150, 151, 152]. For example, TAVI-HSQC experiment specifically selects [^{1}H-^{15}N] correlations in threonine, alanine, valine and isoleucine residues in the spectrum. This type of selection greatly simplifies the recorded spectrum and enables near complete assignment of chemical shifts.

In situations where multidimensional NMR experiments are challenging, one approach to improve resolution in one-dimensional experiments would be to spectroscopically select specific coherences for observation. Such an experiment would be helpful in identification of different features in the spectra without the need for multidimensional experiments. In one strategy, multiple quantum coherence can be generated between two scalar coupled spins of the same type (i.e., ^{13}C–^{13}C coupling). After selection of the required coherences using pulsed field gradients, the multiple quantum coherence can be converted to observable single quantum coherence.

The pulse sequence for an experiment utilizing chemical shift based selection is shown in Figure 4.13. Product operator description of this experiment is similar to the conventional heternuclear multiple quantum coherence (HMQC) experiment which has been described in the literature [153]. The product operator treatment is briefly revisited for the current experiment below. For simplicity, chemical shift evolution of the coherences are not shown and both spins, I & S represent ^{13}C spins in the sample that are bonded. The product operator treatment for the first part of the pulse sequence upto and including the first selective pulse is shown below.

$$I_z \xrightarrow{\frac{\pi}{2}x} -I_y \xrightarrow{2\pi J_{IS}} 2I_xS_z + 2I_yS_z \xrightarrow{sel.\frac{\pi}{2}x} 2I_xS_y - 2I_yS_y \tag{4.3}$$

After the application of selective $\frac{\pi}{2}$ pulse, a double quantum coherence is generated. The π pulse refocusses the evolution of coherences due to chemical shift. The product operator treatment for the last part of pulse sequence including and after the second selective $\frac{\pi}{2}$ pulse is shown here.

$$-2I_xS_y + 2I_yS_y \xrightarrow{sel.\frac{\pi}{2}x} 2I_xS_z + 2I_yS_z \xrightarrow{2\pi J_{IS}} -I_y \tag{4.4}$$

Gradient selection of double quantum coherence can be achieved by using the two pulsed field gradients of opposite signs at 1:-2 ratio.

This experiment was tested using a ^{13}C labeled samples of acetate in a hyperpolarized experiment and using denatured [^{13}C, 50%^{2}H] L23 in a non-hyperpolarized experiment. Spectra obtained using the selective experiment as applied to ^{13}C–acetate is shown in Figure 4.14. Figure 4.14 (b) & (c) show results from a hyperpolarized experiment using 300 mM ^{13}C –acetate. In (b), the selective excitation is on the methyl resonance of acetate. This selects the resonance bonded to it, in this case, the carbonyl resonance while the methyl resonance itself is suppressed yielding a single peak in the spectrum. The spectrum from the excitation of the carbonyl peak leading to observation of the methyl peak is shown in (c).

Figure 4.15 shows the performance of the selective experiment on a sample of denatured [^{13}C, 50%-^{2}H]–L23 sample. Selective excitation on the C$^{\alpha}$ region enabled observa-

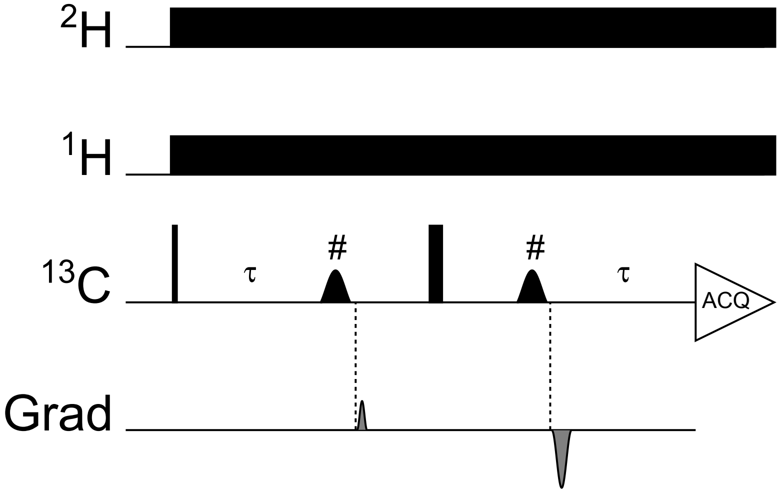


Figure 4.13: Pulse sequence used for the selective multiple quantum coherence experiment. ^{1}H and ^{2}H decoupling were applied for the entire duration of the experiment. The selective pulses (indicated by #) were of Sinc shape with 2500 μs duration. τ, delay for scalar coupling evolution, is set to $\frac{1}{2J}$. All pulses were of phase x and coherence selection is accomplished by applying pulsed field gradients at strengths of 25 G/cm and -50 G/cm.

tion of the carbonyl region and vice versa as shown in Figure 4.15 (b) & (c). In order to test the experiment in a crowded region of the spectrum, selective excitation was applied at ~33 ppm. This yielded a spectrum with resonances that likely correspond to bonded C^{α}, C^{β}, C^{γ} & C^{ϵ}. In comparison with the simulated spectrum of L23 in Figure 3.2 (d), the selective radiofrequency pulse would excite resonances C^{β} of Val, Lys, Met & Pro and C^{δ} of Lys. This excitation would be expected to lead to the observation of C^{α} (V, K, M, P), C^{ϵ}(K) and $C^{\gamma 1}$ & $C^{\gamma 2}$ of Val resonances. Indeed, the observed spectrum shown in Figure 4.15 (d) contains these resonances at ~60, ~40 ppm and ~20 ppm, respectively. L23 contains 11 valine and 16 lysine residues and only one methionine and one proline. This distribution of residues is likely to result in a larger contribution of Lys and Val residues to observed signal than the other two residues.

Such site selection greatly simplifies the spectrum, potentially enabling a detailed anal-

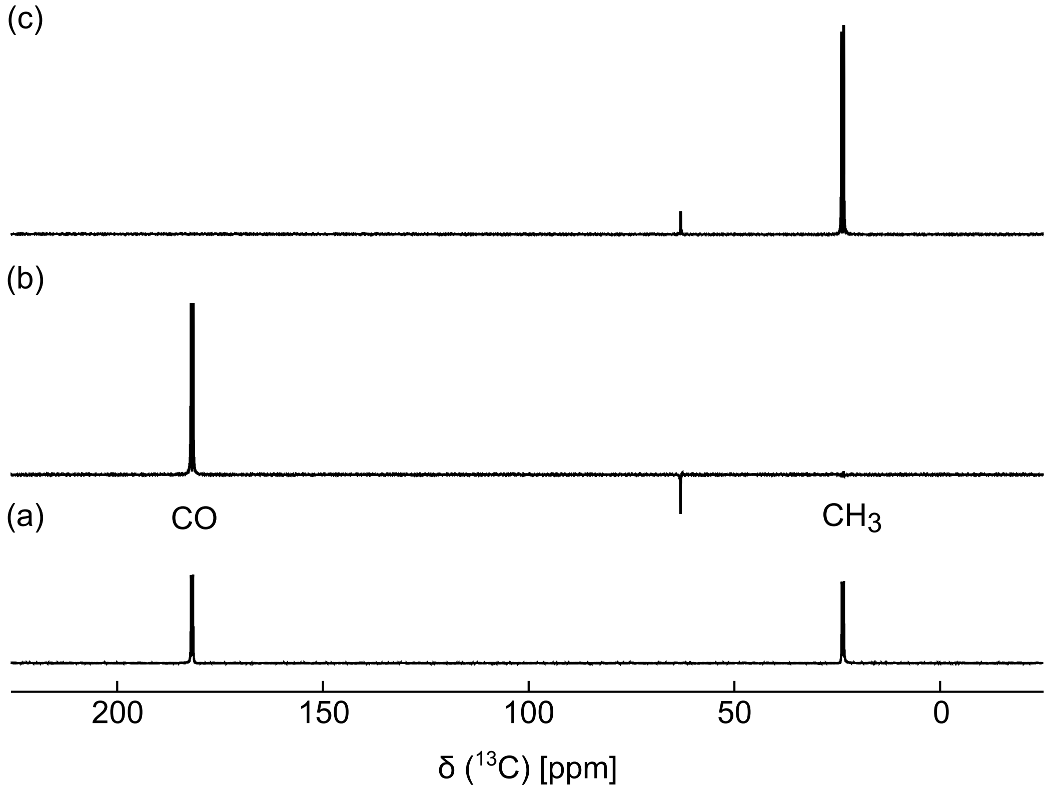


Figure 4.14: (a) Non-hypeprolarized spectrum of 700 mM ^{13}C –acetate. Identities of the peaks are indicated above the peak. Dissolution DNP enhanced NMR experiment showing selection of (b) carbonyl resonance and (c) methyl resonance for observation. Selective excitation was on resonances corresponding to (b) 23.6 and (c) 181.7 ppm. Spectra were measured with ^{1}H and ^{2}H decoupling. The delay, τ for evolution under the influence of scalar coupling was set to 9.1 ms corresponding to coupling constant of 55 Hz.

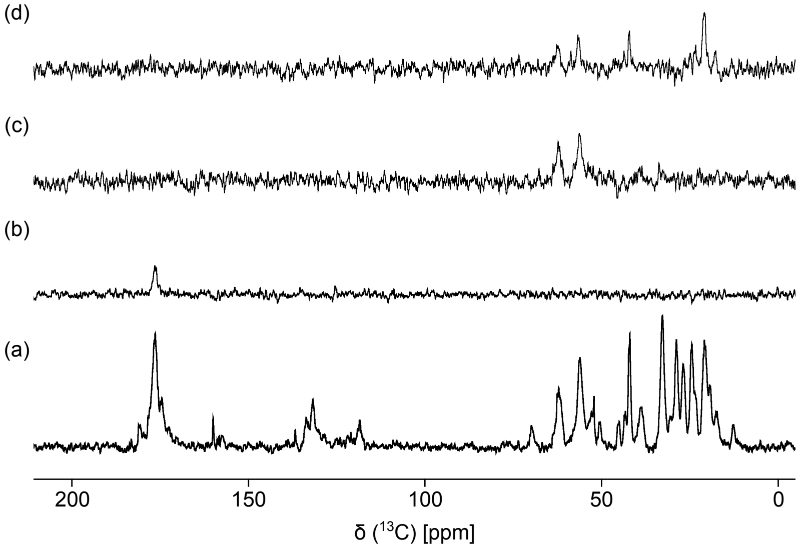


Figure 4.15: (a) Non-hypeprolarized spectrum of 1.5 mM [^{13}C, 50%-^{2}H]–L23. Non-hyperpolarized spectra from selective experiment with excitation at on resonances corresponding to (b) 62.4, (c) 176.5 and (d) 33.2 ppm. Spectra were measured with ^{1}H and ^{2}H decoupling. The delay τ for evolution under the influence of scalar coupling was set to (b) & (c) 9.1 and (d) 14.2 ms (coupling constants of 55 and 35 Hz, respectively). Selective excitation was applied using a single lobed Sinc pulse with 1000 μs pulse length and spectra were acquired using (a) 256 scans and (b), (c) & (d) 1024 scans.

ysis even without the need for special isotopic labels. Signal averaging was employed in these non-hyperpolarized experiments. Since the selective experiments involve a coherence transfer step, a higher number of scans are needed to record spectra of sufficient S/N (Figure 4.15 caption). Nevertheless, this experiment may, with further development, be useful to monitor site specific changes in time resolved experiments using dissolution DNP.

4.4 Dissolution DNP of p27 Using Aqueous Solvents

An important aspect of the protein folding experiment described above is the nature of the solvent used to study folding. In these experiments, water/methanol mixtures were used to overcome the foaming problem during the injection of the sample. However, reports of water/organic mixtures destabilizing protein structure and impeding function are available in the literature [154, 155]. At the same time, several enzymes are known to function as effectively in seemingly hostile conditions as they function in near-native conditions [156, 157]. In fact, some enzymes are known to function more efficiently in organic solvents than in aqueous conditions [158]. Given the complexity of protein solvent interactions, determination of best suited condition for study of protein folding and activity must be carried out experimentally.

The best case, however, would be the use of a purely aqueous dissolution solvent. This might be of particular importance in experiments designed to study protein–protein interactions. Some proteins, in fact, are amenable to dissolution DNP experiment using just aqueous buffers, in the current setup without modification. For example, p27, a 88 amino acid intrinsically disordered protein that binds to Cyclin-dependent Kinase (CDK) – cyclin A complex, can be hyperpolarized and subsequently dissolved and injected using an aqueous buffer containing 100 mM Tris at pH 7.2. Dissolution DNP spectra measured from [u-^{13}C,^{2}H]-p27 is shown in Figure 4.16.

Closer inspection of the spectra shown in Figure 4.16 reveals fine structure in many peaks (for example, peaks at ~183 ppm, ~179, ~170–175 ppm, ~50–56 ppm, etc.) Prominent peaks corresponding to aromatic residues (17 aromatic residues are present in p27) in the protein are present between 120–140 ppm. It is likely that the p27 spectra show more prominent peaks in this region than seen in the hyperpolarized spectra of L23 because of higher level of deuteration in p27. The peak that is likely to correspond to C^{δ} of Glu residues in p27 is well pronounced in these spectra at ~183 ppm.

Since p27 is an intrinsically disordered protein (IDP), it is likely to show random coil secondary structure when existing as an isolated polypeptide in solution. These IDPs typically fold upon binding their targets or act as linkers and play a role in aiding assembly of macromolecular complexes [159]. It may be possible to observe these structural changes

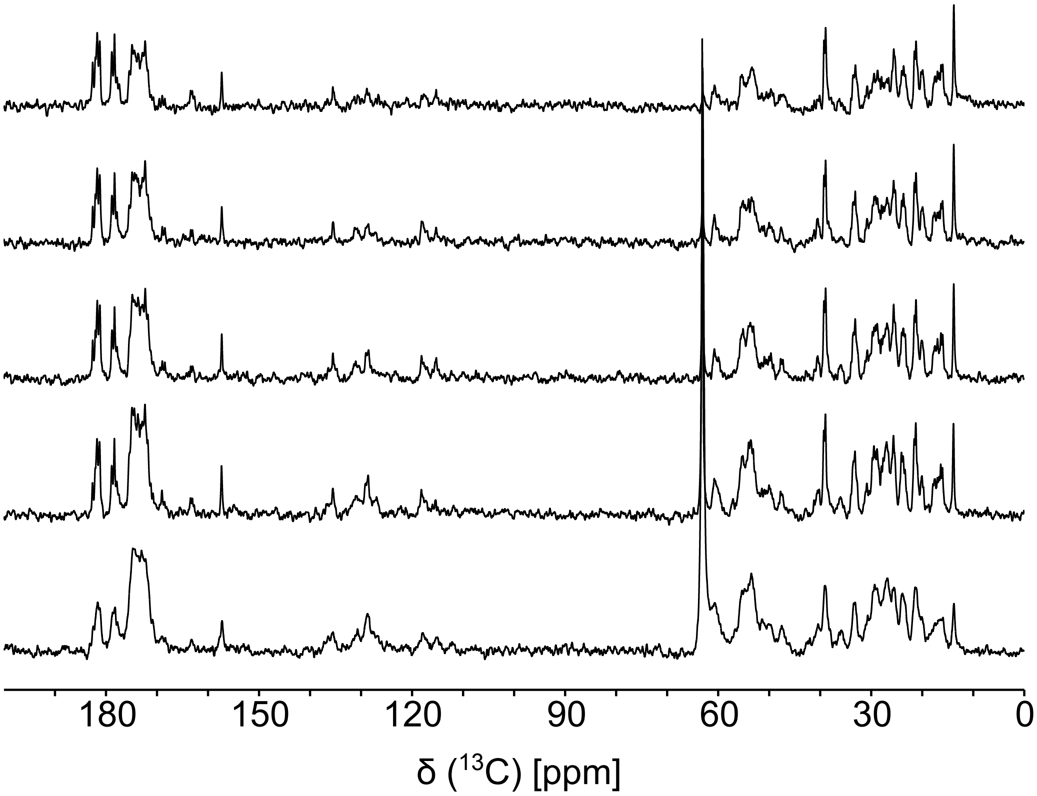


Figure 4.16: Series of spectra of [^{13}C, 70%–^{2}H] – p27 hyperpolarized on ^{13}C nuclei and dissolved using 100 mM Tris buffer, pH 7.2 containing 300 mM sodium chloride and 5 mM dithiotritol (DTT). The spectra were measured using a variable flip angle with ^{2}H decoupling and solvent suppression at 63.2 ppm corresponding to ethylene glycol resonance. The spectrum closest to the chemical shift axis corresponds to the first scan of the experiment and the spectrum farthest from the chemical shift axis, the last scan.

by hyperpolarizing the intrinsically disordered protein and mixing with the hyperpolarized protein with its biological target to monitor the folding process. To this end, in a second experiment, the hyperpolarized sample was dissolved and mixed with 50 μl of the provided CDK2-Cyclin A complex in the sample tube prior to NMR measurement. Multiple scans were then measured using variable flip angle excitation.

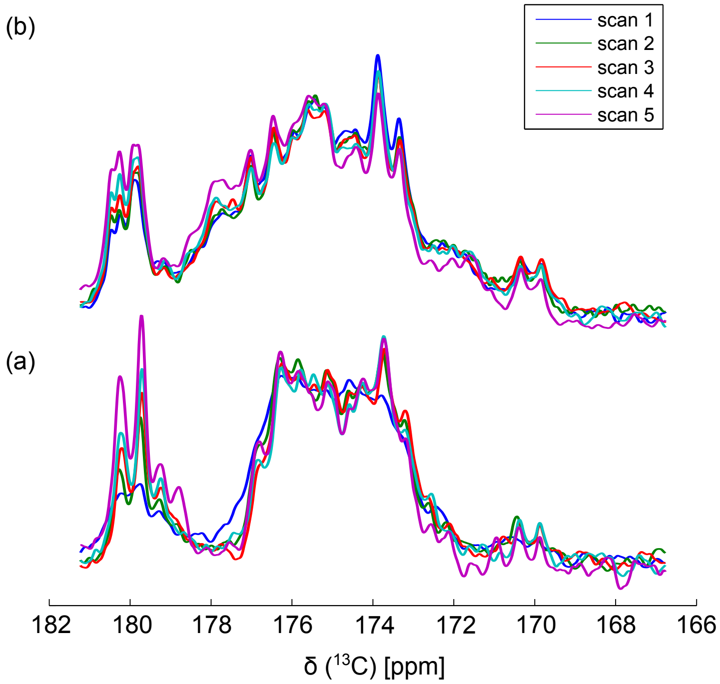


Figure 4.17: DNP-NMR spectra from (a) control experiment (same data as in Figure 4.16) and from (b) experiment where p27 is mixed with CDK2/Cyclin-A complex showing the carbonyl region. Each scan of the experiment is normalized to the same total integral in the 167 and 181.5 ppm. Peaks above 179 ppm correspond to sidechain carbonyl resonances.

The spectra obtained from the two dissolution DNP experiments were analyzed as follows. Signals from the carbonyl region (167 – 181.5 ppm) of each scan of both experiments were normalized to a total integral of 1. Carbonyl region of the normalized spectra from the control experiment (no mixing) and mixing experiment are shown in Figure 4.17 (a) & (b), respectively. Peaks above ~179 ppm are likely to correspond to sidechain carbonyl resonances.

In order to identify potential chemical shift changes, difference spectra were computed using the normalized spectra. Figure 4.18 (a) shows the difference calculated by subtracting the normalized first scan spectrum from normalized spectra of every scan of the control experiment. The same analysis was repeated with data from mixing experiment as shown in Figure 4.18 (b). The differences in the carbonyl region of the DNP-NMR spectra between the control and mixing (Figure 4.18 (a) & (b)) experiment could indicate changes in the structure of p27 upon binding CDK2/Cyclin complex. More specifically, the peaks in the downfield region (~178 ppm) in the mixing experiment suggest the formation of α–helix, which would be consistent with what has been shown in the literature. However, most of the signal at ~178 ppm is already present in the first scan (Figure 4.17 (b)), perhaps because most of the structural change has already occurred at that time and could not be observed in this experiment. Still, some increase of presumably α–helical structure is visible as seen in Figure 4.18 (b). Difference spectra were also calculated between the two experiments by subtracting each scan of the control experiment from the corresponding scan of the mixing experiment (Figure 4.18 (c)). It can be seen here that the differences in the downfield region (~178 ppm) show a relative increase in the later scans suggesting the formation of α–helix elements in p27.

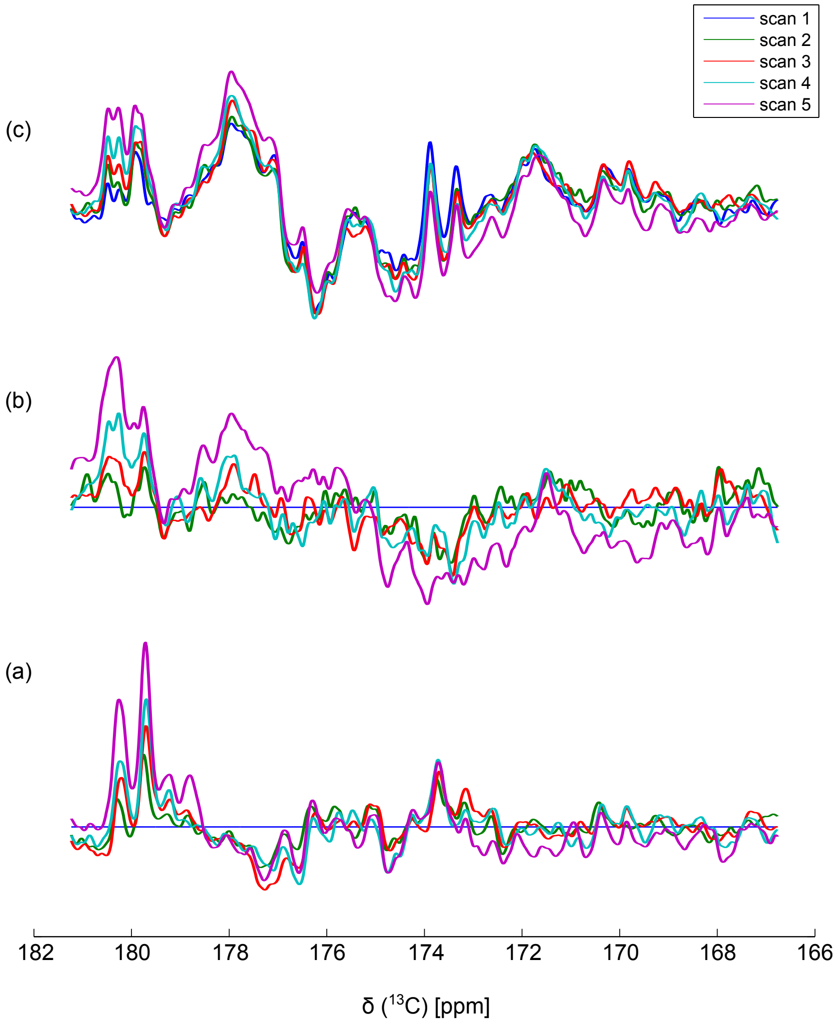


Figure 4.18: Difference spectra showing chemical shift changes in ^{13}C spectra of p27 in (a) control experiment and (b) mixing experiment (hyperpolarized p27 is mixed with CDK2/Cyclin-A complex). (c) Difference spectra between the control and mixing experiment for each scan. Difference spectra shown in (a) & (b) were computed by subtracting the first scan data from each scan of the same experiment. In (c), differences were calculated by subtracting data in each scan of the control experiment from the corresponding scan of the mixing experiment.

4.5 Conclusion

Surface tension of the protein plays a very important rule in the choice of solvent used in dissolution DNP experiments. Factors that affect the surface tension of the protein - pH, presence of salts and temperature, etc. - have been well established for almost 100 years [160]. Studies have also shown that foaming in protein solutions depend not on the surface tension of the protein solution but on the rate of change of surface tension as the protein is solubilized [161]. Such a dependence, when combined with a complete absence of literature on the relationship between the protein sequence and its surface tension, makes it very difficult to determine the type of solvent to be used for sample transfer in dissolution DNP.

Despite these potential difficulties, it is clear based on the results presented above that dissolution DNP enhanced NMR spectroscopy can be used to study protein folding events. Two state folding of L23 and potential structual changes associated with binding of p27 to CDK2/Cyclin-A have been observed. It is of interest that these experiments were possible using just uniformly isotopically enriched samples. Further, site selection obtained by employing selective multiple quantum coherence experiment has demonstrated one approach to improve site resolution in the 1D ^{13}C spectra.

While this study utilized a protein (L23), which reportedly follows the two-state model of protein folding, it is conceivable that proteins with complicated protein folding pathways may also be investigated. Improvements to the sample transfer step that could broaden the application range might be to use a flow setup where the sample is driven using liquid as described in the literature [162].

One approach to study protein–protein interactions was already demonstrated using p27 in this thesis. Another strategy to tackle protein–protein interactions using dissolution DNP would be to employ a flow based system (similar to [162]), where the hyperpolarized protein is flown over an immobilized binding partner. The binding partner could be immobilized in a setup as simple as Ni-affinity beads or in a elaborate setup similar to those used in surface plasmon resonance experiments. A flow based strategy may also be amenable to study protein–lipid interactions. In these experiments, changes in the 1D ^{13}C spectrum could be observed by measuring a series of spectra.

5. ANALYSIS OF DNP NMR DATA USING NON-NEGATIVE MATRIX FACTORIZATION

5.1 Introduction

Dissolution DNP NMR offers the ability to gain temporal resolution in addition to spectral resolution. This enables observation of transient species in chemical processes which can otherwise not be observed in NMR experiments. However, this presents a new challenge - how can the peaks from transient species in the NMR spectra be identified and how best to determine the contribution of each species to the observed signal? Of course, in the best case scenario, the resonances corresponding to the “intermediates” will be unique and the appearance of the new peaks may be followed. Even in these cases, identification of low abundance species may prove to be challenging since signals from these species might be low even with hyperpolarization. While conventional NMR experiments utilize signal averaging as needed to record spectra, this is not possible in the case of dissolution DNP experiments, since the polarization of spins generated by the DNP process is non-renewable in the NMR spectrometer. One approach towards identification of transient species would be to take advantage of machine learning algorithms to analyze NMR spectra from hyperpolarization experiments.

A branch of artificial intelligence, machine learning describes development of systems that can “learn” from the data provided and provide outputs or can “learn” from the data provided to construct models using which new data may be analyzed. The classic problem that is presented in evaluation of machine learning algorithms is the “Cocktail party problem” [163, 164]. The simplest formulation of this problem can read as “how does one identify what a single person is stating when more than one person is speaking at the same time?”. This problem stipulates that there may be overlaps in both time and frequency from which unique intelligible outputs must be generated.

Two of the many different approaches undertaken under the purview of machine learning are (i) supervised learning and (ii) unsupervised learning. While supervised learning outcomes are generally mapping pattern(s) between the provided input and generated output, in unsupervised learning inputs are provided to the algorithm without expected output. In effect, unsupervised learning procedures could be considered as pattern recognition schemes. In supervised learning, given a complete set of inputs and corresponding measured outputs, outputs for a given new set of data are predicted without the need for experimental measurements [165]. A good example of supervised machine learning is the spam filtering in e-mail [166, 167].

In the case of unsupervised learning, the only "input" presented to the learner (the computer performing analysis) is the experimental data. Without providing any further instructions other than an algorithm that specifies the convergence rules, the learner derives the output(s) for the given set of data. Unsupervised learning in the form of clustering analysis has found extensive application in the field of data mining and in genomics using a wide range of clustering algorithms. For example, one of the popular DNA sequence analysis tool, Clustal Omega, uses the clustering analysis algorithm called *k-means* [168]. Weighted correlation network analysis (or Weighted Gene Co-expression Network Analysis (WGCNA)) is extensively used in studying gene expression data. This method has been used successfully to identify a over-expression marker in glioblastoma [169] and to identify age-dependent changes in DNA methylation patterns in human brain and blood tissue [170].

Blind source separation (BSS) is another approach to unsupervised learning, which is used predominantly in applications where the dimensionality of data needs to be reduced. In other words, BSS is used to identify different features (sources) that constitute the experimental data. Several approaches including principal component analysis

(PCA) [171], independent component analysis (ICA) [172], singular value decomposition (SVD) [173] and non-negative matrix factorization [174] have been described. PCA and its statistical derivative, Probabilistic PCA (PPCA) have been extensively used in analysing metabolomics data [175, 176, 177]. ICA and modern implementations such as FastICA [178] have been used in pattern recognition (specifically palmprint recognition) [179] and as a tool for noise reduction in electrocardiogram data [180].

5.2 Methodology

5.2.1 Non-Negative Matrix Factorization

Non-negative matrix factorization (NNMF), also referred to as positive matrix factorization and non-negative matrix approximation, is an algorithm to extract components from the non-negative experimental data [181, 174]. Let V be the experimental data with size $m \times n$ where m is the number of data points in a single measurement and n, the number of measurements. Let p be the number of species that can contribute to the observed signal. Then, this data can be represented as

$$V \approx W \times H \tag{5.1}$$

where, the matrix W represents the sources of size $(m \times p)$ that contribute to the signal. These sources then correspond to co-efficients H of size $(p \times n)$. The symbol × in the equation denotes matrix multiplication.

NNMF is a numerical procedure to extract W and H from the data V. The steps involved in carrying out this procedure is presented as a flowchart in Figure 5.1 [181]. This factorization of the data matrix must be carried out to minimize the error between the reconstructed matrix (V_{recon}) and initial data matrix (V). The mostly used error function, in this application is the euclidean distance norm [182]. The euclidean distance norm is defined as

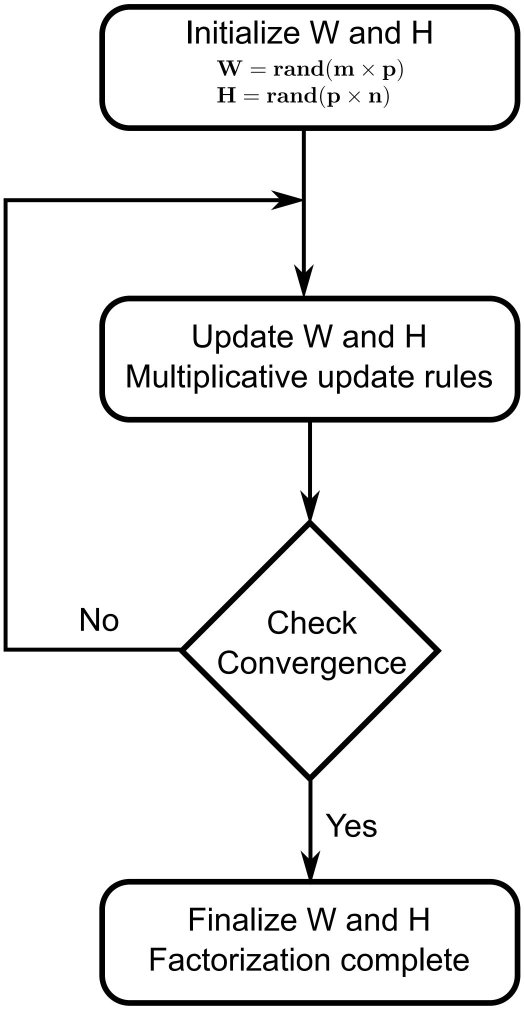


Figure 5.1: Flow chart showing steps involved in Non-Negative Matrix Factorization of experimental data. Multiplicative update rules are shown in equations 5.3 and 5.4 and convergence criterion are shown in equation 5.2.

$$E(W,H) = \parallel V - W{\times}H \parallel^2 = \sum_{i,j} (V_{ij} - (W{\times}H)_{ij})^2 \tag{5.2}$$

In this work, the magnitude of change of euclidean norm between consecutive iterations is used as the "convergence" criterion.

Several update rules for W and H have been proposed in the literature [183]. Of those, the alternating least squares (ALS) update rule and the multiplicative update rule have been reported to be most robust [182]. A desired characteristic of any of these approaches would be that the output is a sparse representation of the data. In other words, the data should be completely described by as few non-zero entries as possible. In this regard, although the ALS update rule offers faster performance and greater sparsity, it can lead to local minima rather than the global minimum due to its greater stress on sparsity [183]. The multiplicative update rule is shown in equations 5.3 & 5.4. This rule does not have problems associated with ALS update rules, although it is comparatively slower. In this work, multiplicative update rules have been used in the implementation of NNMF.

$$W \leftarrow W \cdot \frac{VH^T}{WHH^T + 10^{-9}} \tag{5.3}$$

$$H \leftarrow H \cdot \frac{W^T V}{W^T WH + 10^{-9}} \tag{5.4}$$

Here, the constant 10^{-9} in both equations are present to avoid calculations resulting in division by zero.

5.2.2 Dynamic Nuclear Polarization

Malic acid and oxaloacetatic acid (both from Amresco Inc., Solon, OH) were used without any further purification. 500 mM solutions of malic acid and oxaloacetic acid were prepared in 60%(v/v) ethylene glycol/40% water glassing mixture. The solutions were mixed with OX-63 (final concentration - 15 mM) and Gd-DTPA (final concentration - 1 mM) and used for ^{13}C hyperpolarization. The samples were irradiated with 60 mM microwaves at 93.974 GHz for 3 hours at 1.4 K and subsequently dissolved using 50 mM potassium phosphate, pH 7.4 and NMR spectra were measured (with ^{1}H decoupling) at 298 K using a BBO probe in an NMR spectrometer with 9.4 T magnetic field corresponding to 400 MHz ^{1}H Larmor frequency (Bruker, Billerica, MA).

Enzymatic conversion oxaloacetate to malate was carried out using porcine heart malate dehydrogenase (Sigma Aldrich, St. Louis, MO). In this experiment, oxaloacetatic acid was hyperpolarized as described above. The solution containing polarized oxaloacetic acid was mixed with 40 mM of Nicotinamide adenine dinucleotide (NADH) (Sigma Aldrich, St. Louis, MO) and 200 μM of malate dehydrogenase already present in the NMR tube. The reaction was monitored by measuring a series of 16 spectra using variable flip angle pulse with ^{1}H decoupling at 298 K over a period of 10 seconds. All spectra were referenced against DSS.

5.2.3 Data Analysis

NMR spectra were processed using Topsin 3.1 (Bruker, Billerica, MA) and subsequent data analysis using NNMF was carried out using MATLAB (Mathworks, Natick, MA). For simulations, ^{13}C spectra were calculated using a Lorentzian function with 20 Hz linewidth. All spectra (DNP experiment or simulated) were first scaled such that the highest intensity in the dataset is unity. The input values for data analysis using the NNMF algorithm were absolute values of the rescaled data and the number of expected sources.

5.3 Results and Discussion

5.3.1 Analysis of DNP Enhanced NMR Data using NNMF

Figure 5.2 shows the time resolved spectra of reduction of oxaloacetic acid to malic acid. The spectra show decay of signal intensity of peaks corresponding to oxaloacetic acid and appearance of peaks from malic acid. It must be noted here, that the signal decay is due to spin relaxation, effect of radiofrequency pulses and the chemical reaction. Scans 7 –11 show clearly discernible peaks from both the species present in the solution, whereas by scan 12 almost all signal has relaxed.

This data was used for analysis using NNMF. As described in equation 5.1, matrix V corresponds to the NMR data (matrix size, 32768×16 data points) shown in Figure 5.2. To analyze this data using the NNMF algorithm to separate two sources, the initial values of source (W) and coefficients (H) matrices were generated randomly with 32768×2 and 2×16 entries, respectively. These matrices are updated using the multiplicative update rules shown in equations 5.3 & 5.4 until the convergence condition is met. The results from the separation of sources (here, peaks from oxaloacetic acid and malic acid) is shown in Figure 5.3 (a). It can be seen that reference spectrum 1 also contains (<10%) contributions from peaks representing the second species. This contribution, however, is random and is variable if the calculations are repeated multiple times.

Based on the convergence condition described in equation 5.2, the reconstructed matrix WH should be very similar to the experimental data. The comparison between the spectra from the experiment and reconstruction is presented in Figure 5.3. Reconstructed spectra from the other scans agree with the measured spectra (not shown). Scans 1 and 7 represent two points in the reaction when the sample contains predominantly oxaloacetic acid and measurable amount of malic acid respectively. It can be seen from Figure 5.3 (b) & (c) that the reconstruction faithfully reproduces these time points in the reaction.

Additionally, the robustness of the algorithm was tested by specifying three sources for the same dataset. In equation 5.1, the source matrix, W, has size $(m \times p)$ where p is the number of sources. To test the robustness of separation, the number of sources, p, was set to 3 in this case. This procedure returned a third source with just noise (Figure 5.4).

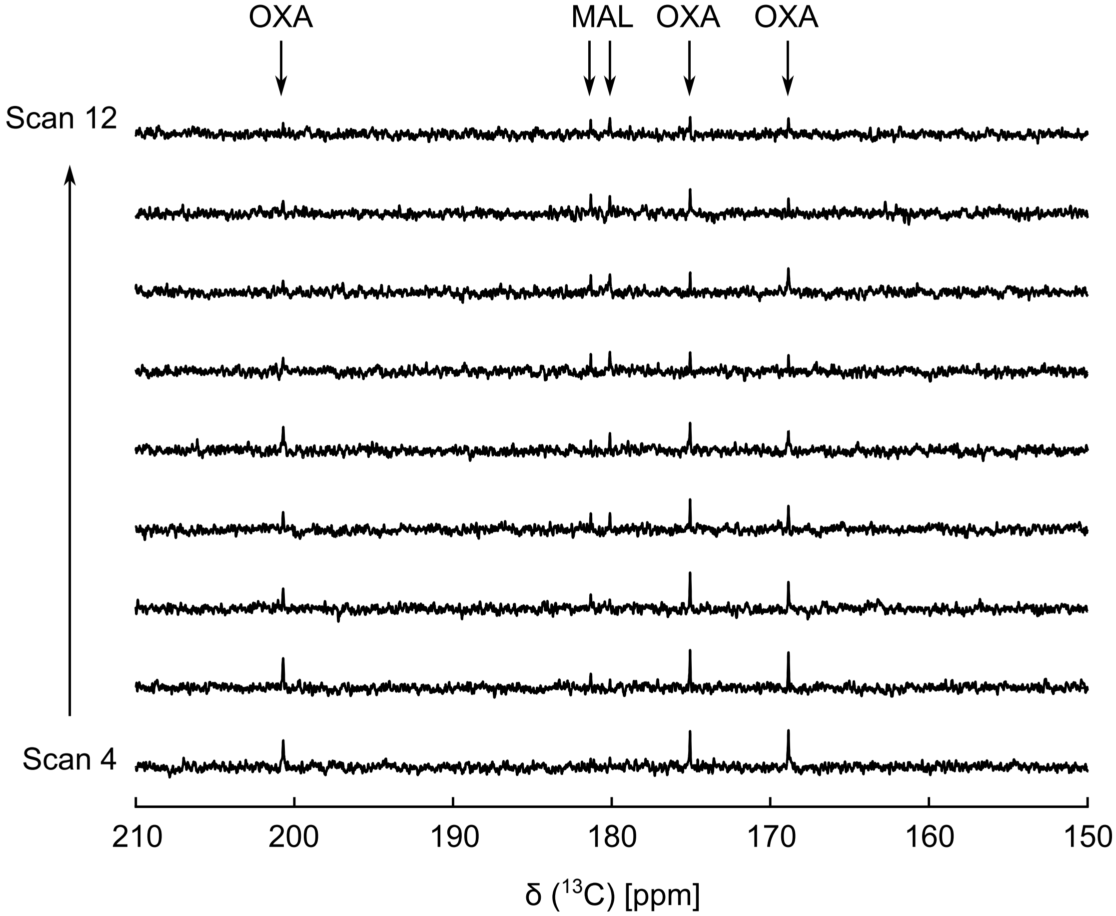


Figure 5.2: Series of low signal-to-noise spectra showing the reduction of oxaloacetic acid to malic acid. Spectra corresponding to scans 4 to 12 (from a total of 16 scans) are shown. Resonances from malic acid and oxaloacetic acid are indicated as MAL and OXA respectively. ^{1}H decoupling was applied during acquisition with a strength of 2.5 kHz. It should be noted that the experimental conditions were not optimized for measuring spectra with the best possible S/N.

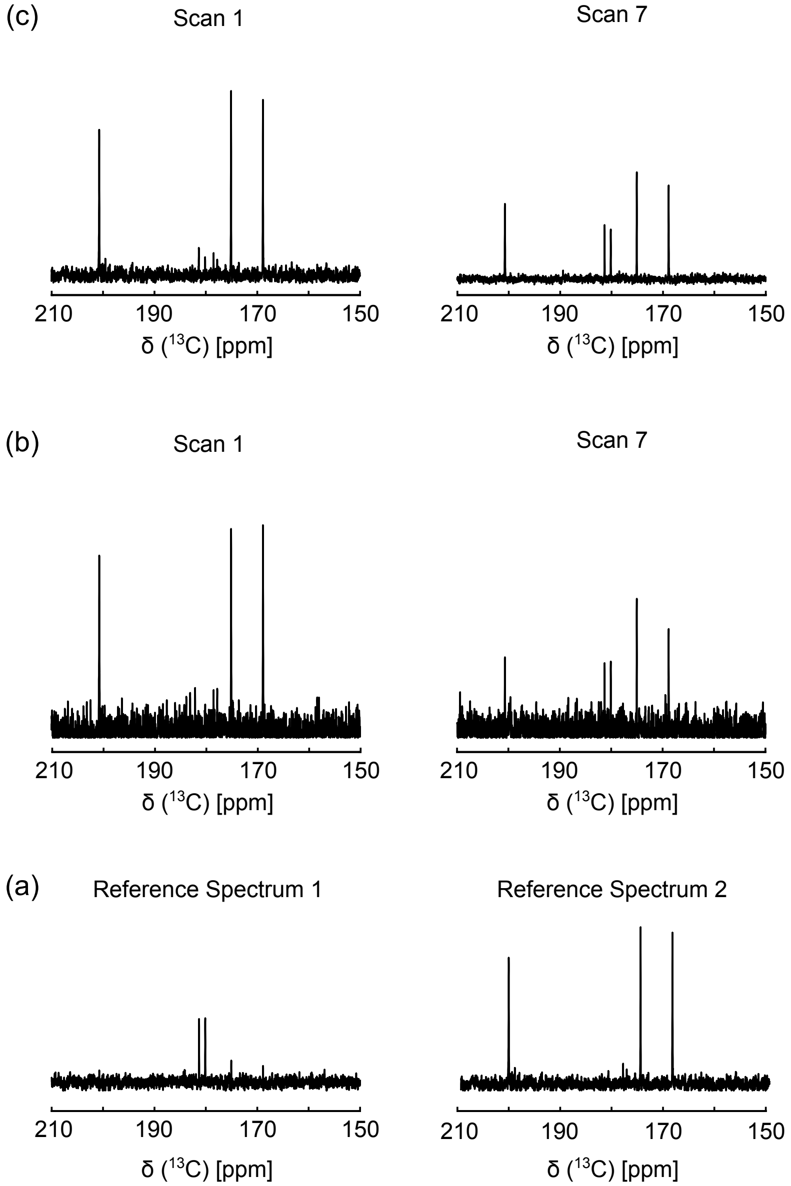


Figure 5.3: (a) Reference spectra obtained using NNMF. Reference spectrum 1 and 2 refer correspond to CO resonances from malic acid and oxaloacetic acid, respectively. The identity of the two spectra are random and may be reversed in another attempt. (b) & (c) Comparison between spectra obtained from dissolution DNP-NMR experiment (b) and spectra reconstructed after analysis using NNMF (c). Spectra from scans 1 and 7 are shown where scan 1 shows peaks from hyperpolarized oxaloacetic acid while scan 7 shows peaks from both oxaloacetic acid and malic acid generated from the reaction.

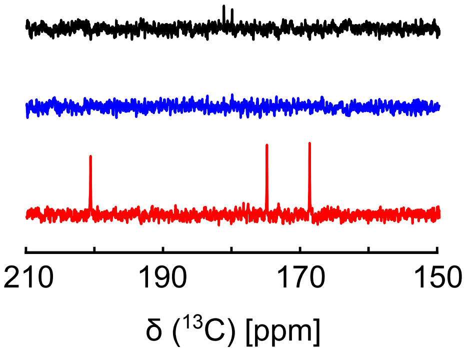


Figure 5.4: Output from NNMF algorithm when subjected to extracting three references instead of two. References in black and red correspond to malic acid and oxaloacetic acid peaks.

5.3.2 Effect of Noise

It is of interest that the reconstructed spectra have significantly higher signal to noise ratio (standard deviation of noise is ∼2 times lower in the reconstructed spectra) when compared to the measurement (Figure 5.3 (b) & (c)). This aspect of the NNMF and spectral reconstruction is of particular significance in DNP enhanced NMR experiments.

To evaluate the effect of noise in the spectra, a dataset with four peaks each for reactants and products were simulated, while ignoring the effect of relaxation. The reaction was simulated for a period of 2 seconds with 6, 16 and 32 scans. The reaction rate constant used in the simulations was 2 s^{-1}. Effect of noise in the spectrum was investigated by adding noise to generate four "noisy" time series of spectra and the performance of NNMF was evaluated using these datasets.

Initially, the dataset containing 6 spectra was reconstructed using the BSS extracted references as shown in Figure 5.5. In all cases, an improvement in S/N was observed in the spectra reconstructed after NNMF analysis including in (d), where the signals are almost indistinguishable from noise in the input data. In order to obtain an estimate of improvement in S/N in the reconstructed spectra, the standard deviation of noise was compared between the simulated spectra and reconstructed spectra. The standard deviation of noise was computed from the absolute value data using the definitions

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{5.5}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5.6}$$

Here, s is the standard deviation, n is the number of points in the region over which noise is calculated, x_i is the i^{th} data point and $\overline{x}$ denotes the calculated mean over the region.

For the simulation with data from 6 scans, Table 5.1 shows that the standard deviation of noise in the case of reconstructed spectra is at least 30% lower than the input data. It is noted here that the standard deviation calculated in this way using absolute value data would be different from the standard deviation calculated from non-absolute value data. While the above definition was used here for simplicity, other approaches to estimate noise in the absolute value spectra may also be employed as reported in the literature [184, 185].

In typical kinetics experiments using dissolution DNP, the number of transients used in the measurement is usually higher. Therefore, the effect of higher numbers of scans on the improvement of signal to noise in the reconstructed spectra was evaluated using datasets with 16 scans and 32 scans. The NNMF algorithm was used to extract sources and reconstruct each of these datasets. Table 5.1 compares the standard deviation of noise between simulations of three datasets with 6 scans, 16 scans and 32 scans. In general, later scans in the simulations, which have higher noise, show greater (and similar) improvement in S/N in both the 16 and 32 scans data when compared with the 6 scans data.

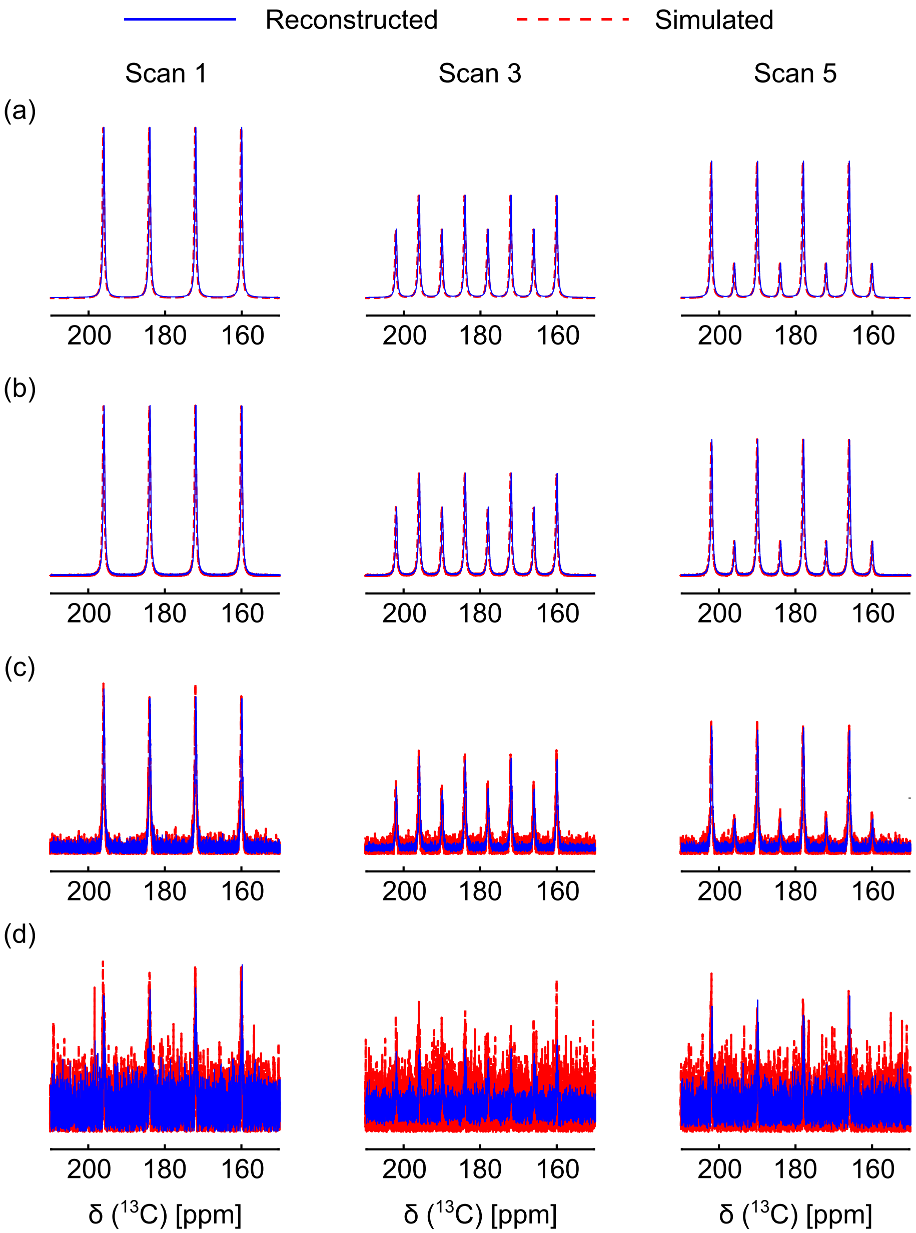


Figure 5.5: Comparison between simulated and NNMF reconstructed spectra. Scans 1, 3 and 5 (out of six scans total), corresponding to scenarios of S/N (a) 100, (b) 50, (c) 25 and (d) 10 are shown. “Simulated spectra” plotted here refers to the intensity rescaled data that is provided as input to the NNMF algorithm.

Table 5.1: Comparison of S/N ratio improvement between simulated and reconstructed spectra for different noise levels

Data	Standard Deviation of Noise					
	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6
6 scans						
S/N ratio = 25						
Simulated	0.0234	0.0234	0.0236	0.0236	0.0236	0.0241
Reconstructed	0.0169	0.0128	0.0099	0.0099	0.0130	0.0169
S/N ratio = 10						
Simulated	0.0925	0.0913	0.0909	0.0893	0.0888	0.0916
Reconstructed	0.0612	0.0522	0.0394	0.0395	0.0498	0.0636
16 scans						
S/N ratio = 25						
Simulated	0.0251	0.0247	0.0247	0.0244	0.0238	0.0251
Reconstructed	0.0115	0.0106	0.0095	0.0085	0.0076	0.0070
S/N ratio = 10						
Simulated	0.0951	0.0957	0.0927	0.0925	0.0921	0.0949
Reconstructed	0.0419	0.0409	0.0344	0.0325	0.0311	0.0273
32 scans						
S/N ratio = 25						
Simulated	0.0246	0.0249	0.0244	0.0245	0.0247	0.0249
Reconstructed	0.0083	0.0082	0.0078	0.0074	0.0071	0.0067
S/N ratio = 10						
Simulated	0.0953	0.0982	0.0984	0.0998	0.0982	0.0970
Reconstructed	0.0316	0.0319	0.0298	0.0294	0.0274	0.0267

5.3.3 Effect of Spin Relaxation

The applicability of dissolution DNP experiments is often limited by the spin lattice relaxation time of the analytes. It follows that the sites most amenable to study are those that have the longest relaxation times. However, the choice of sites for observation is often dictated by the property under investigation rather than spectroscopic considerations. An approach that improves signal to noise (S/N) ratio would be welcome in these circumstances and data analysis using NNMF seems to be a promising approach on this front.

Analysis of DNP enhanced NMR spectra must take into account the different spin lattice relaxation times of various sites on the molecule of interest. To mimic the situation where different peaks in the spectra have varying relaxation times, two situations were considered. In the first case, spin lattice relaxation rate constants used for simulating the spectra were smaller than the reaction rate constant ($R_1 < k$; Figure 5.6(a)) while in the second case relaxation rate constants are larger than the relaxation rate constants ($R_1 > k$; Figure 5.6(b)).

This time series of spectra was subjected to the NNMF algorithm and sources were separated. In the first case ($R_1 < k$), the calculated reference spectra contain all the expected resonances with only minor artifacts as seen in Figure 5.6(b). However, in the case where $R_1 > k$, the performance of the algorithm breaks down with one reference spectrum containing all the peaks in the dataset. It can be seen in this case that only two out of total six scans in the simulate data show peaks since the signal decay happens too quickly and factorization of such datasets may not be possible using NNMF.

Based on these simulations, it can be said that in cases, where peaks from different species are present in most of the scans, separation of sources can be accomplished albeit with minor errors. However, when the input data is missing information, like in the case where $R_1 > k$, the algorithm breaks down. As would be expected from a factorization

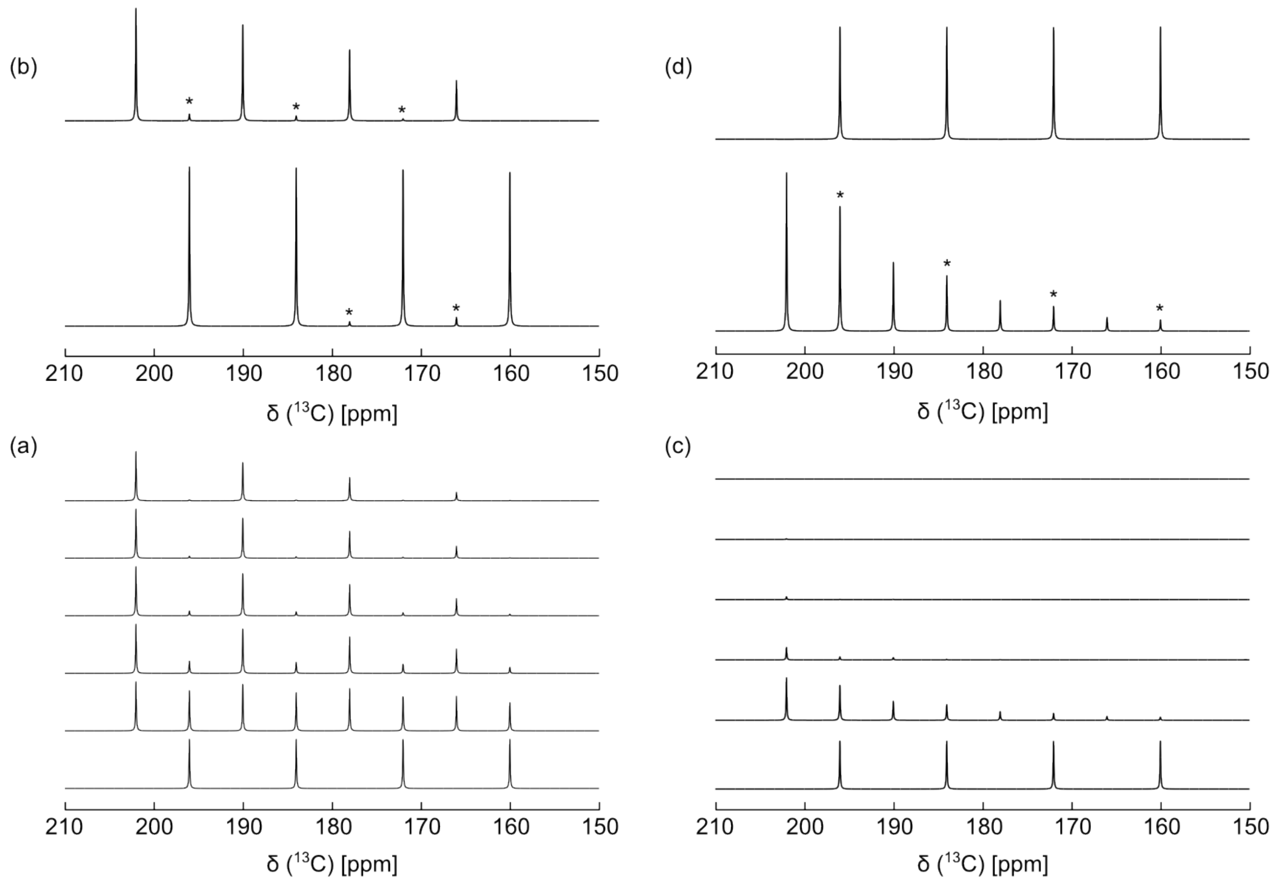


Figure 5.6: Simulations carried out to evaluate the effect of spin relaxation on NNMF analysis. Spectra were simulated with reaction rate constant of k=2 s^{-1}. The relaxation rate constants used in the simulation for 4 pair of peaks 160 ppm & 166 ppm, 172 ppm & 178 ppm, 184 ppm & 190 ppm and 196 ppm & 202 ppm were (a) 1 s^{-1}, 0.5 s^{-1}, 0.25 s^{-1} and 0.125 s^{-1} and (c) 10 s^{-1}, 8 s^{-1}, 6 s^{-1} and 4 s^{-1} respectively. (b) & (d) References extracted using NNMF from the data shown in (a) & (c). Spurious peaks present in references are marked with an asterisk ('*').

algorithm, datasets that can be constructed as a linear combination of two or more sources can be analyzed using NNMF. In other cases, the results may not be representative of the sources that make up the observed data.

In the context of BSS algorithms, it should be noted that the absolute intensities of the references and, by extension, the weighting factors obtained from BSS are not reliable since the intensities are scaled arbitrarily. This indeterminacy has been discussed elsewhere in the literature [186]. While neither the intensity nor the weighting factors may be utilized directly for quantitative analysis, source spectra may be integrated to obtain scaling factors for each analysis. Using the scaling factors thus obtained, it would be possible to scale the weighting factors to represent the concentrations of different species in the sample. It should be noted, that this approach is only feasible if each reference spectrum obtained from NNMF analysis is unique. In such cases, this approach may make it possible to use NNMF as a tool to extract quantitative information when applied to data from dissolution DNP experiments.

5.4 Outlook

An additional consideration in developing BSS techniques for analyzing DNP enhanced NMR data would be to enable handling data where significant overlap of resonances occur. Since multidimensional spectra are not always possible in dissolution DNP experiments, improvements in BSS techniques would enable analysis of spectra similar to those obtained in protein folding experiment presented in Chapter 4. One approach to get better separation of sources is the introduction of sparsity constraints [182]. Multiplicative update rules featuring sparsity constraints, as defined by Pauca and others in [187], are shown in equations 5.7 and 5.8 below.

$$W \leftarrow W \cdot \frac{VH^T}{WHH^T + \alpha W + 10^{-9}} \tag{5.7}$$

$$H \leftarrow H \cdot \frac{W^T V}{W^T W H + \beta H + 10^{-9}} \tag{5.8}$$

Here, the value of α and β are constants that enforce sparsity. This algorithm may carry greater utility in cases where significant overlap exists. An additional complication in using sparsity based rules is that they are inherently parametric and hence the choice of values of α and β affect the outcome, which would require additional systematic investigation. Methodological improvements that simultaneously improve noise reduction and source separation would be of tremendous use for DNP–NMR experiments. A model mathematical framework for such an improvement as applied to acoustic signals [188].

5.5 Conclusion

This work is a new application of NNMF to data obtained using dissolution DNP enhanced NMR spectroscopy. The most visible output of a blind source separation method like NNMF is undoubtedly the different components that make up the observed signal. NNMF was used to separate different components from a series of spectra and robustness of the algorithm was evaluated. Improvement in signal-to-noise ratio was observed in the spectra reconstructed from the matrices obtained from source separation. Several cases which are applicable to dissolution DNP experiments were simulated and the performance of NNMF algorithm was evaluated. In DNP experiments performed to identify and/or characterize reaction intermediates, separation of different components along with an improvement of S/N as obtained using NNMF is likely to be valuable tool for data analysis.

6. SUMMARY AND CONCLUSIONS

The strength of solution NMR spectroscopy as a technique to study macromolecular samples is the almost single atom resolution that is attainable. At the same time, it is also plagued by low sensitivity. To overcome the sensitivity problem, high sample concentrations or long acquisition times or both become necessary. Dissolution DNP effectively addresses the sensitivity problem by providing large signal enhancements. A large number of reports on dissolution DNP–NMR experiments with small molecules are available. The work reported in this thesis is the direct observation of proteins using this technique. By virtue of high S/N obtained in the reported experiments, signal averaging is not needed and it becomes possible to harness the full potential of NMR spectroscopy to gain both temporal and spectral resolution.

Dissolution DNP entails dissolving samples hyperpolarized in the solid state and transferring into an NMR tube for measurement of signal in solution state. Tendency of proteins to foam when subjected to vigorous mixing leads to broad signals in NMR spectra. Binary mixtures of water and organic solvents were explored in this work to reduce foaming and measure ^{13}C spectra from hyperpolarized protein samples. Samples of ribosomal protein L23, were hyperpolarized on the ^{13}C nuclei and more than 500 fold signal enhancement in comparison with a non-hyperpolarized spectrum was obtained. A theoretical ^{13}C spectrum of L23 was computed using published random coil chemical shifts. Comparing the spectra from dissolution DNP experiment with the theoretical spectrum, identity of resonances were obtained which were used to correlate the observed enhancement values with the relaxation properties of specific sites as measured using a multiple scan experiment. ^{1}H hyperpolarization of an unlabeled peptide, bacitracin, was shown to provide 30 and 45 fold enhancement for amide and aliphatic protons respectively.

L23, which follows two state model of protein folding, was used as a model system to develop the technique to monitor protein folding by measuring a series of time-resolved ^{13}C spectra after a pH jump from a single hyperpolarized sample. Protein folding was followed by observing the changes in the carbonyl region of the spectra. Using statistical distribution of ^{13}C chemical shifts, fractions of different secondary structure elements were estimated for each transient. $C^{\delta 1}$ resonances of the five isoleucine residues in L23 were resolved as refolding of L23 occurred.

Because of the non-renewability of spin polarization, DNP–NMR experiments are often limited to measuring 1D spectra. These 1D spectra are often crowded and site resolution is lost. To improve this situation, a selective multiple quantum coherence experiment is demonstrated. This spectroscopic approach is attractive particularly in scenarios where biochemical strategies like selective amino acid labeling or labeling specific site using metabolic precursors are problematic. Finally, ^{13}C spectra from hyperpolarized p27, an intrinsically disordered protein, were measured where the sample transfer was carried out using an aqueous buffer instead of binary mixtures. Hyperpolarized p27 was mixed with CDK-2/Cyclin-A complex and chemical shift changes potentially reporting on structural changes in p27 upon binding were observed.

The temporal resolution offered by DNP experiments enables identification of intermediates in reactions under investigation. However, reference spectra for transient species are difficult to obtain. Further, the population of transient species may be low, resulting in peaks with low S/N ratio. To analyze such data, performance of a machine learning algorithm was evaluated. Non-negative matrix factorization, a blind source separation algorithm, was used to decompose time resolved NMR spectra from the enzymatic conversation of oxaloacetate to malate into different sources that make up the composite signal in each transient. Simulations evaluating various conditions such as the effect of noise in the spectra and variations in the relaxation times of different resonances were carried out and performance of the algorithm evaluated. It has been determined that in many cases, this algorithm shows robust performance and shows tremendous potential for analyzing DNP data.

In summary, results from experiments using dissolution DNP enhanced NMR spectroscopy to observe proteins directly were presented. The ability to perform time resolved measurements with near atomic resolution could be beneficial in gaining understanding of many transient process encountered in biological systems. Since NMR measurements are carried out in solution state, conditions close to those encountered *in vivo* may be reproducible in experiments using this technique.

In another application, interactions of different proteins with cell surface receptors could be carried out in a similar setup as described earlier with the exception of using live cells. It might even be possible to map the interaction interfaces using selective isotopic labeling of the protein being hyperpolarized. In the case of uniformly labeled proteins, the orientation of the protein at the time of binding and any interactions with the lipids during this interactions could also be obtained simultaneously.

There are many more interesting problems that could be studied using dissolution DNP. Numerous facets of biological systems have evaded understanding, sometimes because of lack to tools to study these systems. It is hoped that the work in this thesis provides another set of tools to approach the complexities of biological systems. It is also hoped that this work provides an impetus to use dissolution DNP as a technique to study macromolecules and develop strategies to gain insight into some of the many unanswered questions.

REFERENCES

[1] Timothy D. W. Claridge. *High-resolution NMR Techniques in Organic Chemistry*. Elsevier, Amsterdam, Netherlands, 2009.

[2] Martin Billeter, Gerhard Wagner, and Kurt Wüthrich. Solution NMR structure determination of proteins revisited. *Journal of Biomolecular NMR*, 42:155–158, 2008.

[3] Stanley J. Opella. Structure determination of membrane proteins by nuclear magnetic resonance spectroscopy. *Annual Review of Analytical Chemistry*, 6:305–328, 2013.

[4] Gemma Comellas and Chad M. Rienstra. Protein structure determination by magic-angle spinning solid-state NMR, and insights into the formation, structure, and stability of amyloid fibrils. *Annual Review of Biophysics*, 42:515–536, 2013.

[5] Lewis E. Kay, Dennis A. Torchia, and Ad Bax. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28:8972–8979, 1989.

[6] Rieko Ishima and Dennis A. Torchia. Protein dynamics from NMR. *Nature Structural and Molecular Biology*, 7:740–743, 2000.

[7] Anthony Mittermaier and Lewis E. Kay. New tools provide new insights in NMR studies of protein dynamics. *Science*, 312:224–228, 2006.

[8] Matt A. Bernstein, Kevin F. King, and Xiaohong Joe Zhou. *Handbook of MRI Pulse Sequences*. Elsevier, Amsterdam, Netherlands, 2004.

[9] Michael P Williamson, Timothy F Havel, and Kurt Wüthrich. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology*, 182:295–315, 1985.

[10] John Cavanagh, Wayne J. Fairbrother, Arthur G. Palmer, Mark Rance, and Nicholas J. Skelton. *Protein NMR Spectroscopy: Principles and Practice*. Elsevier, Amsterdam, Netherlands, 2007.

[11] David S. Wishart, Brian D. Sykes, and Fredric M. Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of Molecular Biology*, 222:311–333, 1991.

[12] David S. Wishart and Brian D. Sykes. The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *Journal of Biomolecular NMR*, 4:171–180, 1994.

[13] Wolfgang Jahnke, Matthias Baur, Gerd Gemmecker, and Horst Kessler. Improved accuracy of NMR structures by a modified NOESY-HSQC experiment. *Journal of Magnetic Resonance*, 106:86–88, 1995.

[14] Martin. Karplus. Vicinal proton coupling in nuclear magnetic resonance. *Journal of the American Chemical Society*, 85:2870–2871, 1963.

[15] Christian Griesinger, Ole W. Sørenson, and Richard R. Ernst. Practical aspects of the E.COSY technique. measurement of scalar spin-spin coupling constants in peptides. *Journal of Magnetic Resonance*, 75:474–492, 1987.

[16] Christian Griesinger and Richard R. Ernst. Frequency offset effects and their elim-

ination in NMR rotating-frame cross-relaxation spectroscopy. *Journal of Magnetic Resonance*, 75:261–271, 1987.

[17] James H. Prestegard, Hashim M. Al-Hashimi, and Joel R. Tolman. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quarterly Reviews of Biophysics*, 33:371–424, 2000.

[18] Peter Guntert, Christian Mumenthaler, and Kurt Wüthrich. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology*, 273:283–298, 1997.

[19] Peter Guntert. Automated NMR structure calculation with CYANA. *Methods in molecular biology*, 278:353–378, 2004.

[20] Konstantin Pervushin, Roland Riek, Gerhard Wider, and Kurt Wüthrich. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*, 94:12366–12371, 1997.

[21] Michael Salzmann, Gerhard Wider, Konstantin Pervushin, and Kurt Wüthrich. Improved sensitivity and coherence selection for [^{15}N, ^{1}H]–TROSY elements in triple resonance experiments. *Journal of Biomolecular NMR*, 15:181–184, 1999.

[22] Michael Salzmann, Konstantin Pervushin, Gerhard Wider, Hans Senn, and Kurt Wüthrich. NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *Journal of the American Chemical Society*, 122:7543–7548, 2000.

[23] Dominique P Frueh, Andrew C Goodrich, Subrata H Mishra, and Scott R Nichols.

NMR methods for structural studies of large monomeric and multimeric proteins. *Current Opinion in Structural Biology*, 23:734–739, 2013.

[24] David Rovnyak, Dominique P. Frueh, Mallika Sastry, Zhen-Yu J. Sun, Alan S. Stern, Jeffrey C. Hoch, and Gerhard Wagner. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *Journal of Magnetic Resonance*, 170:15–21, 2004.

[25] Sven G. Hyberts, Haribabu Arthanari, and Gerhard Wagner. Applications of non-uniform sampling and processing. *Topics in Current Chemistry*, 316:125–148, 2012.

[26] Vitali Tugarinov, Voula Kanelis, and Lewis E. Kay. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*, 1:749–754, 2006.

[27] Vitali Tugarinov and Lewis E. Kay. Ile, leu, and val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel nmr methods. *Journal of the American Chemical Society*, 125:13868–13878, 2003.

[28] James H Prestegard, David A Agard, Kelley W Moremen, Laura A Lavery, Laura C Morris, and Kari Pederson. Sparse labeling of proteins: Structural characterization from long range constraints. *Journal of Magnetic Resonance*, 241:32–40, 2014.

[29] Jason R. Schnell, H. Jane Dyson, and Peter E. Wright. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annual Review of Biophysics and Biomolecular Structure*, 33:119–140, 2004.

[30] Phineus R. L. Markwick, Thrse Malliavin, and Michael Nilges. Structural biol-

ogy by NMR: structure, dynamics, and interactions. *PLoS Computational Biology*, 4:e1000168, 2008.

[31] David D. Boehr, H. Jane Dyson, and Peter E. Wright. An NMR perspective on enzyme dynamics. *Chemical Reviews*, 106:3055–3079, 2006.

[32] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society*, 104:4546–4559, 1982.

[33] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society*, 104:4559–4570, 1982.

[34] Francesca Massi, Michael J. Grey, and Arthur G. Palmer. Microsecond timescale backbone conformational dynamics in ubiquitin studied with NMR $R_{1\rho}$ relaxation experiments. *Protein Science*, 14:735–742, 2005.

[35] Vladislav Yu. Orekhov, Dmitry M. Korzhnev, and Lewis E. Kay. Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *Journal of the American Chemical Society*, 126:1886–1891, 2004.

[36] Ian R. Kleckner and Mark P. Foster. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(8):942–968, August 2011.

[37] Tanja Mittag, Brian Schaffhausen, and Ulrich L. Günther. Tracing kinetic intermediates during ligand binding. *Journal of the American Chemical Society*, 126:9017–9023, 2004.

[38] Thad G. Walker and William Happer. Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics*, 69:629–642, 1997.

[39] F William Hersman, Iulian C Ruset, Stephen Ketel, Iga Muradian, Silviu D Covrig, Jan Distelbrink, Walter Porter, David Watt, Jeffrey Ketel, John Brackett, Aaron Hope, and Samuel Patz. Large production system for hyperpolarized ^{129}Xe for human lung imaging studies. *Academic radiology*, 15:683–692, 2008.

[40] Jason C. Leawoods, Dmitriy A. Yablonskiy, Brian Saam, David S. Gierada, and Mark S. Conradi. Hyperpolarized ^{3}He gas production and MR imaging of the lung. *Concepts in Magnetic Resonance*, 13:277–293, 2001.

[41] Thomas J. Lowery, Sandra Garcia, Lana Chavez, E. Janette Ruiz, Tom Wu, Thierry Brotin, Jean-Pierre Dutasta, David S. King, Peter G. Schultz, Alex Pines, and David E. Wemmer. Optimization of xenon biosensors for detection of protein interactions. *ChemBioChem*, 7:65–73, 2006.

[42] Megan M. Spence, Seth M. Rubin, Ivan E. Dimitrov, E. Janette Ruiz, David E. Wemmer, Alexander Pines, Shao Qin Yao, Feng Tian, and Peter G. Schultz. Functionalized xenon as a biosensor. *Proceedings of the National Academy of Sciences of the United States of America*, 98:10654–10657, 2001.

[43] Jennifer M. Chambers, P. Aru Hill, Julie A. Aaron, Zhaohui Han, David W. Christianson, Nicholas N. Kuzma, and Ivan J. Dmochowski. Cryptophane xenon-129 nuclear magnetic resonance biosensors targeting human carbonic anhydrase. *Journal of the American Chemical Society*, 131:563–569, 2009.

[44] Johannes Natterer and Joachim Bargon. Parahydrogen induced polarization. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 31:293–315, 1997.

[45] Simon B Duckett and Christopher J Sleigh. Applications of the parahydrogen phenomenon: A chemical perspective. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34:71–92, 1999.

[46] Stefan Glöggler, Rafael Müller, Johannes Colell, Meike Emondts, Martin Dabrowski, Bernhard Blümich, and Stephan Appelt. Para-hydrogen induced polarization of amino acids, peptides and deuterium–hydrogen gas. *Physical Chemistry Chemical Physics*, 13:13759–13764, 2011.

[47] Mathias Haake, Johannes Natterer, and Joachim Bargon. Efficient NMR pulse sequences to transfer the parahydrogen-induced polarization to hetero nuclei. *Journal of the American Chemical Society*, 118:8688–8691, 1996.

[48] Lars T. Kuhn, Ute Bommerich, and Joachim Bargon. Transfer of parahydrogen-induced hyperpolarization to ^{19}F. *The Journal of Physical Chemistry A*, 110:3521–3526, 2006.

[49] Ralph W. Adams, Juan A. Aguilar, Kevin D. Atkinson, Michael J. Cowley, Paul I. P. Elliott, Simon B. Duckett, Gary G. R. Green, Iman G. Khazal, Joaqun Lopez-Serrano, and David C. Williamson. Reversible interactions with para-hydrogen enhance NMR sensitivity by polarization transfer. *Science*, 323:1708–1711, 2009.

[50] Kevin D Atkinson, Michael J Cowley, Paul I P Elliott, Simon B Duckett, Gary G R Green, Joaquin Lopez-Serrano, and Adrian C Whitwood. Spontaneous transfer of parahydrogen derived spin order to pyridine at low magnetic field. *Journal of the American Chemical Society*, 131:13362–13368, 2009.

[51] Patrick Habler and Joachim Bargon. In situ transfer of parahydrogen-induced nuclear spin Polarization–Structural characterization of hydrogenation intermediates. *Angewandte Chemie International Edition*, 39:3701–3703, 2000.

[52] Silvio Aime, W. Dastrú, R. Gobetto, A. Russo, A. Viale, and D. Canet. A novel application of para H_2: the reversible Addition/Elimination of H_2 at a Ru_3 cluster revealed by the enhanced NMR emission resonance from molecular hydrogen. *The Journal of Physical Chemistry A*, 103:9702–9705, 1999.

[53] Eduard Y. Chekmenev, Jan Hovener, Valerie A. Norton, Kent Harris, Lynne S. Batchelder, Pratip Bhattacharya, Brian D. Ross, and Daniel P. Weitekamp. PASADENA hyperpolarization of succinic acid for MRI and NMR spectroscopy. *Journal of the American Chemical Society*, 130:4212–4213, 2008.

[54] Stefan Glöggler, Meike Emondts, Johannes Colell, Rafael Müller, Bernhard Blümich, and Stephan Appelt. Selective drug trace detection with low-field NMR. *The Analyst*, 136:1566, 2011.

[55] J. Hore and R. W. Broadhurst. Photo-CIDNP of biopolymers. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 25:345–402, 1993.

[56] Christopher M. Dobson and Peter J. Hore. Kinetic studies of protein folding using NMR spectroscopy. *Nature Structural Biology*, 5:504–507, 1998.

[57] Robert Kaptein, Klaas Dijkstra, and Klaas Nicolay. Laser photo-CIDNP as a surface probe for proteins in solution. *Nature*, 274:293–294, 1978.

[58] K. Hun Mok, Lars T. Kuhn, Martin Goez, Iain J. Day, Jasper C. Lin, Niels H. Andersen, and Peter J. Hore. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, 447:106–109, 2007.

[59] Anatole Abragam and Maurice Goldman. Principles of dynamic nuclear polarization. *Reports of Progress in Physics*, 41:395–467, 1978.

[60] Albert W. Overhauser. Polarization of nuclei in metals. *Physical Review*, 92:411–415, 1953.

[61] Thomas R. Carver and Charles P. Slichter. Experimental verification of the Overhauser nuclear polarization effect. *Physical Review*, 102:975–980, 1956.

[62] Tom R. Carver and Charles P. Slichter. Polarization of nuclear spins in metals. *Physical Review*, 92:212–213, 1953.

[63] Lino R. Becerra, Gary J. Gerfen, Brenden F. Bellew, Jeff A. Bryant, Dennis A. Hall, Souheil J. Inati, Ralph T. Weber, Sun Un, Thomas F. Prisner, Ann E. Mcdermott, Ken W. Fishbein, Kenneth E. Kreischer, Richard J. Temkin, David J. Singel, and Robert G. Griffin. A spectrometer for dynamic nuclear polarization and electron paramagnetic resonance at high frequencies. *Journal of Magnetic Resonance*, 117:28–40, 1995.

[64] Nikolaus M. Loening, Melanie Rosay, Volker Weis, and Robert G. Griffin. Solution-state dynamic nuclear polarization at high magnetic field. *Journal of the American Chemical Society*, 124:8808–8809, 2002.

[65] H. G. Beljers, L. van der Kint, and J. S. van Wieringen. Overhauser effect in a free radical. *Physical Review*, 95:1683–1683, 1954.

[66] Christian Griesinger, Marina Bennati, Hans-Martin Vieth, Claudio Luchinat, Giacomo Parigi, Peter Höfer, Frank Engelke, Stefeen J. Glaser, Vasyl Denysenkov, and Thomas F. Prisner. Dynamic nuclear polarization at high magnetic fields in liquids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 64:4–28, 2012.

[67] Brandon D. Armstrong and Songi Han. Overhauser dynamic nuclear polarization to

study local water dynamics. *Journal of the American Chemical Society*, 131:4641–4647, 2009.

[68] Qing Zhe Ni, Eugenio Daviso, Thach V. Can, Evgeny Markhasin, Sudheer K. Jawla, Timothy M. Swager, Richard J. Temkin, Judith Herzfeld, and Robert G. Griffin. High frequency dynamic nuclear polarization. *Accounts of Chemical Research*, 46:1933–1941, 2013.

[69] Kan-Nian Hu, Hsiao-hua Yu, Timothy M. Swager, and Robert G. Griffin. Dynamic nuclear polarization with biradicals. *Journal of the American Chemical Society*, 126:10844–10845, 2004.

[70] Gary J. Gerfen, Lino R. Becerra, Dennis A. Hall, Robert G. Griffin, Richard J. Temkin, and David J. Singel. High frequency (140 GHz) dynamic nuclear polarization: Polarization transfer to a solute in frozen aqueous solution. *The Journal of Chemical Physics*, 102:9494–9497, 1995.

[71] Kan-Nian Hu, Changsik Song, Hsiao-Hua Yu, Timothy M Swager, and Robert G Griffin. High-frequency dynamic nuclear polarization using biradicals: a multi-frequency EPR lineshape analysis. *The Journal of chemical physics*, 128:052302, 2008.

[72] Changsik Song, Kan-Nian Hu, Chan-Gyu Joo, Timothy M. Swager, and Robert G. Griffin. TOTAPOL: a biradical polarizing agent for dynamic nuclear polarization experiments in aqueous media. *Journal of the American Chemical Society*, 128:11385–11390, 2006.

[73] Matthew K. Kiesewetter, Björn Corzilius, Albert A. Smith, Robert G. Griffin, and Timothy M. Swager. Dynamic nuclear polarization with a water-soluble rigid biradical. *Journal of the American Chemical Society*, 134:4537–4540, 2012.

[74] Eric L. Dane, Björn Corzilius, Egon Rizzato, Pierre Stocker, Thorsten Maly, Albert A. Smith, Robert G. Griffin, Olivier Ouari, Paul Tordo, and Timothy M. Swager. Rigid orthogonal bis-TEMPO biradicals with improved solubility for dynamic nuclear polarization. *The Journal of Organic Chemistry*, 77:1789–1797, 2012.

[75] R. A. Wind, M. J. Duijvestijn, C. van der Lugt, A. Manenschijn, and J. Vriend. Applications of dynamic nuclear polarization in ^{13}C NMR in solids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 17:33–67, 1985.

[76] Yonatan Hovav, Akiva Feintuch, and Shimon Vega. Theoretical aspects of dynamic nuclear polarization in the solid state - spin temperature and thermal mixing. *Physical Chemistry Chemical Physics*, 15:188–203, 2013.

[77] Dennis A. Hall, Douglas C. Maus, Gary J. Gerfen, Souheil J. Inati, Lino R. Becerra, Frederick W. Dahlquist, and Robert G. Griffin. Polarization-enhanced NMR spectroscopy of biomolecules in frozen solution. *Science*, 276:930–932, 1997.

[78] Patrick C. van der Wel, Kan-Nian Hu, J. Lewandowski, and R. G. Griffin. Dynamic nuclear polarization of amyloidogenic peptide nanocrystals: GNNQQNY, a core segment of the yeast prion protein Sup35p. *Journal of the American Chemical Society*, 128:10840–10846, 2006.

[79] Ioannis Gelis, Veronika Vitzthum, Neha Dhimole, Marc A. Caporini, Andreas Schedlbauer, Diego Carnevale, Sean R. Connell, Paola Fucini, and Geoffrey Bodenhausen. Solid-state NMR enhanced by dynamic nuclear polarization as a novel tool for ribosome structural biology. *Journal of Biomolecular NMR*, 56:85–93, 2013.

[80] Tomas Jacso, W. Trent Franks, Honor Rose, Uwe Fink, Jana Broecker, Sandro Keller, Hartmut Oschkinat, and Bernd Reif. Characterization of membrane proteins

in isolated native cellular membranes by dynamic nuclear polarization solid-state NMR spectroscopy without purification and reconstitution. *Angewandte Chemie International Edition*, 124:447–450, 2012.

[81] Marie Renault, Shane Pawsey, Martine P. Bos, Eline J. Koers, Deepak Nand, Ria Tommassen-van Boxtel, Melanie Rosay, Jan Tommassen, Werner E. Maas, and Marc Baldus. Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angewandte Chemie International Edition*, 51:2998–3001, 2012.

[82] Jan H Ardenkjaer-Larsen, Bjorn Fridlund, Andreas Gram, Georg Hansson, Lennart Hansson, Mathilde H Lerche, Rolf Servin, Mikkel Thaning, and Klaes Golman. Increase in Signal-to-Noise ratio of > 10,000 times in Liquid-State NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 100:10158–10163, 2003.

[83] Sean Bowen and Christian Hilty. Rapid sample injection for hyperpolarized NMR spectroscopy. *Physical Chemistry Chemical Physics*, 12:5766–5770, 2010.

[84] James Leggett, Robert Hunter, Josef Granwehr, Rafal Panek, Angel J. Perez-Linde, Anthony J. Horsewill, Jonathan McMaster, Graham Smith, and Walter Kockenberger. A dedicated spectrometer for dissolution DNP NMR spectroscopy. *Physical Chemistry Chemical Physics*, 12:5883–5892, 2010.

[85] Ronald Kühne, Tomas Schaffhauser, Alexander Wokaun, and Richard R Ernst. Study of transient chemical reactions by NMR: Fast stopped-flow fourier transform experiments. *Journal of Magnetic Resonance*, 35:39 – 67, 1979.

[86] Sydney D. Hoeltzli and Carl Frieden. Stopped-flow NMR spectroscopy: Real-time unfolding studies of 6-^{19}F- tryptophan-labeled *escherichia coli* dihydrofolate

reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 92:9318–9322, 1995.

[87] John J. Grimaldi and Brian D. Sykes. Stopped flow fourier transform nuclear magnetic resonance spectroscopy. application to the α-chymotrypsin-catalyzed hydrolysis of tert-butyl-l-phenylalanine. *Journal of the American Chemical Society*, 97:273–276, 1975.

[88] Matthew D. Christianson, Emily H. P. Tan, and Clark R. Landis. Stopped-flow NMR: Determining the kinetics of [rac-(C_2H_4(1-indenyl)$_2$)ZrMe][MeB(C_6F_5)$_3$]-catalyzed polymerization of 1-hexene by direct observation. *Journal of the American Chemical Society*, 132:11461–11463, 2010.

[89] Janina Buck, Boris Fürtig, Jonas Noeske, Jens Wöhnert, and Harald Schwalbe. Time-resolved NMR spectroscopy: Ligand-induced refolding of riboswitches. *Methods in Molecular Biology*, 540:161–171, 2009.

[90] Masaya Kakuta, Dimuthu A. Jayawickrama, Andrew M. Wolters, Andreas Manz, and Jonathan V. Sweedler. Micromixer-based time-resolved NMR: applications to ubiquitin protein conformation. *Analytical Chemistry*, 75:956–960, 2003.

[91] Pieter E. S. Smith, Kevin J. Donovan, Or Szekely, Maria Baias, and Lucio Frydman. Ultrafast NMR T_1 relaxation measurements: Probing molecular properties in real time. *ChemPhysChem*, 14:3138–3145, 2013.

[92] Lucio Frydman and Damir Blazina. Ultrafast two-dimensional nuclear magnetic resonance spectroscopy of hyperpolarized solutions. *Nature Physics*, 3:415–419, 2007.

[93] Paul Schanda, Eriks Kupce, and Bernhard Brutscher. SOFAST-HMQC experiments

for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR*, 33:199–211, 2005.

[94] Paul Schanda and Bernhard Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127:8014–8015, 2005.

[95] Paul Schanda and Bernhard Brutscher. Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR. *Journal of Magnetic Resonance*, 178:334–339, 2006.

[96] Sebastian Meier, Magnus Karlsson, Pernille R. Jensen, Mathilde H. Lerche, and Jens . Duus. Metabolic pathway visualization in living yeast by DNP-NMR. *Molecular BioSystems*, 7:2834, 2011.

[97] Mathilde H. Lerche, Sebastian Meier, Pernille R. Jensen, Svein-Olaf Hustvedt, Magnus Karlsson, Jens O. Duus, and Jan H. Ardenkjaer-Larsen. Quantitative dynamic nuclear polarization-NMR on blood plasma for assays of drug metabolism. *NMR in Biomedicine*, 24:96103, 2011.

[98] T. Harris, H. Degani, and L. Frydman. Hyperpolarized ^{13}C NMR studies of glucose metabolism in living breast cancer cell cultures. *NMR in Biomedicine*, 26:1831–1843, 2013.

[99] Matthew E Merritt, Crystal Harrison, A Dean Sherry, Craig R Malloy, and Shawn C Burgess. Flux through hepatic pyruvate carboxylase and phosphoenolpyruvate carboxykinase detected by hyperpolarized ^{13}C magnetic resonance. *Proceedings of the National Academy of Sciences of the United States of America*, 108:19084–19089, 2011.

[100] Pernille R. Jensen, Magnus Karlsson, Sebastian Meier, Jens Ø. Duus, and Mathilde H. Lerche. Hyperpolarized amino acids for in vivo assays of transaminase activity. *Chemistry - A European Journal*, 15:10010–10012, 2009.

[101] Haifeng Zeng, Youngbok Lee, and Christian Hilty. Quantitative rate determination by Dynamic Nuclear Polarization enhanced NMR of a Diels-Alder reaction. *Analytical Chemistry*, 82:8897–8902, 2010.

[102] Sean Bowen, Giridhar Sekar, and Christian Hilty. Rapid determination of biosynthetic pathways using fractional isotope enrichment and high-resolution dynamic nuclear polarization enhanced NMR. *NMR in Biomedicine*, 24:1016–1022, 2011.

[103] Youngbok Lee, Haifeng Zeng, Adam Mazur, Melanie Wegstroth, Teresa Carlomagno, Marcel Reese, Donghan Lee, Stefan Becker, Christian Griesinger, and Christian Hilty. Hyperpolarized binding pocket nuclear Overhauser effect for determination of competitive ligand binding. *Angewandte Chemie International Edition*, 51:5179–5182, 2012.

[104] Youngbok Lee, Haifeng Zeng, Simon Ruedisser, Alvar D. Gossert, and Christian Hilty. Nuclear magnetic resonance of hyperpolarized fluorine for characterization of protein–ligand interactions. *Journal of the American Chemical Society*, 134:17448–17451, 2012.

[105] Youngbok Lee, Gyu Seong Heo, Haifeng Zeng, Karen L. Wooley, and Christian Hilty. Detection of living anionic species in polymerization reactions using hyperpolarized NMR. *Journal of the American Chemical Society*, 135:4636–4639, 2013.

[106] Melanie Rosay, Jonathan C. Lansing, Kristin C. Haddad, William W. Bachovchin, Judith Herzfeld, Richard J. Temkin, and Robert G. Griffin. High-frequency dynamic

nuclear polarization in MAS spectra of membrane and soluble proteins. *Journal of the American Chemical Society*, 125:13626–13627, 2003.

[107] Sean Bowen and Christian Hilty. Time–resolved dynamic nuclear polarization enhanced NMR spectroscopy. *Angewandte Chemie International Edition*, 120:5313–5315, 2008.

[108] Michel Hochuli, Thomas Szyperski, and Kurt Wüthrich. Deuterium isotope effects on the central carbon metabolism of escherichia coli cells grown on a D_2O-containing minimal medium. *Journal of Biomolecular NMR*, 17:33–42, 2000.

[109] César Fernandez, Koba Adeishvili, and Kurt Wüthrich. Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. *Proceedings of the National Academy of Sciences of the United States of America*, 98:2358–2363, 2001.

[110] Anders Öhman, Alexey Rak, Maria Dontsova, Maria B. Garber, and Torleif Härd. NMR structure of the ribosomal protein L23 from *Thermus thermophilus*. *Journal of Biomolecular NMR*, 26:131–137, 2003.

[111] Haifeng Zeng, Sean Bowen, and Christian Hilty. Sequentially acquired two-dimensional NMR spectra from hyperpolarized sample. *Journal of Magnetic Resonance*, 199:159–165, 2009.

[112] David Wishart, Colin Bigam, Arne Holm, Robert Hodges, and Brian Sykes. ^{1}H, ^{13}C and ^{15}N random coil nmr chemical shifts of the common amino acids. I. investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, 5:67–81, 1995.

[113] Yoh Matsuki, Thorsten Maly, Olivier Ouari, Hakim Karoui, Franois LeMoigne,

Egon Rizzato, Sevdalina Lyubenova, Judith Herzfeld, Thomas Prisner, Paul Tordo, and Robert G Griffin. Dynamic nuclear polarization with a rigid biradical. *Angewandte Chemie International Edition*, 48:4996–5000, 2009.

[114] Gonzalo Vazquez, Estrella Alvarez, and Jose M. Navaza. Surface tension of alcohol water + water from 20 to 50°C. *Journal of Chemical and Engineering Data*, 40:611–614, 1995.

[115] Cezary M. Kinart, Wojciech J. Kinart, and Andrzej Kolasinski. Acetonitryle - water binary mixtures and their assumed internal structures. *Physics and Chemistry of Liquids*, 35:201–208, 1998.

[116] Kevin H. Gardner and Lewis E. Kay. The use of ^{1}H, ^{13}C, ^{15}N multidimensional nmr to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure*, 27:357–406, 1998.

[117] Kevin H. Gardner and Lewis E. Kay. Multidimensional ^{2}h-based NMR methods for resonance assignment, structure determination, and the study of protein dynamics. In N. Rama Krishna and Lawrence J. Berliner, editors, *Modern Techniques in Protein NMR*. Springer US, 2002.

[118] Alexander Krahn, Philip Lottmann, Thorsten Marquardsen, Andreas Tavernier, Maria-Teresa Trke, Marcel Reese, Andrei Leonov, Marina Bennati, Peter Hoefer, Frank Engelke, and Christian Griesinger. Shuttle DNP spectrometer with a two-center magnet. *Physical Chemistry Chemical Physics*, 12:5830, 2010.

[119] Renee Otten, Byron Chu, Karla D. Krewulak, Hans J. Vogel, and Frans A. A. Mulder. Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. *Journal of the American Chemical Society*, 132:2952–2960, 2010.

[120] Robert Zwanzig, Attila Szabo, and Biman Bagchi. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*, 89:20–22, 1992.

[121] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[122] Boyd Hardesty and Gisela Kramer. Folding of a nascent peptide on the ribosome. *Progress in nucleic acid research and molecular biology*, 66:41–66, 2001.

[123] Bernd Bukau and Arthur L Horwich. The hsp70 and hsp60 chaperone machines. *Cell*, 92:351–366, 1998.

[124] F. Ulrich Hartl and Manajit Hayer-Hartl. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, 295:1852–1858, 2002.

[125] Martin Karplus and David L. Weaver. Protein-folding dynamics. *Nature*, 260:404–406, 1976.

[126] Martin Karplus and David L. Weaver. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Science*, 3:650–668, 1994.

[127] Chris Beck, Xavier Siemens, and David L. Weaver. Diffusion-collision model study of misfolding in a Four-Helix bundle protein. *Biophysical Journal*, 81:3105–3115, 2001.

[128] Donald Bashford, David L. Weaver, and Martin Karplus. Diffusion-collision model for the folding kinetics of the γ-repressor operator-binding domain. *Journal of Biomolecular Structure and Dynamics*, 1:1243–1255, 1984.

[129] Gerhard Schwarz. On the kinetics of the helix-coil transition of polypeptides in solution. *Journal of Molecular Biology*, 11:64–77, 1965.

[130] E. M. Bradbury, B. G. Carpenter, and H. Goldman. Conformational studies of polymers and copolymers of L-aspartate esters. I. preparation and solution studies. *Biopolymers*, 6:837–850, 1968.

[131] Kevin W Plaxco and Christopher M Dobson. Time-resolved biophysical methods in the study of protein folding. *Current Opinion in Structural Biology*, 6:630–636, 1996.

[132] Martin Engelhard and Philip A. Evans. Experimental investigation of sidechain interactions in early folding intermediates. *Folding and Design*, 1:R31–37, 1996.

[133] James T. Vivian and Patrik R. Callis. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal*, 80:2093–2109, 2001.

[134] Bryan E Jones, Joseph M Beechem, and C Robert Matthews. Local and global dynamics during the folding of *Escherichia coli* dihydrofolate reductase by time-resolved fluorescence spectroscopy. *Biochemistry*, 34:1867–1877, 1995.

[135] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas, and R. I. Gilmanshin. Study of the molten globule intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, 31:119–128, 1991.

[136] Philip A. Evans and Sheena E. Radford. Probing the structure of folding intermediates. *Current Opinion in Structural Biology*, 4:100–106, 1994.

[137] David T Clarke. Circular dichroism and its use in protein-folding studies. *Methods in molecular biology*, 752:59–72, 2011.

[138] R A Goldbeck, T D Dawes, O Einarsdottir, W H Woodruff, and D S Kliger. Time-resolved magnetic circular dichroism spectroscopy of photolyzed carbonmonoxy cytochrome c oxidase (cytochrome aa3). *Biophysical Journal*, 60:125–134, 1991.

[139] Nikolay N. Kalnin and Kunihiro Kuwajima. Kinetic folding and unfolding of staphylococcal nuclease and its six mutants studied by stopped-flow circular dichroism. *Proteins*, 23:163–176, 1995.

[140] Kunihiro Kuwajima, Hidetoshi Yamaya, and Shintaro Sugai. The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. *Journal of Molecular Biology*, 264:806–822, 1996.

[141] Gulnur A. Elove, Alain F. Chaffotte, Heinrich Roder, and Michel E. Goldberg. Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry*, 31:6876–6883, 1992.

[142] Jos Nelson Onuchic and Peter G Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14:70–75, 2004.

[143] Wenfei Li, Tsuyoshi Terakawa, Wei Wang, and Shoji Takada. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proceedings of the National Academy of Sciences of the United States of America*, 109:17789–17794, 2012.

[144] Andreas G. Ladurner, Laura S. Itzhaki, Valerie Daggett, and Alan R. Fersht. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 95:8473–8478, 1998.

[145] Vijay S. Pande, Ian Baker, Jarrod Chapman, Sidney P. Elmer, Siraj Khaliq, Stefan M. Larson, Young Min Rhee, Michael R. Shirts, Christopher D. Snow, Eric J. Sorin, and Bojan Zagrovic. Atomistic protein folding simulations on the submil-

lisecond time scale using worldwide distributed computing. *Biopolymers*, 68:91–109, 2003.

[146] Bojan Zagrovic, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande. Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing. *Journal of Molecular Biology*, 323:927–937, 2002.

[147] Karen L. Maxwell, David Wildes, Arash Zarrine-Afsar, Miguel A. De Los Rios, Andrew G. Brown, Claire T. Friel, Linda Hedberg, Jia-Cherng Horng, Diane Bona, Erik J. Miller, Alexis Vallée-Blisle, Ewan R.G. Main, Francesco Bemporad, Linlin Qiu, Kaare Teilum, Ngoc-Diep Vu, Aled M. Edwards, Ingo Ruczinski, Flemming M. Poulsen, Birthe B. Kragelund, Stephen W. Michnick, Fabrizio Chiti, Yawen Bai, Stephen J. Hagen, Luis Serrano, Mikael Oliveberg, Daniel P. Raleigh, Pernilla Wittung-Sta fshede, Sheena E. Radford, Sophie E. Jackson, Tobin R. Sosnick, Susan Marqusee, Alan R. Davidson, and Kevin W. Plaxco. Protein folding: Defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Science*, 14:602–616, 2005.

[148] David S. Wishart, Brian D. Sykes, and Fredric M. Richards. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31:1647–1651, 1992.

[149] Haiyan Zhang, Stephen Neal, and David S Wishart. RefDB: a database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR*, 25:173–195, 2003.

[150] Mario Schubert, Maika Smalla, Peter Schmieder, and Hartmut Oschkinat. MUSIC in triple-resonance experiments: Amino acid type-selective ^{1}H–^{15}N correlations. *Journal of Magnetic Resonance*, 141:34–43, 1999.

[151] Mario Schubert, Hartmut Oschkinat, and Peter Schmieder. MUSIC, selective pulses, and tuned delays: Amino acid Type-Selective ^{1}H–^{15}N correlations, II. *Journal of Magnetic Resonance*, 148:61–72, 2001.

[152] Mario Schubert, Hartmut Oschkinat, and Peter Schmieder. MUSIC and aromatic residues: Amino acid type-selective ^{1}H–^{15}N correlations, III. *Journal of Magnetic Resonance*, 153:186–192, 2001.

[153] Pravat K. Mandal and Ananya Majumdar. A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts in Magnetic Resonance Part A*, 20A:1–23, 2004.

[154] Ariel Fernandez and Oktay Sinanoglu. Denaturation of proteins in methanol/water mixtures. *Biophysical Chemistry*, 21:163–166, 1985.

[155] Kai Griebenow and Alexander M. Klibanov. On protein denaturation in Aqueous–Organic mixtures but not in pure organic solvents. *Journal of the American Chemical Society*, 118:11695–11700, 1996.

[156] A. Zaks and A. M. Klibanov. Enzyme-catalyzed processes in organic solvents. *Proceedings of the National Academy of Sciences of the United States of America*, 82:3192–3196, 1985.

[157] Gerald Kirchner, Mark P. Scollar, and Alexander M. Klibanov. Resolution of racemic mixtures via lipase catalysis in organic solvents. *Journal of the American Chemical Society*, 107:7072–7076, 1985.

[158] Alexander M. Klibanov. Improving enzymes by using them in organic solvents. *Nature*, 409:241–246, 2001.

[159] H Jane Dyson and Peter E Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6:197–208, 2005.

[160] James Hallewell Johnston. The surface tension of protein solutions. part III. *Biochemical Journal*, 21:1314–1328, 1927.

[161] Naofumi Kitabatake and Etsushiro Doi. Surface tension and foaming of protein solutions. *Journal of Food Science*, 47:1218–1221, 1982.

[162] Hsueh-Ying Chen and Christian Hilty. Hyperpolarized hadamard spectroscopy using flow NMR. *Analytical Chemistry*, 85:7385–7390, 2013.

[163] Te-Won Lee, A. Ziehe, R. Orglmeister, and T. Sejnowski. Combining time-delayed decorrelation and ICA: towards solving the cocktail party problem. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing,*, pages 1249–1252, 1998.

[164] Simon Haykin and Zhe Chen. The cocktail party problem. *Neural Computation*, 17:1875–1902, 2005.

[165] Stotiris B. Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268, 2007.

[166] Enrico Blanzieri and Anton Bryl. A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, 29:63–92, 2008.

[167] Gordon V. Cormack. Email spam filtering: A systematic review. *Foundations and Trends in Information Retrieval*, 1:335–455, 2008.

[168] Fabian Sievers, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes

Sading, Julie D. Thompson, and Desmond G. Higgins. Fast, scalable generation of high–quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7, 2011.

[169] S. Horvath, B. Zhang, M. Carlson, K. V. Lu, S. Zhu, R. M. Felciano, M. F. Laurance, W. Zhao, S. Qi, Z. Chen, Y. Lee, A. C. Scheck, L. M. Liau, H. Wu, D. H. Geschwind, P. G. Febbo, H. I. Kornblum, T. F. Cloughesy, S. F. Nelson, and P. S. Mischel. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*, 103:17402–17407, 2006.

[170] Steve Horvath, Yafeng Zhang, Peter Langfelder, Ren S. Kahn, Marco PM Boks, Kristel van Eijk, Leonard H. van den Berg, and Roel A. Ophoff. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biology*, 13:R97, 2012.

[171] Markus Ringnar. What is principal component analysis? *Nature Biotechnology*, 26:303–304, 2008.

[172] Wei Kong, Charles R Vanderburg, Hiromi Gunshin, Jack T Rogers, and Xudong Huang. A review of independent component analysis application to microarray gene expression data. *BioTechniques*, 45:501–520, 2008.

[173] Gilbert W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35:551–566, 1993.

[174] Yu-Xiong Wang and Yu-Jin Zhang. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25:1336–1353, 2013.

[175] Gift Nyamundanda, Lorraine Brennan, and Isobel C. Gormley. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics*, 11:571, 2010.

[176] Danielle Nuzillard, Simon Bourg, and James M. Nuzillard. Model-free analysis of mixtures by NMR using blind source separation. *Journal of Magnetic Resonance*, 133:358–363, 1998.

[177] David A. Snyder, Fengli Zhang, Steven L. Robinette, Lei Bruschweiler-Li, and Rafael Brüschweiler. Non-negative matrix factorization of two-dimensional NMR spectra: Application to complex mixture analysis. *The Journal of Chemical Physics*, 128:052313, 2008.

[178] Dinesh Patil, Niva Das, and Aurobinda Routray. Implementation of fast-ICA: a performance based comparison between floating point and fixed point DSP platform. *Measurement Science Review*, 11:118–124, 2011.

[179] Li Shang, De-Shuang Huang, Ji-Xiang Du, and Chun-Hou Zheng. Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network. *Neurocomputing*, 69:1782–1786, 2006.

[180] A. Rashid, Zahooruddin, I.M. Qureshi, and A. Saleem. Electrocardiogram signal processing for baseline noise removal using blind source separation techniques: A comparative analysis. In *2011 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 4, pages 1756–1761, 2011.

[181] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[182] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

[183] Michael W. Berry, Murray Browne, Amy N. Langville, V. Paul Pauca, and Robert J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. In *Computational Statistics and Data Analysis*, pages 155–173, 2006.

[184] Chris D. Constantinides, Ergin Atalar, and Elliot R. McVeigh. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magnetic resonance in medicine*, 38:852–857, 1997.

[185] J Sijbers and A J den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic resonance in medicine*, 51:586–594, 2004.

[186] Lang Tong, R. Liu, V.C. Soon, and Yih-Fang Huang. Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 38:499–509, 1991.

[187] V. Paul Pauca, Jon Piper, and Robert J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, 416:29–47, 2006.

[188] M. Souden, S. Araki, K. Kinoshita, T. Nakatani, and H. Sawada. Simultaneous speech source separation and noise reduction via clustering and MMSE-based filtering. In *2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–6, 2011.