Solid-State NMR Studies of Solvent-Accessible Fragments of a Seven-Helical Transmembrane Protein Proteorhodopsin

by

Meaghan Ward

A Thesis
presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Master of Science
in
Biophysics

Guelph, Ontario, Canada

© Meaghan Ward, August, 2011
ABSTRACT

SOLID-STATE NMR STUDIES OF SOLVENT-ACCESSIBLE FRAGMENTS OF A SEVEN-HELICAL TRANSMEMBRANE PROTEIN PROTEORHODOPSIN

Meaghan Ward
University of Guelph, 2011

Advisors: Dr. V. Ladizhansky Dr. L. Brown

High–resolution multidimensional proton-detected NMR was used to study the solvent-exposed regions of a seven-helical integral membrane proton pump proteorhodopsin (PR). Fully deuterated PR samples with protons reintroduced to solvent-accessible sites through back exchange were prepared and found to produce NMR spectra with acceptable proton resolution (~0.2 ppm). Novel three-dimensional proton-detected chemical shift correlation spectroscopy was used for the identification and resonance assignment of the solvent–exposed regions of PR. Though most of the observed residues were located at the membrane interface there were notable exceptions, particularly in helix G. This helix contains the Schiff base-forming Lys231 and many conserved polar residues in the extracellular half. Solvent accessibility of helix G supports the hypothesis that high mobility of the F-G loop could transiently expose a hydrophilic cavity in the extracellular half of PR, and implies that such a cavity may be part of the protein’s proton-conduction pathway.
Acknowledgements

I owe many thanks to the people who have helped me in various ways throughout my degree. To begin, I would like to thank the boss-men, my advisors Dr. Vladimir Ladizhanksy and Dr. Leonid Brown, for all their eternal patience, help, support, and infallible good attitudes over the last few years. I would also like to thank Dr. Steffan Graether, for being on my advisory committee, Dr. Jim Davis for being on my advisory and examining committees, and Dr. John Dutcher for being on my examining committee. In addition, I would also like to thank Dr. George Harauz for continuing to take an interest in my research, even after I had left his group, and Dr. Michelle Oliver and Dr. Hermann Eberl for always being available to answer the BIG questions.

Many thanks are owed to all those who assisted me in this project. Specifically, I would like to thank the soon to be Dr. Ying Fan, Dr. Lichi Shi and Evelyn Lake for their help with protein sample preparation and Dr. Mumdooh Ahmed, Dr. Lichi Shi, and Dr. Shenlin Wang for their extensive help with data processing and analysis. I sincerely appreciate all of your time and patience.

I would also like to thank Valerie Robertson, Joe Meissner, and Peter Scheffer of the NMR centre for their technical assistance and willingness to help find a solution to whatever problem was tackling. Many thanks also go to Bill Teesdale, for his constant technical support, and to the lovely ladies in the Physics office, Reggie Vaillilee, Janice Hall, and Paige Clarke, and Tanya Qureshi for all of their administrative support. I would also like to thank all of the members of Bruker Canada and USA, specifically Sridevi Krishnamurthy, Howard Hutchins, Michael Fey, Werner Maas, Martine Monette, Henry Stronks, for their amazing technical support and for providing us with such high-quality equipment.
Finally, I would also like to thank my office mates, Iain Braithwaite, Miranda Schmidt, Emily Ritz, Andrew Gravelle, Mumdooh Ahmed, Lichi Shi, and Ivana Komljenovic for being a constant source of help, support, entertainment, and food.

This research was supported by funding in the form of a Canadian Graduate Scholarship from the Natural Sciences and Engineering Council of Canada from 2009-2010, and an Ontario Graduate Scholarship from the Ontario Ministry of Training, Colleges, and Universities from 2010-2011.
# Table of Contents

1.0 Introduction .................................................................................................................. 1
1.1 Solid-State NMR of Membrane Proteins ..................................................................... 1
1.2 Proteorhodopsin .......................................................................................................... 3
1.3 Proton Detection in SSNMR ....................................................................................... 8
2.0 Theory ............................................................................................................................ 12
  2.1 NMR Background ...................................................................................................... 12
    Zeeman Interaction ........................................................................................................ 12
    Chemical Shift Interaction ............................................................................................ 14
    Dipolar Interaction ....................................................................................................... 14
    J-Interactions ................................................................................................................ 15
    RF Field ......................................................................................................................... 16
  2.2 Magic Angle Spinning NMR .................................................................................... 16
  2.3 Proton Linewidth ...................................................................................................... 20
  2.4 Deuteration ............................................................................................................... 22
  2.5 Multidimensional NMR .......................................................................................... 23
    Cross Polarization ....................................................................................................... 23
    Double Quantum Excitation ....................................................................................... 25
3.0 Materials and Methods ............................................................................................... 29
  3.1 Materials .................................................................................................................... 29
  3.2 Expression, purification, and reconstitution of PR ..................................................... 29
  3.3 Assignment Strategy ................................................................................................. 31
  3.4 MAS SSNMR Spectroscopy ..................................................................................... 32
  3.5 Data Processing and Analysis .................................................................................. 36
4.0 Results and Discussion ............................................................................................... 37
  4.1 Spectral resolution in deuterated samples .................................................................. 37
  4.2 Peak picking and assignment .................................................................................... 43
  4.3 Amino Acid Assignments and Structural Implications ............................................. 48
5.0 Conclusions .................................................................................................................... 51
6.0 Future Directions .......................................................................................................... 52
7.0 References ...................................................................................................................... 55
List of Tables and Figures

**Figure 1-1:** a) 1D b) 2D and c) 3D SSNMR spectra .................................................. 2

**Figure 1-2:** Structural model of (A) bacteriorhodopsin and (B) proteorhodopsin .................. 5

**Figure 1-3:** Topology model of PR .................................................................................. 7

**Figure 1-4:** Amino acid systems .................................................................................... 10

**Figure 2-1:** Magic angle spinning .................................................................................. 17

**Figure 2-2:** Spin echo experiment .................................................................................. 21

**Figure 2-3:** Representation of the tilted frame for SPECIFIC CP .................................... 24

**Figure 2-4:** 2D INADEQUATE-type experiment using SPC-5$_3$ for DQ excitation and reconversion ................................................................. 26

**Figure 2-5:** Experimental $^{13}$C-$^{13}$C spectrum of uniformly $^{13}$C-labeled L-threonine ...................... 27

**Figure 3-1:** Assignment strategy ................................................................................... 32

**Figure 3-2:** Pulse sequences for a) 2D $^{13}$C-$^{13}$C and b) 2D $^1$H-$^{15}$N experiments ............... 34

**Figure 3-3:** Pulse sequences for a) CANH & CONH and b) DQCANH & DQCONH ............. 35

**Table 1:** Summary of chemical shift differences between UCN PR and UCND PR .................. 38

**Figure 4-1:** 2D $^{13}$C-$^{13}$C spectra of UCND labeled PR at 800 MHz frequency obtained with SPC-5$_3$ mixing ................................................................. 39

**Figure 4-2:** $^1$H-$^{15}$N heteronuclear correlation spectra collected from a U-$[^2]$H,$^{15}$N labeled sample of PR .................................................................................. 41

**Figure 4-3:** Comparison of predicted linewidths calculated from $T_2$ measurements to experimentally measured linewidths .................................................. 42

**Table 2:** Comparison of experimental linewidths and theoretical linewidths calculated from $T_2$ measurements on individual amino acids ........................................... 43

**Figure 4-4:** 2D planes of 3D CONH and DQCONH experiments performed on UCND PR .......... 44

**Figure 4-5:** Sequential walk from G171 to T177 through the CANH, CONH, DQCANH, and DQCONH spectra ........................................................................... 46

**Figure 4-6:** a) Topological representation and b) 3D model of PR ........................................ 49
Abbreviations

2/3D – two/three dimensional
ASR – Anabaena sensory rhodopsin
ATP – adenosine triphosphate
BR – bacteriorhodopsin
CARA – computer aided resonance assignment
CHES – N-Cyclohexyl-2-aminoethanesulfonic acid
CP – cross polarization
CSA – chemical shift anisotropy
CSI – chemical shift index
DMPA – 1,2-dimyristoyl-sn-glycero-3-phosphate
DMPC – 1,2-dimyristoyl-sn-glycero-3-phosphocholine
DSS – 2,2-dimethyl-2-silapentane-5-sulfonic acid
DQ – double quantum
E. coli – Escherichia coli
FID – free induction decay
FTIR – Fourier-transform infrared spectroscopy
GB1 – β1 immunoglobulin binding domain of protein G
INADEQUATE – incredible natural abundance double quantum transfer experiment
IPTG – Isopropyl β-D-thiogalactopyranoside
MAS – magic angle spinning
MS – mass spectrometry
NMR – nuclear magnetic resonance
PR – proteorhodopsin
ppm – parts per million
RF – radio frequency
S/N – signal to noise
SH3 – SRC homology 3 domain
SPECIFIC CP – spectrally induced filtering in combination with cross polarization
SPC-53 – Supercycled POST-C5
SQ – single quantum
SR-II – sensory rhodopsin II
SSNMR – solid state NMR
TM – transmembrane
TPPI – time proportional phase incrementation
UCN – uniformly-[13C,15N]-labeled
UCND – uniformly-[13C,15N,2H]-labeled
UND – uniformly-[15N,2H]-labeled
WS – water suppression
XR – xanthorhodopsin of Salinibacter ruber
1.0 Introduction

1.1 Solid-State NMR of Membrane Proteins

Membrane proteins account for ~30% of all proteins, and represent almost half of known drug targets. This class of proteins includes enzymes, ion channels, transporters, and receptors. The knowledge of their structures is key to the understanding of their functions. Despite this, few structures are known, largely because membrane proteins are not readily amenable to two of the major techniques currently used for protein structure determination; x-ray crystallography and solution NMR. Though x-ray crystallography has been successful in determining the structure of several membrane proteins, the high quality, well-diffracting protein crystals necessary for these experiments are often difficult to obtain. Solution NMR has also increasingly been successfully applied to membrane proteins [3-16], however the large protein-detergent micelle complexes, which are necessary for solubility purposes, generally do not tumble rapidly enough to average anisotropic interactions.

Magic angle spinning [17] (MAS) solid state nuclear magnetic resonance (SSNMR) is quickly becoming a powerful tool for investigating the structure and function of membrane proteins in a native-like lipid environment. Unlike the two previously mentioned methods, SSNMR does not require protein crystals and is not limited by size. Using SSNMR, amino acid assignments for several membrane proteins have been determined and partial assignments have been obtained for many more [1, 2, 18-29]. SSNMR has furthermore been used to provide accurate structural information in the form of distance and torsional constraints [30-34], solvent
Considerable progress has also been made into the investigation of ligand binding to membrane proteins [30, 39-44].

Before structural studies can be performed, it is necessary to identify individual amino acids in the spectra through the sequential assignment of spectral resonances. This is often done by first establishing intra-residue connectivity between nuclei to determine the amino acid type and then determining inter-residue connections to link systems and identify the sequence-specific amino acid assignments. Although it is possible to assign small proteins with just three 2D experiments [45], spectral overlap in more complicated systems often precludes the application of this method and a more rigorous approach is necessary. Using 3D experiments (Figure 1-1) both the spin system identification and spin system linking become less ambiguous due to the addition of an extra dimension, and this method has been applied to both soluble and membrane proteins [2, 21, 28, 46].
As the limits of MAS SSNMR are pushed and larger membrane proteins are studied, complications arise due to the length of the amino acid sequence, the high repetitiveness of hydrophobic residues, and the limit in spectral resolution arising from the dominance of one type of secondary structure. A variety of methods have been developed to resolve these issues including advanced labeling approaches [47-50], novel polarization transfer experiments [51], and 4D experiments [1, 52]. Investigations into the use of proton detection, a method previously not amenable to SSNMR due to the broad lines created by the large anisotropic interactions between protons, have also been initiated on microcrystalline proteins for signal enhancement, and these methods have the potential to be applied to membrane proteins as well, as will be discussed further in Section 1.3. In this work, SSNMR studies on the light-driven 7TM helical retinal-binding proton pump green proteorhodopsin are extended using proton detection experiments.

1.2 Proteorhodopsin

Rhodopsins are a class of integral membrane proteins which bind a retinal chromophore to form light-absorbing pigments [53]. The general structure of these proteins is a seven transmembrane \( \alpha \)-helical bundle that spans the cell membrane and encompasses the retinal molecule, which is bound covalently to the protein through a Schiff base formed with lysine. Rhodopsins have been found to perform a wide range of vital functions in many different organisms including bacteria, archaea, and eukaryotes [53]. For example, one of the most studied rhodopsins, bacteriorhodopsin (BR), acts as a proton pump which creates an electrochemical transmembrane ion gradient for the generation of ATP. Green proteorhodopsin, the protein investigated in these studies, belongs to a large family of eubacterial proton pumps [54].
Proteorhodopsins are found in the genomes of many species of marine bacteria in the photic zones of the oceans [55, 56]. Variants of this protein share up to 80% sequence identity, yet often exhibit significant differences in their absorption maxima [57], with variations of up to 35 nm in the blue-green region. Green proteorhodopsin (henceforth referred to simply as PR), which absorbs light at ~520 nm, has been the focus of many studies. PR acts as a light driven proton pump [55, 58, 59], which suggests that it could play an important energetic role in the oceans, even though there is controversy about the pH-dependent vectoriality of the proton transport [60, 61]. However, in the first studies on PR function the predicted differences in growth rates or cell yields between dark and light cultures were not observed [54], and the function of PR remained elusive.

Recently, PR has been shown to play an important role in a complex cellular response that maintains cell functions during periods of carbon starvation [62]. In carbon-starved cells exposure to light has been shown to cause changes in physiology and gene expression which ultimately allow cells to maintain their normal size and ATP content, while cells kept in the dark become spheres approaching the minimum size predicted for cells. The ability to produce ATP under carbon-starvation conditions allows cells to continue ATP hydrolysis and the import of substrates, which may enable cells to rapidly restart respiratory metabolism when oxidizable organic compounds are once again available. In this way PR provides cells with metabolic support during carbon starvation, a condition cells are likely to experience in the oceanic environment.

The photochemical cycle and proton-transporting mechanism of PR share many features with those of its archaeal homolog, BR, but also display some unique characteristics (Figure 1-2) [55, 58, 63, 64]. As in BR, photon absorption triggers retinal isomerization from all-trans to 13-cis,
rotating around the C13=C14 double bond. This facilitates a proton transfer from the Schiff base to D97 (primary proton acceptor), which is located on the extracellular side of PR. A second proton from E108 (primary proton donor) on the intracellular side replaces the first proton. In time, D97 will release its proton to the extracellular medium, E108 will take up another proton from the cytoplasm [63], and the retinal will reisomerize back to the all-trans configuration. However, while in BR the primary proton acceptor is known to pass its proton to a secondary proton acceptor pair (E194 and E204), homologous polar residues are not found in PR. Therefore, the mechanism of proton release in PR is not well understood.

Figure 1-2: Structural model of (A) bacteriorhodopsin and (B) proteorhodopsin. The PR structure is modeled with the 3D-JIGSAW program [65] using BR (PDB ID: 1C3W) as a template. Helices are lettered from A-G as labeled, and the proteins are oriented with the cytoplasmic side on top.
Another significant difference between PR and BR is the presence of a highly conserved histidine residue in PR (H75) which is not seen in BR, but is similar to H62 in xanthorhodopsin (XR) [66], a distant carotenoid–binding homolog. In both PR and XR this histidine has been found to form a hydrogen bond with the primary proton acceptor, and is responsible for the high pK_a of this residue [66, 67]. Like PR, XR lacks the final proton acceptor pair of BR. The recently solved structure of XR [66] has revealed displacements of the B–C and F–G interhelical segments (helices in XR are labeled as seen in Figure 1-2 in PR and BR) relative to their positions in BR, which expose a deep cavity extending from the extracellular side of the protein halfway towards the Schiff base. While PR may contain a similar cavity and possibly use it for proton release, considerable sequence differences separate XR and PR and it has been hypothesized that the primary structure of PR is intermediate between those of XR and BR [66].

Currently, there is little structural information available for PR. Topology analysis, homology modeling, and solid-state NMR studies have confirmed that PR follows the typical seven transmembrane helix topology [1, 2, 58, 64], and low-resolution atomic force and cryo-electron microscopy data show that PR can assemble into pentameric or hexameric complexes [68, 69]. Though PR produces 2D crystals, it does not form well-diffracting 3D crystals [70], which precludes x-ray crystallography studies. Solution NMR experiments have been attempted on detergent-solubilized PR, but HSQC spectra contained an insufficient number of resolvable peaks [70].
Figure 1-3: Topology model of PR with the cytoplasmic side on top. Filled circles represent residues for which resonance assignments have been obtained using solid state NMR [1, 2].

Solid-state NMR studies on PR reconstituted into lipid vesicles with the aim of solving a high-resolution structure for PR are currently underway [1, 2]. To date, 168 of 238 amino acids have been assigned (Figure 1-3), with a disproportionate number of the residues being located in the transmembrane regions. Through chemical shift index (CSI) analysis of the assigned regions and further analysis of the data, several structural details have been elucidated. To begin, helical boundaries could be established, and distortions to the helices, including proline and non-proline kinks, similar to those known for BR were observed. In addition, the presence of a short β-turn in the B-C loop, similar to that seen in sensory rhodopsin II (SR-II) [20] was discovered, and the cytoplasmic E-F loop was shown to form a short interfacial α-helix, similar to that suggested for BR by solution NMR [71]. Furthermore, it has been shown that many of the loop regions of the protein undergo increased motions, which could account for the high level of difficulty in obtaining assignments for these regions [1, 2, 72].
1.3 Proton Detection in SSNMR

Proton detection is an indispensable component in most biological and organic solution NMR pulse sequences. Compared to the direct detection of $^{13}\text{C}$ and $^{15}\text{N}$, proton detection experiments offer theoretical increases in sensitivity of up to 8 and 30 fold, respectively [73]. In addition, proton detection allows for another dimension in which to resolve systems, provides additional chemical shift data, and introduces the ability to measure proton-proton distances and the possibility to probe proton-proton interactions. In SSNMR the broad lines caused by strong proton-proton dipolar interactions preclude the easy application of proton detection, as at standard MAS conditions the proton dimension is completely unresolved. To improve the resolution of proton detected SSNMR experiments homonuclear decoupling sequences [74, 75], fast MAS [73, 76], and reduction of the proton homonuclear coupling network through extensive deuteration [77-82] have been utilized.

For the purpose of structural analysis, ultra-fast MAS experiments on fully protonated proteins are advantageous as complete chemical shift information is obtainable, the largest number of proton-proton distance constraints are available, sample preparation is simplified, and signal intensity per unit sample is maximized due to the large number of protons available for initial polarization. These experiments are preferred in cases where obtaining large amounts of sample is difficult, or isotopic labeling of the sample is not possible. Such experiments on microcrystalline GB1 spun at a MAS rate of 40 kHz have been shown to yield fully resolvable spectra, with linewidths of ~500 Hz, and full backbone assignments were possible [76]. Unfortunately, the resolution available from these experiments is often not sufficient to completely resolve the proton dimension in more complicated systems with congested spectra, such as multi-spanning $\alpha$-helical proteins.
Proton linewidths can be further decreased through deuteration of the sample and the selective reintroduction of protons. Though replacing large numbers of protons with deuterium leads to a reduction in the available chemical shift information, the number of observable proton-proton interactions, and signal intensity per unit sample, the increased resolution of deuterated samples is necessary to resolve the proton dimension in more complicated systems where spectral overlap is a problem. Furthermore, narrow linewidths can be obtained on deuterated samples spun at lower spinning frequencies, and thus larger diameter rotors, which can help compensate for the signal loss due to the reduction in proton density, can be utilized.

In perdeuterated proteins, protons can be easily reintroduced at exchangeable positions through back exchange with a protonated solvent. Though low levels of back exchange protonation have been shown to yield the best resolution [83-86], the linewidths in this regime are often dominated by inhomogeneous factors, and it is likely that higher proton concentrations can result in the optimal S/N ratio. For example, in recent studies on microcrystalline α-spectrin SH3 optimal S/N was observed at proton concentrations of 30-40% [87]. It is likely that this number is highly variable between different systems.

Proton detected experiments on deuterated microcrystalline proteins have been shown to yield linewidths as low as 20 Hz [80, 87, 88] and have furthermore been shown to yield acceptable resolution in membrane and fibrillar proteins [89]. Currently, no universal method for the assignment and linking of amino acid systems has yet been established for proton detected experiments on deuterated proteins in SSNMR, as polarization transfer pathways are limited in such samples. In order to be confidently linked, the spin systems created from NMR experiments must contain at least two inter-residue connections, as seen in Figure 1-4, as well as sufficient intra-residue information for amino acid type assignment. Creating spin systems with adequate
information for sequential assignments in deuterated proteins is complicated by the inefficient excitation of side chain carbons and the necessity to transfer polarization back to the exchangeable amide protons for detection. Thus, weak polarization and the many polarization transfer steps required make collecting side chain information difficult.

Figure 1-4: Amino acid systems created from a series of proton detected experiments, which are displayed in Figure 3-1. As indicated by the red ellipses, adjacent systems are linked through simultaneously linking of CA (CA=Cα) and CO resonances. Amino acid type is determined through the CB (CB=Cβ) resonance values.

One assignment strategy presented [88] uses scalar-based polarization transfer to create amino acid systems containing CA[i]/[i-1] and CB[i]/[i-1] shifts. These systems are linked through the matching of CA and CB resonances, and amino acid type can be determined for many systems using CB information. However, scalar-based polarization transfer is not always efficient in large proteins, and thus other assignment strategies are necessary.

This thesis presents a set of dipolar-based proton detection experiments, which will be explained in full in Section 3.3, for the sequential amino acid assignment of deuterated proteins with protons selectively reintroduced through back-exchange at the backbone amide position. As
seen in Figure 1-4, the four proton detected experiments are used to create amino acid systems which contain two overlapping residues for system linking, as well as CB information which can be used for amino acid type identification. These experiments are applied to the solvent exposed regions of PR.

As many studies have shown that the protein core is not susceptible to hydrogen-deuterium exchange [28, 90-96], without employing additional unfolding-refolding protocols [3], deuterated membrane proteins will have protons selectively reintroduced at exchangeable positions in solvent exposed regions of the protein only. Thus, proton detection experiments on these proteins benefit from not only an increase in signal intensity, but also from the reduction of overlap in the spectra due to the absence of solvent-shielded residues. The solvent exposed regions of PR are of particular interest, as they include the loop and termini regions, which are often hard to assign due to motions [1, 8], as well as residues located in the protein core which are likely involved in proton transport. Through these experiments it could therefore be possible to not only obtain resonance assignments not available though carbon-detected methods, but also to identify and study potentially important residues in PR’s proton transport pathway.
2.0 Theory

2.1 NMR Background

Nuclear Magnetic Resonance (NMR) uses the spin property of atoms to probe the atomic structure of molecules. In most cases, biological samples are isotopically labeled with $^{13}$C and $^{15}$N. These nuclei, as well as $^1$H, have a spin value of $\frac{1}{2}$, and thus the quadrupolar interaction, which is present in nuclei with spin values greater than $\frac{1}{2}$, need not be considered. In the theoretical description of an NMR experiment the Hamiltonian, $H(t)$, describes the interactions occurring within the system and can be written as:

$$H = H_Z + H_{CS} + H_D + H_J + H_{RF}$$

where

$$H_Z = \text{the Zeeman interaction}$$

$$H_{CS} = \text{chemical shift interaction}$$

$$H_D = \text{the dipolar interaction}$$

$$H_J = J – \text{interaction}$$

$$H_{RF} = \text{radio frequency interaction}.$$  

While the chemical shift, dipolar, and J-interactions are intrinsic to the system, the Zeeman and radio frequency interactions are the result of applied fields.

Zeeman Interaction

The Zeeman interaction is described by the Hamiltonian

$$H_Z = \omega_0 I_Z$$

$$\omega_0 = -\gamma \hbar B_0$$
where $\gamma$ is the gyromagnetic ratio of the nucleus, $B_0$ is the strength of the magnetic field which is chosen to be along the $z$-axis, $\vec{B} = (0,0,B_0)$, $I_Z$ is a spin $\frac{1}{2}$ operator, $\hbar$ is Planck’s constant, and $\omega_0$ is the Larmor frequency of the nucleus. By convention, energy and field strength are measured in frequency units and $\hbar$ is set to equal 1.

The Larmor frequency is equal to the energy difference between the two eigenstates of the Hamiltonian, $|\alpha\rangle, |\beta\rangle$, which correspond to the $\pm 1/2$ projections of the $z$-component of the spin. These states have energies, in frequency units, of

$$E_\alpha = -\frac{1}{2} \gamma B_0, \quad E_\beta = \frac{1}{2} \gamma B_0$$

with their respective populations $(n_\alpha, n_\beta)$ defined by the Boltzmann distribution,

$$n_\alpha = e^{-E_\beta/k_B T}, \quad n_\beta = e^{-E_\alpha/k_B T}.$$  \hspace{1cm} (2.4)

Inequalities in the above populations creates a net magnetization ($M_0$) along the $z$-direction:

$$M_0 = \frac{\gamma}{2} (n_\alpha - n_\beta)$$

which at high temperatures can be approximated as

$$M_0 \approx \frac{N \gamma^2 B_0}{4 k_B T}.$$ \hspace{1cm} (2.5)

This magnetization is dependent on the gyromagnetic ratio, the strength of the magnetic field, the number of atoms, $N$, and the temperature, $T$. As the population difference between $n_\alpha$ and $n_\beta$ is on the order of $1$ in $10^5$ for protons at high magnetic field and physiological temperature [97], the nuclear magnetization is generally small.

The Zeeman interaction is by far the strongest interaction occurring in the system, and all non-secular interactions which do not commute with this interaction can generally be neglected. Secular terms, which do commute will correct energies but do not perturb eigenstates.
**Chemical Shift Interaction**

When a population of spins is placed in an external magnetic field, the motions of electrons which are induced generate secondary magnetic fields. Because of this the local magnetic field at each spin will be slightly different and each spin will have a slightly different Larmor frequency. Because of these differences, individual nuclei, which without the chemical shift interaction would appear identical in the spectra, can be identified. The difference between the Larmor frequency of a particular spin and the expected Larmor frequency of the same spin without the presence of other spins is called the chemical shift. The chemical shift interaction can be represented as

\[ H_{CS} = (\omega_{ISO} + \omega_{CSA}) l_z \]  \hspace{1cm} (2.7)

where \( \omega_{ISO} \) is the isotropic component of the chemical shift, which does not depend on the orientation, and \( \omega_{CSA} \) is the anisotropic component of the chemical shift and is averaged by molecular tumbling in solution NMR, or by MAS in SSNMR. The first step in a NMR structural study is to assign chemical shift values to individual atoms in the protein. Chemical shifts can be related to structure and also allow for specific sites in molecules to be studied. The chemical shift is most often reported in units of parts per million (ppm), which can be calculated using

\[ \delta = \left( \frac{\omega_{ISO} - \omega_0}{\omega_0} \right) \times 10^6. \]  \hspace{1cm} (2.8)

**Dipolar Interaction**

The dipolar interaction is the direct through space coupling of two magnetic moments. The dipolar interaction can be represented as
where 

\[ d_{jk} = \frac{\mu_0 \gamma_j \gamma_k}{4\pi r_{jk}^3} \]  

and \( \gamma_{jk} \) is the gyromagnetic ratio of the \( k \) or \( j \) nucleus, and \( r_{jk} \) is the distance between the \( j \) and \( k \) nuclei. By neglecting non-secular terms, which do not commute with the Zeeman Hamiltonian, this expression can be simplified to

\[ H_D = \sum_{j<k} d_{jk} (\vec{I}_j \cdot \vec{I}_k - 3 \frac{\vec{I}_j \cdot \vec{r}_{jk} \vec{I}_k \cdot \vec{r}_{jk}}{r_{jk}^2}) \]

\[ \text{(2.9)} \]

in the homonuclear case and to

\[ H_D = \sum_{j<k} \omega_D^{jk} \left( 2I_{jz}I_{kz} \right) \]

\[ \text{(2.11)} \]

in the heteronuclear case, in which

\[ \omega_D^{jk} = d_{jk} (1 - 3\cos^2 \theta_{jk}) \]

\[ \text{(2.13)} \]

and \( \theta_{jk} \) is the angle between \( \vec{r}_{jk} \) and the static magnetic field \( B_0 \). Since the dipolar interaction is orientation dependent, it can lead to peak splitting and anisotropic broadening if this interaction is not eliminated by isotropic tumbling of the molecules or by MAS, as will be described further in Section 2.2.

**J-Interactions**

J-coupling characterizes indirect dipole-dipole interactions that occur through bonding electrons. The J-coupling Hamiltonian can be approximated as

\[ H_J = 2\pi J_{jk} (\vec{I}_j \cdot \vec{I}_k) \]

\[ \text{(2.14)} \]

in the homonuclear case and as

\[ H_J = 2\pi J_{jk} (I_{jz}I_{kz}) \]

\[ \text{(2.15)} \]
in the heteronuclear case, where $J_{jk}$ is the isotropic J-coupling. This is the weakest of the interactions and in the context of this work can be neglected.

**RF Field**

To create magnetization detectable using NMR the system must be perturbed from its equilibrium position. This can be accomplished by applying a radio-frequency (RF) pulse to the sample at or close to the Larmor frequency of the nuclear spins to be excited. The RF Hamiltonian can be represented as

$$H_{RF} = 2\omega_1 I_c \cos(\omega_{RF} t + \varphi)$$  \hspace{1cm} (2.16)

where $\omega_1$ is the strength of the field, $\omega_{RF}$ is the frequency at which the pulse is applied, and $\varphi$ is the phase of the pulse. When this pulse is removed the spins will relax back to equilibrium. During this process energy is lost in the form of radio waves, which can be detected. The obtained signal is called a Free Induction Decay (FID) and if a Fourier transformation of this data is taken, a frequency spectrum can be obtained.

**2.2 Magic Angle Spinning NMR**

In solid samples, unlike the liquid samples used in solution NMR, the anisotropic interactions experienced by nuclear spins are not averaged by molecular tumbling. The most common approach used to average these interactions is magic angle spinning (MAS) [17] (Figure 2-1). In a basic MAS experiment the sample is tilted at an angle of 54.7 ° with respect to the external magnetic field.
When a sample is spun around an axis tilted at an angle with respect to the external magnetic field the anisotropic interactions become periodically time dependent. In the case of dipolar interactions this can be represented as:

\[
\omega_{jk}^d(t) = d_{jk}(g_0 + g_1 \cos(\omega_r t + \varphi) + g_2 \cos(2\omega_r t + 2\varphi))
\]  
(2.17)

\[
g_0 = \frac{(3\cos^2\theta_{jk}-1)(3\cos^2\theta_m-1)}{2}
\]  
(2.18a)

\[
g_1 = \frac{3}{4}\sin 2\theta_m \sin 2\theta_{jk}
\]  
(2.18b)

\[
g_2 = -\frac{3}{4}\sin^2 \theta_m \sin^2 \theta_{jk}
\]  
(2.18c)

where \(\omega_r\) is the spinning frequency, \(\theta_m\) is the angle between the external magnetic field and the rotor axis, and \(\theta_{jk}\) is a polar angle of \(\vec{r}_{jk}\) in a rotating reference frame as defined by the sample rotor at a time \(t = 0\). By choosing to tilt the axis at an angle of \(\cos^{-1} (\sqrt{1/3}) \approx 54.7\) with respect to the magnetic field, the time-independent anisotropic contribution of \(g_0\) disappears. The remaining terms, \(g_1\) and \(g_2\), are oscillatory and are averaged to zero over a rotor period when MAS is applied, as can be shown by examining the time evolution of the system.

---

**Figure 2-1**: Magic angle spinning. Sample is packed in a rotor that spins about an axis pointing at the magic angle (54.7°) relative to the external magnetic field.
The time evolution of an ensemble of spins is determined by the Liouville-von Neumann equation:

$$\frac{d\rho(t)}{dt} = -i[H(t), \rho(t)].$$  \hspace{1cm} (2.19)

In the case of a time-independent Hamiltonian, the solution to this equation is simply

$$\rho(t) = e^{-iHt} \rho(0) e^{iHt}. \hspace{1cm} (2.20)$$

However as previously noted, MAS creates a time-dependent Hamiltonian. In this case, the density matrix must be propagated through a series of small time intervals, \((\Delta t = t/n, \text{where } n \text{ is the number of divisions})\), such that the Hamiltonian can be approximated to be time independent during each time interval:

$$\rho(t) = e^{-iH_n \Delta t} e^{-iH_{n-1} \Delta t} \ldots e^{-iH_1 \Delta t} e^{-iH_0 \Delta t} \rho(0) e^{iH_0 \Delta t} e^{iH_1 \Delta t} \ldots e^{iH_{n-1} \Delta t} e^{iH_n \Delta t}. \hspace{1cm} (2.21)$$

As \(n \to \infty\) this becomes:

$$\rho(t) = U(t) \rho(0) U^{-1}(t) \hspace{1cm} (2.22)$$

where

$$U(t) = T e^{-i \int_0^t H(t) dt} \hspace{1cm} (2.23)$$

and \(T\) is the Dyson time-ordering operator [98], which defines the order of the exponential functions for a Hamiltonian that is not self-commuting at different times. This expression can often be approximated as

$$U(t) = e^{-i\bar{H}t} \hspace{1cm} (2.24)$$

where \(\bar{H}\) is the time-independent average Hamiltonian and can be defined using the Magnus expansion [98]:

$$\bar{H} = \bar{H}^{(0)} + \bar{H}^{(1)} + \bar{H}^{(2)} + \ldots \hspace{1cm} (2.25)$$

in which
If this Hamiltonian is periodic in time, as is the case with systems undergoing MAS, it can be used to describe the time evolution over long periods:

$$U(nt_c) = U(t_c)^n = e^{-i\mathcal{H}nt_c}$$  \hspace{1cm} (2.27)

where $t_c$ is the length of the period.

When the Hamiltonian is self commuting at all times, such as with the CSA and the heteronuclear dipolar interactions, it is easy to see that all higher order terms will be zero, and thus the evolution will proceed as seen by solving for $\mathcal{H}^{(0)}$. For example, using the Hamiltonian for heteronuclear dipolar interactions:

$$U(t_1, t_2) = e^{-i\theta(t_1, t_2)J_{Jz}Kz}$$  \hspace{1cm} (2.28)

$$\theta(t_1, t_2) = \Omega_1\{\sin(\omega_{r1}t_2 + \varphi_1) - \sin(\omega_{r1}t_1 + \varphi_1)\} + \Omega_2\{2\sin(\omega_{r2}t_2 + \varphi_2) - \sin(2\omega_{r1}t_1 + \varphi_2)\}$$  \hspace{1cm} (2.29)

$$\Omega_1 = \frac{d_{jk}\alpha_1}{\omega_{r1}}, \Omega_2 = \frac{d_{jk}\alpha_2}{\omega_{r2}}$$  \hspace{1cm} (2.30)

As $\omega_r$ is increased, $\Omega_1$ and $\Omega_2$ will become very small, as does the time-dependent precession angle $\theta(t_1, t_2)$, thus the effect of these interactions is reduced to zero. Therefore, in the case of the CSA and heteronuclear dipolar interactions, MAS is sufficient to average the anisotropic interactions.

In some cases the Hamiltonian is not self-commuting. These interactions, which include multi-spin homonuclear dipolar interactions, are said to be homogeneous in the sense of Maricq...
& Waugh (and self-commuting Hamiltonians are called inhomogeneous) [99]. Since the Hamiltonian is not self-commuting, higher order terms in Eq. 2.27 will contribute to the average Hamiltonian. This effect is small and can generally be neglected for homonuclear carbon and nitrogen interactions, since the Larmor frequencies of neighboring carbon atoms are well dispersed, nitrogen has a low population, and therefore large distances between atoms, and both atoms have low gyromagnetic ratios. This is not true, however, for the homonuclear proton dipolar interactions which are strong due to the high gyromagnetic ratio and the number of different interactions involved. Thus, the contribution from the higher order terms cannot be ignored and MAS is generally not sufficient to average homonuclear $^1\text{H}-^1\text{H}$ interactions. In SSNMR further measures must often be taken in order to resolve proton resonances.

2.3 Proton Linewidth

In addition to the dipolar interactions, many other factors contribute to the proton linewidth, such as sample heterogeneity, $B_0$ field inhomogeneity, molecular dynamics, and temperature effects due to temperature gradients within the sample. The line broadening caused by these interactions can be classified as either homogenous or inhomogeneous. Homogeneous line broadening is caused by fluctuating microscopic magnetic field, includes molecular motions and homonuclear dipolar interactions, and is quantified by the transverse relaxation time, $T_2$. Inhomogeneous line broadening is caused by variation in the macroscopic magnetic field over the volume of the sample and includes $B_0$ field inhomogeneity and sample heterogeneity [97].

The effects of homogeneous versus inhomogeneous line broadening can be differentiated using a series of spin-echo experiments (Figure 2–2). During the echo-period, inhomogeneous interactions are refocused by the $180^\circ$ pulse, and thus do not contribute to the signal decay
during this time. By performing a series of spin-echo experiments with increasing mixing time, $\tau$ [97], the observed signal decay in the resulting signal can be modeled by

$$\text{Signal} \propto e^{-\frac{\tau}{T_2}}$$

where $T_2$ is the transverse relaxation time. According to theory, the linewidth (LW), or full width at half height, of a spectral peak due to homogeneous interactions can be defined as

$$LW_h = \frac{1}{\pi T_2}.$$  \hspace{1cm} (2.32)

By comparing this value to the experimentally determined linewidth ($LW_{exp}$), line-broadening due to inhomogeneous interactions ($LW_i$) can be determined using

$$LW_i = LW_{exp} - LW_h.$$ \hspace{1cm} (2.33)

---

**Figure 2-2**: Spin echo experiment.

---

In order for proton detection experiments on large proteins to be feasible, the homogeneous broadening caused by proton-proton dipolar interactions must be further reduced.
As seen previously, the strength of the dipolar interactions is proportional to $r^{3}_{ij}$ and is summed over all interacting pairs. Line broadening due to dipolar interactions can therefore be reduced by reducing the number of protons in the sample and by increasing the distance between the protons which are present. This is often done through deuteration of the sample. In general, a fully deuterated sample is created and protons are reintroduced at exchangeable positions through back-exchange with a protonated solvent.

### 2.4 Deuteration

In addition to narrowing proton linewidths, the use of deuterated samples with protons reintroduced at exchangeable positions has the additional benefits of eliminating the need for high-powered decoupling, which can lead to sample degradation, increasing the transverse relaxation time of carbon atoms, which allows for longer experiments to be performed, and reducing the water signal, which can often obscure peaks in the proton dimension, when low proton back-exchange concentrations are used. Though it could be predicted that initial $^{1}$H-$^{13}$C cross polarization would suffer due to the dilution of the proton bath, it has been shown that this transfer is still reasonably efficient [100, 101].

A possible disadvantage to using deuterated samples is the deuterium-induced changes in chemical shifts which have been reported for carbon and nitrogen resonances in studies of deuterated proteins [100, 102]. These shifts can complicate the transfer of resonance assignments from protonated to deuterated samples. Though it has been possible to transfer assignments between protonated and deuterated samples while taking into account the average changes in chemical shift for different atom types, in complicated spectra it is still preferable to confirm amino acid assignments through spectra obtained from the deuterated sample.
2.5 Multidimensional NMR

Though individual resonances are dispersed as a result of the chemical shift interaction, and MAS can be used to significantly narrow linewidths, the sheer number of resonances in large proteins most often leads to spectral overlap. This overlap can be resolved using multidimensional experiments in which the chemical shifts of different spins are correlated. These experiments are essential in the assignment of spectral resonances to individual nuclei. In order to correlate resonances, polarization must be transferred between nuclear species. In SSNMR this is generally done by recoupling the dipolar interaction. In this thesis the two main types of recoupling used are cross polarization, for heteronuclear dipolar recoupling, and double-quantum recoupling, for homonuclear $^{13}$C-$^{13}$C recoupling.

Cross Polarization

Cross polarization (CP) is a common technique in MAS used to transfer polarization between spins. Two of the most useful applications of CP are its use in enhancing the signals of low gamma nuclei ($^{13}$C or $^{15}$N) through polarization transfer with dipolar coupled proton baths, and its use in establishing frequency selective transfers using SPECIFIC-CP [103]. To transfer magnetization through cross polarization under MAS conditions the two nuclei involved must be irradiated simultaneously such that the Hartmann-Hahn condition [104] \((\omega_I \pm \omega_S = n\omega_r)\), is satisfied. This experiment can be described using the CP Hamiltonian

$$H(t) = \omega_{1I}I_x + \omega_{1S}S_x + \Delta^I I_z + \Delta^S S_z + 2\omega_d(t)(I_zS_z)$$

(2.34)

where \(\omega_{1I}\) and \(\omega_{1S}\) are the RF intensities for spin I and S, \(\Delta^I\) and \(\Delta^S\) are the resonance offsets. The Hartmann-Hahn condition holds only when the RF strengths used in these experiments
greatly exceed the offset of the isotropic chemical shifts \((\omega_{1S}^{iS} \gg \Delta^{iS})\). In this case the experiment becomes insensitive to the offset of the isotropic chemical shifts, and all resonances are excited. However, when the resonance offsets and the RF powers are of comparable intensities, the Hartmann-Hahn condition becomes

\[
(\omega_{l,\text{eff}} \pm \omega_{s,\text{eff}} = n\omega_r)
\]

where the effective field experienced by a spin is expressed as

\[
\omega_{l,s}^{\text{eff}} = \sqrt{(\Delta^{l,s})^2 + (\omega_{1S}^{iS})^2}
\]

Figure 2-3: Representation of the tilted frame for SPECIFIC CP. Off resonance RF fields are employed in two channels such that the sum or difference of the effective fields matches a multiple of the spinning frequency.

as represented in Figure 2-3. In this case, the dependence of the matching condition on the isotropic chemical shifts allows for frequency-selective transfers to be performed. This type of CP experiment is particularly useful for selectively cross polarizing either CO or CA from N. Backbone nitrogens are coupled to a CA \((\delta \approx 45-60 \text{ ppm})\) and a CO \((\delta \approx 170-180 \text{ ppm})\) with approximately equal strengths. By irradiating directly between the two frequency ranges at a high power, it is possible to simultaneously recouple both interactions. However, one of these
interactions can be eliminated by centering the irradiation at the frequency of one type of atom and using a lower RF power. For example, if the carbon carrier frequency is set at CA, CO will have an off-resonance component of ~120 ppm, or 24 kHz at 800 MHz proton frequency. Under 21 kHz MAS spinning, the Hartmann-Hahn condition can be matched for CA and N by choosing
\[ \frac{\omega_{1N}}{2\pi} = 31.5 \text{ kHz}, \quad \text{and} \quad \frac{\omega_{1C}}{2\pi} = 52.5 \text{ kHz}. \]
However, the Hartmann-Hahn condition for CO and N will be mismatched by ~5.2 kHz, which is much larger than the dipolar coupling strength between CO and N (~1 kHz). Thus the effective Hamiltonian is established for the N-CA pair only, while the N-CO interaction is decoupled by MAS.

**Double Quantum Excitation**

Double quantum (DQ) coherence is excited by establishing the DQ dipolar Hamiltonian,
\[ H = D(I_1^+I_2^+ + I_1^-I_2^-) \]
between spin pairs. In an INADEQUATE (incredible natural abundance double quantum transfer experiment) -type [105] experiment as seen in Figure 2-4, the DQ coherence is excited before the evolution period, and then reconverted to observable single quantum (SQ) coherence for detection or further polarization transfer. As seen in Figure 2-5, in a 2D INADEQUATE-type experiment the sum of the chemical shifts of coupled spins are collected in the indirect dimension, and are correlated to detectable SQ coherences in the direct dimension. These experiments have the benefits of suppressing signal from isolated spins and eliminating diagonal peaks which can lead to spectral overlap in SQ spectra.

A variety of methods for exciting DQ coherence have been developed, including HORROR [106], DREAM [107], and synchronous helical pulse sequences such as C7 [108], SC14 [109], SPC5 [110], and its band selective version SPC-53 [111]. The selection of the ideal recoupling sequence depends on a variety of factors, including required spinning frequency,
power requirements, and sensitivity to chemical shift dispersion. As many of these recoupling techniques require power levels which are integer multiples of the spinning frequency, they are non-ideal for use in proton-detection experiments, which are generally performed at higher MAS frequencies.

Figure 2-4: 2D INADEQUATE-type experiment using SPC-53 for DQ excitation and reconversion. The diagonal lines on the reconversion block indicate the use of a DQ phase cycle.

The SPC-53 pulse sequence has a relatively low RF power requirement of \( \frac{10}{3} \omega_r \), which makes it an ideal candidate for use at high spinning frequencies. Furthermore, it is sufficiently broadband such as to be able to recouple the CA-CO dipolar interactions, as is required for our experiments. The SPC-53 pulse sequence (Figure 2-4) is a rotor-synchronized spin-locking pulse sequence with fivefold symmetry. The basic unit of the SPC-53 sequence is timed to span three rotor cycles, and is broken into 5 C-elements, with \( t_c \) being the length of a single element, which are sequentially phase shifted by \( \frac{2\pi}{5} \). As seen in Figure 2-4, each C-element is further subcycled
in order to ensure stability towards isotropic offset and RF-field inhomogeneity. The phase inversion super-cycle, achieved through the inversion of the phases of the C-elements in every other basic unit, as shown in Figure 2-4, is used to further filter components with unwanted symmetry.

---

**Figure 2-5:** Experimental $^{13}$C-$^{13}$C spectrum of uniformly $^{13}$C-labeled L-threonine. The assignments of the DQ coherences to the molecular site pairs are indicated. CA+CB and CB+CG coherences are folded in the spectrum. The spectrum was recorded using TPPI[112] for quadrature detection.

---

Difficulties in the application of DQ spectroscopy to proteins are encountered due to the large spectral width of the double quantum dimension and the necessity for the reconversion block to be rotor-synchronous with the excitation block. Without applying further phase shifting to the reconversion block, the $t_1$ increment is limited to multiples of the rotor period ($t_R$), which simultaneously limits the spectra width of the DQ dimension. In these experiments States [113] or States-TPPI [114] is used for phase sensitive detection in order to maximize the spectral
width. However, at some spinning frequencies this can lead to spectral overlap between DQ coherences. It has been shown, that with the implementation of proper phase shifting of the reconversion block to maintain rotor synchrony, that spectral width in the DQ dimension can be increased by setting the $t_1$ increment equal to the length of once C-element ($t_C$) [115] and that the position of folded coherences can be altered through the use of time proportional phase incrementation (TPPI) [112] for phase sensitive detection [109]. Through the implementation of these techniques, spectral overlap in the DQ dimension can be avoided.
3.0 Materials and Methods

3.1 Materials

Common chemicals of reagent grade were purchased from either Fisher Scientific (Unionville, Ontario, Canada) or Sigma-Aldrich (Oakville, Ontario, Canada). Isotopically labeled compounds, such as $^{15}$NH$_4$Cl, $^2$H,$^{13}$C$_6$–glucose, $^2$H-glucose, and deuterium oxide, were obtained from Cambridge Isotope Laboratories (Andover, MA). The Ni$^{2+}$-NTA (nitrilotriacetic acid) agarose resin was purchased from Qiagen (Mississauga, Ontario, Canada). Lipids were purchased from Avanti Polar Lipids (Alabaster, AL).

3.2 Expression, purification, and reconstitution of PR

In the following, [U-$^{15}$N,$^2$H], [U-$^{13}$C,$^{15}$N,$^2$H], and [U-$^{13}$C,$^{15}$N]-labeled PR samples are referred to as UND, UCND, and UCN, respectively. UND and UCND samples were produced by optimizing a previously described protocol [1, 2] for expression in 100% D$_2$O. BL21-Codonplus-RIL E. coli cells were transformed with a plasmid encoding C-terminally 6×His-tagged PR, and cultured in 100% D$_2$O M9 minimal medium at 30°C, using 4 g of either uniformly $^2$H,$^{13}$C$_6$-labelled or $^2$H-labeled glucose, and 1 g of $^{15}$N-labeled ammonium chloride per litre of culture as the sole carbon and nitrogen sources. As it was found that the cells could not survive at low densities in a 100% D$_2$O environment, the cells were grown to a high cell density in 2 ml and then 25 ml volumes before being added to the 1 L culture to create a target cell density of $A_{600}$= 0.1 OD. These growth steps took approximately 24 and 15 hrs, respectively. The cells were then grown for ~6 hrs to reach a target cell density of $A_{600}$= 0.4 OD, at which point the expression of PR was induced by the addition of IPTG to a concentration of 1 mM. Protonated retinal was
added exogenously at this time and twice after at intervals of ~7 hrs to a final concentration of 7.5 μM. After ~21 hrs the cells were collected by centrifugation, then pre-treated with lysozyme (12 mg/L of culture) and DNAase I (600 units/L culture) before being broken by sonication. The membrane fraction was then solubilized in 1% Triton X-100 at 4 °C, and purified following the batch procedure described in the Qiagen Ni\(^{2+}\)-NTA resin manual. Approximately 15 mg of PR was purified from one litre of culture. The molar amount of PR was determined by the absorbance of opsin-bound retinal, using the extinction coefficient of 44,000 M\(^{-1}\)cm\(^{-1}\) [58].

PR was reconstituted into liposomes, prepared by hydrating dried DMPC and DMPA mixed at a 9:1 ratio (w/w), at a protein:lipid ratio of 2:1 (w/w). The detergent solubilized protein was mixed with the liposomes and incubated for 1 h at room temperature before Bio-Beads SM (Bio-Rad Laboratories) were added for detergent removal. The functionality of the reconstituted protein has been tested using visible and FTIR spectroscopy as described previously [1, 2].

To achieve the desired level of proton back-exchange, proteoliposomes were incubated in partially deuterated buffer for ~24 hours, as was optimized previously [28], before being collected by ultracentrifugation at 150,000×g for 1 h. Reconstituted protein samples were packed into either a 2.5 or a 3.2 mm NMR rotor (4 and 7 mg of protein respectively). It was found that proton concentration at the exchangeable sites was higher than in the buffer (perhaps owing to the exchange effects with water in the air during sample packing). To minimize these effects, or to alter the proton concentration, ~10 μl of buffer was placed directly on top of the packed protein, and then the packed rotor was soaked in ~1 ml of buffer overnight at 4°C. The rotor was the removed from the buffer and blotted dry before the spacer and cap were reinserted. In previous studies on Anabaena sensory rhodopsin (ASR) [28] in which a fully protonated ASR sample was back-exchanged with 100% D\(_2\)O, this method was shown to result in a uniform
deuteration over the volume of the sample, as was evident by the complete elimination of signal from solvent exposed regions of the protein.

In the discussion below the proton concentrations discussed refer to the effective proton concentration, which was assessed based on the relative intensity of the $^{15}$N spectrum, and on the intensity of the amide proton band of the proton spectrum.

### 3.3 Assignment Strategy

The assignment strategy is based on four three-dimensional chemical shift correlation proton-detected experiments (Figure 3-1). The dipolar-based CANH and CONH experiments provide CA[i]-N[i]-H[i] intra- and CO[i-1]-N[i]-H[i] inter-residue correlations, respectively. By comparing nitrogen and proton chemical shifts these experiments can be matched to result in basic spin systems, CO[i-1]-N[i]-H[i]-CA[i]. To extend these systems into overlapping fragments which can be linked to form contiguous chains, two additional experiments, in which the N[i] and H[i] shifts are correlated to the DQ$_{COCA}$, DQ$_{CACB}$ frequencies (DQ$_{COCA}$ and DQ$_{CACB}$ denotes CO+CA and CA+CB DQ coherences), are utilized. The N[i] and H[i] shifts are correlated to the DQ$_{CACO}$[i] and the DQ$_{CACB}$[i] frequencies simultaneously in the DQCANH experiment, and to the DQ$_{COCA}$[i-1] frequency in the DQCONH experiment.

Though the DQ$_{CACO}$[i]/DQ$_{COCA}$[i-1] shifts will match between systems, in order to more confidently link systems and to obtain CB information the DQ$_{COCA}$[i-1] and DQ$_{CACO}$[i]/DQ$_{CACB}$[i] frequencies must be compared to the CO[i-1] and CA[i] frequencies respectively to calculate CA[i-1] and CO[i]/CB[i] shifts to yield extended CA[i-1]-CO[i-1]-N[i]-H[i]-CA[i]–CO[i]-CB[i] systems. These systems can then be linked through simultaneously matching CO[i]/CO[i-1] and CA[i]/CA[i-1] shifts between extended systems, as seen in Figure 1-4, and amino acid type assignments can often be determined using CB[i] information.
3.4 MAS SSNMR Spectroscopy

All NMR experiments were performed on a Bruker Biospin Avance III spectrometer operating at 800.230 MHz using either a 2.5 mm $^1$H-$^{13}$C-$^{15}$N or a 3.2 mm $^2$H-$^1$H-$^{13}$C-$^{15}$N probe set in triple resonance mode. Approximately 4 mg of UND PR was center-packed in a 2.5 mm rotor for an initial investigation of spectral quality as a function of spinning frequency and the level of back-exchange. Carbon-carbon correlation experiments and 3D chemical shift correlation spectroscopy were performed using a 3.2 mm probe. About 7 mg of UCND PR was center-packed in a 3.2 mm rotor. The effective temperature was kept at 5°C in all experiments. Typical $\pi/2$ pulses were 2.5 $\mu$s for $^1$H, 5 $\mu$s for $^{13}$C, and 7 $\mu$s for $^{15}$N. The $^1$H/X (where X is $^{15}$N
or $^{13}\text{C}$) cross-polarization (CP) [116] contact times were 2 ms, with a constant RF field of 50 kHz and 37.5 kHz on carbon and nitrogen, respectively, while the proton lock field was ramped linearly around the $n=1$ Hartmann-Hahn condition [104]. $^{13}\text{CA}/^{15}\text{N}$ and $^{13}\text{CO}/^{15}\text{N}$ band-selective transfers [103] were implemented with a contact time of 6 ms. For the $^{13}\text{CA}/^{15}\text{N}$ CP, a constant lock field of $2.5\times v_r$ ($v_r=\omega_r/2\pi$, spinning frequency) strength was applied on $^{15}\text{N}$, while the $^{13}\text{C}$ field was ramped linearly (5% ramp) around $1.5\times v_r$. For the $^{13}\text{CO}/^{15}\text{N}$ transfer, a constant lock field of $3.5\times v_r$ field strength was applied on $^{13}\text{C}$, while the $^{15}\text{N}$ field was ramped linearly (10% ramp) around $2.5\times v_r$. Optimized SPINAL64 decoupling [117] was used during indirect $^{15}\text{N}$ and $^{13}\text{C}$ chemical shift evolutions. The water suppression (WS) period was composed of a train of pulses applied at an RF field of 25 kHz and phase shifted by $\pi/2$ every 12 ms for a total of 240 ms. This strategy does not completely suppress the water signal, but rather significantly reduces it. In all proton detected experiments the recycle delay was 2.5 s.

To determine the feasibility of 3D proton-detected experiments on PR, a series of 2D $^1\text{H}-^{15}\text{N}$ experiments were performed at various back-exchanged proton concentrations and spinning frequencies using the pulse sequence seen in Figure 3-2b. The effect of deuteriation on nitrogen chemical shift values and resolution was also determined from these spectra. To determine the effect of deuteriation on carbon linewidths and chemical shifts, and to perform a primary investigation into the location of solvent-exposed regions, a 2D $^{13}\text{C}-^{13}\text{C}$ experiment using the SPC-5$_3$ sequence for $^{13}\text{C}-^{13}\text{C}$ mixing was performed using the pulse sequence seen in Figure 3-2a.
Three-dimensional CANH and CONH chemical shift correlation experiments were recorded using the pulse sequence shown in Figure 3-3a. A total of 128, 96, and 1278 points were acquired in the t_1, t_2, t_3 dimensions, leading to acquisition lengths of 6.4 ms, 12.4 ms, and 10.0 ms, respectively. Three-dimensional DQCANH and DQCONH experiments were recorded using the pulse sequence shown in Figure 3-3b. For the DQCANH a total of 440, 68, and 1994 points were acquired in the t_1, t_2, t_3 dimensions, leading to acquisition lengths of 6.4 ms, 10.5 ms, and 10.0 ms, respectively. For the DQCONH a total of 260, 64, and 1994 points were acquired in the t_1, t_2, t_3 dimensions, leading to acquisition lengths of 6.4 ms, 9.9 ms, and 10.0 ms respectively.
Figure 3-3: Pulse sequences for a) CANH and CONH and b) DQCANH and DQCONH experiments. Filled and hollow bars represent $\pi/2$ and $\pi$ pulses, respectively. Both the CANH and DQCANH experiments are identical to the CONH and DQCONH experiments, respectively, except for the position of the $^{13}$C carrier frequency which is set on CA for the CANH, and CO for the CONH, and is shifted to these values for the $^{13}$C-$^{15}$N cross polarization step in the corresponding DQ experiments. All experiments begin with a 90° pulse on both carbon and proton in order to increase initial polarization. During direct proton detection $^{15}$N-decoupling is performed using WALTZ-16.[118] In a) the following phase cycling was used $\varphi_1 = (y, y, -x, -x, -y, y, x, x)$, $\varphi_2 = (x, x, y, y, -x, -x, -y, -y)$, $\varphi_3 = (x)$, $\varphi_4 = (y, -y)$, $\varphi_{rec} = (x, -x)$ and phase-sensitive detection is obtained in the indirect $t_1$ and $t_2$ dimensions by incrementing the phases $\varphi_3$ and $\varphi_4$, respectively, by 90°. In b) the following phase cycle was used $\varphi_3 = (x, y, -x, -y)$ $\varphi_4 = (y)$, $\varphi_{rec} = (-y, y, -y, y)$ and phase-sensitive detection is obtained in the indirect $t_1$ and $t_2$ dimensions by incrementing the phases $\varphi_3$ by 45° and $\varphi_4$ by 90°, respectively. To ensure rotor synchrony in the DQCANH experiment [109, 115], the reconversion block was further modified by incrementing the entire reconversion block by one C-element as shown in b, and by adding a phase shift of 144° to $\varphi_C$ with every $t_1$ increment.
The excitation bandwidth of the SPC-5\textsubscript{3} sequence at 20.5 kHz was sufficient to excite DQ coherences between CA and CB atoms, and this consideration dictated the choice of the t\textsubscript{1} increment. In the DQCONH experiment the DQ\textsubscript{CACB} was not sampled because of the selectivity of the NCO CP transfer step. Thus, the t\textsubscript{1} increment was synchronized with one rotor cycle, States-TPPI [114] was used for phase sensitive detection, and the resulting spectral width was fully sufficient to acquire the indirect DQ\textsubscript{COCA} dimension without any spectral folding or aliasing. In the DQCANH experiment, both the DQ\textsubscript{CACO} and DQ\textsubscript{CACB} were transferred through the NCA CP transfer step. To sample both types of coherences and avoid their spectral overlap, the carrier frequency was chosen at 112 ppm, phase sensitive detection of the t\textsubscript{1} indirect dimension was done using time-proportional phase incrementation (TPPI) [112] and the t\textsubscript{1} increment was set to the length of a single C-element of the SPC-5\textsubscript{3} block. Under these conditions the DQ\textsubscript{CACO} state is sampled correctly in one half of the spectrum ranging from ~ 220-250 ppm, while the cross-peaks due to the DQ\textsubscript{CACB} coherences are aliased into the empty half from ~ 150-210 ppm. Thus, both types of coherences do not overlap and their frequencies can be extracted unambiguously. Further details of the phase cycle are given in the figure caption.

3.5 Data Processing and Analysis

Chemical shifts were referenced to DSS using the \textsuperscript{13}C adamantane downfield peak resonating at 40.48 ppm as a secondary external standard [119]. Data processing was performed using NMRPipe [120]. All 2D and 3D proton detected data was apodized with a sine-squared-bell function shifted by π/3 prior to taking the Fourier transform. Noise analysis, peak picking, and assignments were performed in the CARA [121] environment utilizing AutoLink [122].
4.0 Results and Discussion

4.1 Spectral resolution in deuterated samples

To evaluate how deuteration affects carbon spectral resolution, and to assess water accessibility of various residues, a 2D $^{13}$C-$^{13}$C correlation spectrum using SPC-53 as a mixing sequence was recorded on a UCND sample. The 2D spectrum shown in Figure 4-1a exhibits a reduced number of peaks as compared to the UCN sample shown in Figure 4-1b [1, 2]. Although water molecules are known to reside within the protein in a hydrogen bonded network with polar amino acids in the retinal binding pocket [123, 124], many residues in the hydrophobic core are well protected from contacts with water by inter-helical and protein-lipid hydrophobic interactions. They are not susceptible to the H/D back-exchange, do not contribute to the $^1$H/$^{13}$C CP spectra, and do not show up in the 2 or 3D spectra. Apparent lopsidedness of the spectrum in Figure 4-1a is related to the non-uniform excitation of $^{13}$C resonances, in which atoms more proximate to amide protons are excited to a higher degree. As a consequence, many cross peaks in the upper half of the spectrum (e.g. CB-CA, CG-CB) are weaker than their counterparts (CA-CB, CB-CG) in the lower half. One notable exception to this is observed for threonine and serine residues (Figure 4-1c), whose side chain CB atoms are likely excited by oxygen bound protons, and therefore appear in this spectra with comparable intensity to the backbone carbons.

The linewidths of individually resolved resonances in Figure 4-1a are on the order of 0.5-0.7 ppm, and are very similar to those in the UCN sample (Figure 4-1b) [1, 2]. Similarly, the linewidths of $^{15}$N resonances are on the order of 0.7 ppm (data not shown), similar to what has been seen in non-deuterated samples. It can therefore be concluded that the presence of
deuterium does not cause any significant line broadening of these nuclei, which is in agreement with previously reported observations in deuterated ubiquitin [101] and GB1 [102].

As seen in Figure 4-1c-d, the presence of deuterium produces a significant isotope effect [125], causing shifts in carbon and nitrogen resonances. The upfield deuterium-induced chemical shift changes in PR are on the order of ~0.5-1.0 ppm for CA and CB spins, while CO resonances are shifted downfield by ~0.3 ppm and HN resonances are upfield shifted by ~0.3 ppm. All shifts appear to be larger than those reported previously for GB1 [102]. Table 1 summarizes the chemical shift differences induced by the isotope effect for various amino acid types and carbon atoms.

Table 1: Summary of chemical shift differences between UCN PR and UCND PR ($\Delta \delta(D) = \delta(D) - \delta(H)$)

<table>
<thead>
<tr>
<th>Residues</th>
<th>Site</th>
<th>$\Delta \delta(D)$ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>C'</td>
<td>0.3 ± 0.1</td>
</tr>
<tr>
<td>All</td>
<td>N</td>
<td>-0.3 ± 0.2</td>
</tr>
<tr>
<td>CD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>$C_a$</td>
<td>-0.6 ± 0.1</td>
</tr>
<tr>
<td>Ala</td>
<td>$C_a$</td>
<td>-0.57 ± 0.08</td>
</tr>
<tr>
<td>Val, Ile</td>
<td>$C_a$</td>
<td>-0.71 ± 0.07</td>
</tr>
<tr>
<td>Asp, Asn, Phe, Tyr, Trp</td>
<td>$C_a$</td>
<td>-0.5 ± 0.2</td>
</tr>
<tr>
<td>Leu</td>
<td>$C_a$</td>
<td>-0.6 ± 0.1</td>
</tr>
<tr>
<td>Glu, Gln, Lys, Met</td>
<td>$C_a$</td>
<td>-0.6 ± 0.1</td>
</tr>
<tr>
<td>Thr</td>
<td>$C_\beta$</td>
<td>-0.64 ± 0.09</td>
</tr>
<tr>
<td>Val</td>
<td>$C_\beta$</td>
<td>-1.09 ± 0.02</td>
</tr>
<tr>
<td>CD₂</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>$C_a$</td>
<td>-0.60 ± 0.07</td>
</tr>
<tr>
<td>Asp, Asn, Phe, Tyr, Trp</td>
<td>$C_\beta$</td>
<td>-0.9 ± 0.2</td>
</tr>
<tr>
<td>Glu, Gln, Met</td>
<td>$C_\beta$</td>
<td>-1.1 ± 0.1</td>
</tr>
<tr>
<td>Leu</td>
<td>$C_\beta$</td>
<td>-1.0 ± 0.1</td>
</tr>
<tr>
<td>CD₃</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>$C_\beta$</td>
<td>-0.96 ± 0.05</td>
</tr>
</tbody>
</table>
Figure 4-1: 2D $^{13}$C-$^{13}$C spectra of a) UCND labeled PR and b) F,L,Y-reversely labeled, fully protonated PR [2]. Both spectra were collected at 800 MHz proton frequency with SPC-5 mixing and were processed identically using Lorentzian-to-Gaussian apodization in both dimensions. c) Enlarged Thr and Ser CA/CB regions and d) enlarged Ala CA/CB region of the spectra in a) (color) overlaid with the spectra in b) (grey).
Despite these shifts, many isolated peaks can be readily reassigned by simply comparing the 2D spectrum with the one of the UCN sample. Most of the resolved residues in the 2D $^{13}$C-$^{13}$C spectrum are located in the loop and interfacial regions of the protein, which are expected to be solvent exposed. The S/N ratios of peaks corresponding to exposed residues vary significantly. For example, A168 and A178 are both a part of the E-F loop, yet the intensity of the A168 peak relative to the A178 peak is greatly diminished in the UCND sample as compared to the UCN. Both alanines have CA chemical shifts typical of a β-strand, as predicted from the chemical shift index [1, 2], but very different backbone nitrogen shifts. While the A178 nitrogen chemical shift is 125.4 ppm (125.0 in UCND) and is consistent with a β-strand, Ala168 has an anomalous chemical shift of 117.3 ppm (117.0 in UCND), which is much lower than a typical alanine α-helical shift of 122 ± 3 ppm [126]. It is possible that such a low value of chemical shift is caused by a local electrostatic interaction, possibly with negatively charged E165 [1], and this interaction may result in a low pKa value for Ala168, short residence time of the amide proton, and reduced intensity of Ala168 cross peaks. The detection of residues that are predicted to be buried in the TM regions, e.g., C156, M189 and F234, indicates that there are intramembrane portions of the protein that are accessible to water. Further evidence for water accessibility of specific TM regions of PR is discussed below.

Proton resolution shows a strong dependence on the effective amide proton concentration and on the spinning frequency. Figure 4-2 compares 2D $^1$H-$^{15}$N correlation spectra of a UND sample collected at different effective proton concentrations and spinning frequencies. When the sample is hydrated with 100% H$_2$O and spun at a moderate spinning rate (14 kHz) spectra lack any site-specific resolution. The resolution improves dramatically at faster spinning, due to better
Figure 4-2: $^1$H-$^{15}$N heteronuclear correlation spectra collected from a U-$[^2]$H,$^{15}$N labeled sample of PR. Panels a and b were collected at a spinning frequency of 14 kHz, while c and d were collected at a spinning frequency of 28 kHz. For panels a and c the exchangeable sites were protonated at 100%, while for panels b and d the protonation level was 40%.
averaging of the dipolar interactions, and/or lower proton concentrations, due to the reduction of the proton bath. In a sample with exchangeable sites protonated at 40% and measured at 28 kHz spinning (Figure 4-2 c, d) the linewidths are in the range of 0.1-0.18 ppm, with many resolved peaks appearing in the spectrum.

Overall, proton resolution is worse than that measured in microcrystalline SH3 or GB1 samples [87, 127]. To further investigate the proton linewidth, a series of 2D $^1$H-$^{15}$N spin-echo experiments were performed to estimate residue-specific T$_2$ values. Due to overlap in the 2D spectra, T$_2$ values could be obtained for only 22 residues. T$_2$ values and their respective errors were determined using fitting in SPARKY [128]. The values measured ranged from ~3 to 6.5 ms, and the homogeneous linewidths predicted by the T$_2$ measurements were much smaller than those observed (Figure 4-3, Table 2), leading us to the conclusion that the residual linewidth is caused by structural heterogeneity.

![Figure 4-3: Comparison of predicted linewidths calculated from T$_2$ measurements to experimentally measured linewidths (see Table 2 for identity of the atoms).]
Table 2: Comparison of experimental linewidths and theoretical linewidths calculated from $T_2$ measurements on individual amino acids

<table>
<thead>
<tr>
<th>Residue</th>
<th>$1/\pi T_2$ (ppm)</th>
<th>$1/\pi T_2$ (Hz)</th>
<th>Measured linewidth (ppm)*</th>
<th>Measured linewidth (Hz)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>S184</td>
<td>0.090 ± 0.003</td>
<td>72 ± 2</td>
<td>0.23</td>
<td>180</td>
</tr>
<tr>
<td>G81</td>
<td>0.11 ± 0.05</td>
<td>90 ± 40</td>
<td>0.16</td>
<td>130</td>
</tr>
<tr>
<td>T177</td>
<td>0.110 ± 0.003</td>
<td>88 ± 2</td>
<td>0.17</td>
<td>130</td>
</tr>
<tr>
<td>G66</td>
<td>0.100 ± 0.005</td>
<td>80 ± 4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>K172</td>
<td>0.08 ± 0.03</td>
<td>60 ± 20</td>
<td>0.16</td>
<td>130</td>
</tr>
<tr>
<td>Y110</td>
<td>0.078 ± 0.008</td>
<td>63 ± 6</td>
<td>0.11</td>
<td>90</td>
</tr>
<tr>
<td>N176</td>
<td>0.129 ± 0.006</td>
<td>104 ± 5</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>G203</td>
<td>0.08 ± 0.01</td>
<td>59 ± 7</td>
<td>0.18</td>
<td>140</td>
</tr>
<tr>
<td>G169</td>
<td>0.08 ± 0.01</td>
<td>67 ± 7</td>
<td>0.19</td>
<td>150</td>
</tr>
<tr>
<td>V182</td>
<td>0.120 ± 0.003</td>
<td>96 ± 2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>G30</td>
<td>0.096 ± 0.011</td>
<td>77 ± 9</td>
<td>0.23</td>
<td>190</td>
</tr>
<tr>
<td>G144</td>
<td>0.087 ± 0.010</td>
<td>70 ± 8</td>
<td>0.17</td>
<td>140</td>
</tr>
<tr>
<td>G141</td>
<td>0.104 ± 0.009</td>
<td>83 ± 7</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>G171</td>
<td>0.094 ± 0.003</td>
<td>75 ± 2</td>
<td>0.19</td>
<td>150</td>
</tr>
<tr>
<td>A116</td>
<td>0.098 ± 0.009</td>
<td>79 ± 7</td>
<td>0.21</td>
<td>170</td>
</tr>
<tr>
<td>E170</td>
<td>0.084 ± 0.016</td>
<td>67 ± 13</td>
<td>0.26</td>
<td>210</td>
</tr>
<tr>
<td>A185</td>
<td>0.076 ± 0.014</td>
<td>60 ± 11</td>
<td>0.21</td>
<td>170</td>
</tr>
<tr>
<td>A115</td>
<td>0.090 ± 0.015</td>
<td>72 ± 12</td>
<td>0.18</td>
<td>140</td>
</tr>
<tr>
<td>A242</td>
<td>0.071 ± 0.003</td>
<td>57 ± 3</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>N240</td>
<td>0.088 ± 0.003</td>
<td>70 ± 3</td>
<td>0.28</td>
<td>220</td>
</tr>
<tr>
<td>A174</td>
<td>0.102 ± 0.008</td>
<td>81 ± 6</td>
<td>0.22</td>
<td>180</td>
</tr>
<tr>
<td>Q183Hε21</td>
<td>0.082 ± 0.013</td>
<td>66 ± 11</td>
<td>0.15</td>
<td>120</td>
</tr>
<tr>
<td>Q183Hε22</td>
<td>0.055 ± 0.014</td>
<td>44 ± 11</td>
<td>0.19</td>
<td>150</td>
</tr>
<tr>
<td>N176Hδ21</td>
<td>0.077 ± 0.010</td>
<td>62 ± 8</td>
<td>0.17</td>
<td>140</td>
</tr>
<tr>
<td>N176Hδ22</td>
<td>0.061 ± 0.005</td>
<td>49 ± 4</td>
<td>0.21</td>
<td>160</td>
</tr>
</tbody>
</table>

* Based upon comparisons between the linewidths measured in the different spectra obtained for $T_2$ measurements, errors in the measured linewidths were determined to be no greater than 0.02 - 0.03 ppm
4.2 Peak picking and assignment

Since spectral resolution was not sufficient to resolve most peaks in the 2D experiments, 3D experiments were necessary to unambiguously assign residues. As a compromise between sample volume and maximum achievable spinning frequency, a 3.2 mm rotor was used and experiments were performed at a spinning frequency of 20.5 kHz to avoid rotational resonance conditions between CO and CA, and CO and CB resonances. To increase signal intensity, the effective protonation at exchangeable sites was ~70%, as proton peaks were found to be still fully resolvable at this proton concentration.

Figure 4-4: 2D planes of 3D CONH and DQCONH experiments performed on UCND PR. All peaks are labeled according to the assignments. The first contour of each peak is cut at 4σ with each additional level multiplied by 1.2.

As explained in Section 3.3, basic spin systems were built through matching the CA[i]-N[i]-H[i] and CO[i-1]-N[i]-H[i] peaks of the CANH and COHN experiments, respectively. Approximately 100 of these systems were constructed and, as in the $^{13}$C-$^{13}$C spectrum discussed previously, were found to vary significantly in their S/N ratios. Approximately half of the systems could be extended by the N[i]-H[i]-DQ_CACO[i] and N[i]-H[i]-DQ_COCA[i-1] peaks.
Representative 2D planes of the 3D experiments are shown in Figure 4-4. As shown in Figure 4-5, fragments were built by simultaneously matching CO[i] and CA[i] shifts between extended systems.

As mentioned previously, the SPC-53 sequence is sufficiently broadband to excite DQ coherences between CA and CB atoms. These DQ coherences can be detected in the DQCANH experiment, and provide important information on the amino acid type through CB shifts, albeit the efficiency of DQ excitation for these types of coherences is generally lower. A total of 15 systems could be extended by adding the N[i]-H[i]-DQCA[i] peaks, which allowed for the assignment of the amino acid type, or narrowing it down to fewer choices.

Through the identification of Gly, Ala, Ser, Thr, Val, and Ile residues and the categorization of the remaining residues, 25 residues divided among three contiguous fragments (Figure 4-6a) could be unambiguously assigned. The fragment M140-M146 was identifiable by the unique G-X-X-G-I motif present from G141-I145. These assignments were strengthened by the amino acid type identification of A143 determined through CB data. The fragment G169-S179 was identified through the G-X-G-X pattern, which is unique to this segment. These assignments were strengthened by the amino acid type identifications of S173, A174, N176, and S179, which were determined through the CB data extracted from the DQCANH. The A181-N187 fragment was identified through the A-V-X-S pattern from A181-S184. These amino acid types were all determined using CA and CB shifts calculated from DQCANH data.

Approximately 60 other extended systems were created and could not be assigned to the amino acid sequence due to insufficient information such as a lack of CB data or an absence of the system in either or both of the DQ spectra. Many of these systems were, however, unambiguously identified through matching the available chemical shift information (CO[i-1]-
Figure 4-5: Sequential walk from G171 to T177 through the CANH, CONH, DQCANH, and DQCONH spectra. Each strip is a $^1$H-$^13$C 2D plane cut at the $^15$N shift value indicated in each panel. The $^1$H axis is always 0.5 ppm wide and centered on the shifts labeled on the left. A dashed line separates the DQ$_{CACO}$ and DQ$_{CACB}$ regions of the DQ spectra. The axis on the DQ$_{CACB}$ portion of the spectra is folded, and thus runs from right to left. CA[i] and CO[i-1] values are detected directly in the CANH and CONH experiments, respectively, while the CA[i-1] and the CO[i] values (shown as red circles) are calculated by subtracting the CO[i-1] or CA[i] shifts from the DQ$_{COCO[i-1]}$ or DQ$_{COCO[i]}$ shifts detected in the DQCONH and DQCANH experiments, respectively. Systems are then linked together through simultaneously matching the CA[i-1] and CO[i-1] of the i$^{th}$ system to the CA[i] and the CO[i] of the i-1$^{th}$ system, as indicated with vertical lines. With the addition of the CB[i] information, which is also obtained in the DQCANH, it is possible to identify many amino acid types and to construct a backbone walk and confidently make sequential assignments.
N[i]-CA[i] or longer fragments) with those previously reported for UCN PR [1, 2] while taking into account the average deuterium induced changes in chemical shift. In addition, four peaks in the CONH and DQCONH spectra were identified as the side chain resonances of Asn176 and Gln183 through matching $^{15}$N and $^{13}$C shifts of these systems with those assigned previously.

4.3 Amino Acid Assignments and Structural Implications

The findings are summarized in Figure 4-6, which shows an overall topology diagram of PR, as well as a BR-based structural homology model. In both representations, solvent-exposed residues are colored blue. As expected, the majority of identified signals correspond to surface residues, which directly interact with solvent. Specifically, the cytoplasmic α-helical E-F loop and extracellular D-E loop are completely detectable and can be readily assigned in these experiments. Both loops give strong signals, suggesting that they have more rigid structures than other loop regions. In contrast, only a few residues could be seen in the β-hairpin forming B-C loop, with their resonances being generally weaker.

From the NMR assignment pattern, it appears that in PR water molecules can typically penetrate into the intramembrane portion of the protein to the depth of one or two helical turns (Figure 4-6a). The penetration depth appears to be larger on the cytoplasmic side, as strong signals from the cytoplasmic ends of helices C (Y110-A116), D (K126-V129), F (A181-N187), and G (F234-A242) are observed. The cytoplasmic ends of helices C, D, and F, along with the E-F helical loop, appear to form a localized cluster of solvent exposed residues (Figure 4-6b), which likely extends further out of the lipid membrane into the cytoplasmic environment, forming a turret-like structure. Water penetration is generally shallower on the extracellular side, as only the extracellular tips of helices D (M140-E142) and E (W149-F152) could be seen.
Recent studies have shown that *Anabaena* sensory rhodopsin (ASR) is likely asymmetrically located in the bilayer, with the cytoplasmic side being more exposed to solvent [28], and it appears that PR may have a similar position in the bilayer.

Interestingly, several residues that are located in the transmembrane regions of the protein were detected, implying water accessibility of portions of the protein core. Most notably, the backbone atoms from a large number of residues from helix G have been detected. A network of water accessible residues within the transmembrane region of helix G extends from the retinal - Schiff base-forming Lys231 to the extracellular interface, and includes polar residues such as Asn220, Tyr223, Asn224, Asp227, and Asn230, many of which are unique and conserved within the PR family (e.g., N224, N230) [129]. These residues are spaced with the periodicity of an α-helix, and likely form a hydrophilic face on the interior side of helix G which may line the proton-conducting pathway of PR.

Whereas proteorhodopsin and bacteriorhodopsin share some similarities in the initial stages of proton transport, i.e., retinal photoisomerization and deprotonation of the Schiff base, the proton release steps may be very different. While BR has a proton releasing E194-E204 glutamate pair (in complex with water) located at the extracellular side of the protein, PR lacks homologous polar residues. It has been hypothesized previously that PR may contain a hydrophilic cavity on its extracellular side, similar to that found in another homologous protein, xanthisrhodopsin (XR) [66], and that access to this cavity may be modulated by conformational changes in the F-G loop [1]. The hydrophilic face of helix G could be connected to this cavity by a H-bonded network, which may be a part of the proton release mechanism. Homology modeling predicts that in PR this cavity would include the polar residues Glu142 and Tyr223, both of which are seen to be solvent accessible in our experiments. Interestingly, homologs of these two
residues in XR have their sidechains hydrogen-bonded via a water molecule. Earlier data has show that the sidechain of Glu142 is protonated even at high pH [2]. It is possible that this proton is stabilized by hydrogen-bonding and is used for release to the extracellular bulk. It has also been shown that the E142Q mutation in PR produces a 15 nm spectral shift in visible absorption and suppresses conformational changes of the protein backbone in the late stages of the photocycle [130], which further supports the idea that Glu142 may play an important role in PR function.

**Figure 4-6:** a) Topological representation of PR. All filled circles (blue or green) represent assigned residues. Blue circles represent residues with backbone atoms exposed to solvent. Helices are lettered A-G from left to right. b) 3D model of PR based on homology modeling using the 3D-JIGSAW[65] program and using the BR structure (PDB ID: 1C3W) as a template. Colors are as in a) and helices are labeled by letters. Helices A, B, C, and D are behind helices E, F, and G. In both a) and b) PR is oriented with the cytoplasmic side on top.
FTIR studies on PR have shown that a H-bonded network of water molecules exists within the TM region [123, 124]. The side chain of His75, a highly conserved residue within the PR family that is predicted to be close to the photoactive site and potentially important in proton transfer [67, 131], is seen to be solvent exposed, as it is visible in the 1D $^{15}$N and 2D $^1$H-$^{15}$N heteronuclear correlation spectra of the UNCD sample (data not shown). As His75 is H-bonded to the primary proton acceptor of the Schiff base Asp97 [67], a residue known to be exchangeable by FTIR [63], at least under illumination, this result is not unexpected. Based on homology with XR, both His75 and Asp97 could be a part of the above mentioned hydrogen-bonded network supported by the polar residues in helix G and connected to the cavity facing F-G loop. In fact, such a network in XR includes the homologs of Lys231, Asp227, Asp97, and His75, and terminates by the homolog of Asn220 (Gln229), all of which are shown to be solvent accessible in these experiments. Thus, it is feasible that the injection of the Schiff base proton into a similar network in PR, enriched by additional unique Asn residues, will result in proton release on the extracellular side.
5.0 Conclusions

In this work, the feasibility of proton-detected multidimensional experiments in a perdeuterated 7TM membrane protein Proteorhodopsin, with protons reintroduced at the exchangeable sites through back-exchange, has been demonstrated. Proton resolution in PR was worse than that previously observed in SH3, GB1, or BR, [87, 89, 127] with approximately equal contributions to the linewidths coming from homogeneous and heterogeneous sources. Large heterogeneous contributions suggest the presence of structural heterogeneity, while dynamics and residual proton-proton interactions between the fully protonated retinal and amide protons of the protein may be responsible for the additional homogeneous line broadening. Even with the resolution available, the majority of resonances could be resolved site-specifically in 3D chemical shift correlation experiments, and solvent-accessible residues could be identified. Amide proton resonances were site-specifically assigned in 74 PR residues, most of which are located in the loop and termini regions of the protein, yet some residues in the protein core were found to be solvent accessible as well. In particular, the majority of the residues of transmembrane helix G were found to be H/D exchangeable, suggesting that this helix lines the proton-conducting pathway of PR.
6.0 Future Directions

In this thesis it was demonstrated that through the use of extensive deuteration and 3D experiments proton detection experiments can be used to independently assign stretches of amino acids in a relatively large membrane protein. Solvent exposed residues were found mainly in the loop and termini regions of the proteins, though several exceptions were found and peaks corresponding to solvent exposed intramembrane residues were identified.

The use of proton detection experiments can potentially extend the range of proteins currently available for study by SSNMR. The experiments presented in this thesis could be extended to obtain full protein amino acid assignments and restraints for structure calculations. To begin, with the use of protein refolding/unfolding protocols protons could be selectively introduced to solvent shielded portions of the protein. Though the development of a refolding/unfolding protocol can be quite challenging, it has nevertheless been previously demonstrated [3, 9]. If, after the reintroduction of protons to the core and the subsequent refolding, the protein were put into fully deuterated buffer, the protein core could be studied in isolation from the solvent exposed regions, and spectra would benefit from a reduction in peak overlap. By combining experiments on such a sample with those done on proteins with protons selectively introduced at exchangeable sites in solvent exposed regions, it is possible that the entire protein backbone could be assigned independently using proton detected experiments. This could reduce the need for multiply selectively $^{13}$C-labelled samples and simplify the assignment process. Experiments on samples with both protonation schemes could then be performed to collect $^1$H-$^1$H distances and the resulting data used to determine high-resolution protein
structures. Such experiments have recently been used to solve a high-resolution structure of GB1 [127].

Proton detection experiments could be used to selectively study functionally relevant side chain protons in membrane proteins, which could include those found in important hydrophilic cavities and channels conducting ions and other polar substrates. Residues involved in proton conduction generally have exchangeable protons in their side chains, and by selectively $^1$H and $^{13}$C-labeling these amino acids in an otherwise deuterated sample studies could be targeted to these residues specifically. Protocols for the specific incorporation of protons to side chain methyls have been demonstrated [132] and protocols could be developed for the selective incorporation of protons into other amino acid types. Through line broadening could be expected due to the full protonation of individual amino acids, it is likely that the enhanced signal obtained from the increased available initial polarization along the side chain and the isolation of these residues due their $^{13}$C labeling will compensate for this.

As previously noted, the observed proton linewidths in the presented spectra are broader than those previously reported [87, 89, 127]. Several things could be done to investigate the origin of this broadening. By creating a sample identical to the one used in these studies with deuterated retinal, which could be created chemically or biosynthetically [133], incorporated into the protein, the linewidth contribution of the protonated retinal could be determined by comparing linewidths obtained on this sample to those we obtain in this study. Secondly, investigations into the motions experienced by amino acids could be performed in order to determine the line broadening effect caused by these processes. For example, using a Lee-Goldberg CP experiment the $^1$H-$^{15}$N dipolar coupling strength can be measured site specifically
and line shape analysis can be used to provide information on the amplitude and geometry of motions [134].

The sensitivity of the presented proton detected experiments could potentially be improved through the implementation of gradient magnetic fields for water suppression [135, 136], and narrower linewidths could be obtained through the use of deuterium decoupling [79]. Linewidths could also be further reduced through the use of faster MAS (~60 kHz), though signal loss could be expected due to the required lower sample volume, and the resulting pressure changes in the sample could potentially cause structural changes in the protein [137]. This technique would also likely require the selection of a different pulse sequence for DQ excitation and reconversion, due to power limitations. However, a wide range of synchronous helical pulse sequences is available [138] and a suitable sequence could be developed to specifically suit our purposes. We have begun initial investigations into the use of these techniques, but significant technical issues still need to be addressed.
7.0 References


83.
81.
80.
79.
78.
77.
76.
74.
73.
72.
71.
70.
69.
68.
67.
66.
65.
64.
63.


