Amino Acid-Type Specific Incorporation of Stable $^{17}$O Isotopes into

Yeast Ubiquitin

BY

BINYANG LIN

A thesis submitted to the Graduate Program in Chemistry
in conformity with the requirement for the
Degree of Master of Science

Queen’s University
Kingston, ON, Canada
September 2019
Copyright © Binyang Lin, 2019
Abstract

Stable isotopic labeling is an important part of structural biology by multidimensional NMR spectroscopy. Novel stable isotopic labeling methods have continuously been developed to enhance protein NMR spectroscopy. Protein structural determination based on $^{15}$N- and $^{13}$C-labelled proteins is already well established. However, oxygen, being one of the most abundant elements in biological molecules, has not yet been utilized in NMR studies of proteins. This is largely because of two factors. First, the only NMR-active oxygen isotope, $^{17}$O, has a very low natural abundance (0.037%). Second, $^{17}$O has a nuclear spin of 5/2 (known as quadrupolar), which usually results in broad $^{17}$O NMR signals even for small organic molecules. Despite the technical difficulties, recent studies have shown $^{17}$O NMR to be a valuable tool for studying biological molecules. In this thesis, we used yeast ubiquitin (8.6 kDa) as a model system to investigate general strategies of incorporating $^{17}$O-labelled amino acids into proteins. We have used an auxotrophic $E.coli$ strain DL39 GlyA λDE3 ($aspC^-\ tyrB^-\ ilvE^-\ glyA^-\ λDE3$) to successfully synthesize yeast ubiquitin where glycine, phenylalanine, tyrosine, and alanine residues are selectively $^{17}$O-labelled. This is the first demonstration of $^{17}$O-labeling of a protein in both backbone groups and sidechain. Some preliminary solid-state $^{13}$C and $^{17}$O NMR results on these protein samples are also reported in this thesis. The synthetic methodology for $^{17}$O-labeling yeast ubiquitin reported in this thesis will be useful in future solid-state $^{17}$O NMR studies.
Acknowledgement

I would like to sincerely thank my supervisor, Dr. Gang Wu, for his generous help and patience during my study for the past two years. This thesis would not be possible without his knowledge, support and inspiration. Your trust, encouragement and dedication since my undergraduate studies have made my research experience very memorable and rewarding. I have had the opportunity to broaden my horizons in different areas and learnt making the most out of the available resources. I want to thank Dr. Xianqi Kong for his selfless help in teaching me synthetic techniques at the early stage when I was a summer volunteer. His work ethics and professionalism motivated me to go into the research field. Dr. Steven P. Smith is truly appreciated for his collaboration and enlightening consultations in my project.

Special thanks go to Alison Upsdell and Holly Spencer for their kind help in teaching me lab skills in biochemistry field. Many thanks to my housemate, colleague, but more a good friend, Sherry Dai, for cheering me up when I had doubts in myself and teaching me NMR theory. I am very grateful to other lab mates, Jiahui Shen, Andrew Rinald, Keegan and Marina in Dr. Gang Wu’s and Dr. Smith’s lab for making my lab experience full of joy and laughter.

None of my achievement would have been possible without the support of my family and friends. Their love, compassion and mental support has cheered me up whenever I faced a challenge. I am thankful for the best parents in the world. Life without me in China must be hard for them, but they still insisted on sending me all the way to North America and respect whatever decision I made. I would like to give a special thanks to my husband, Chris Wang, for all his support all along accompanying me setting up NMR experiments and purifying proteins at late nights. Bearing my mood swings whenever I encountered obstacles doing research requires lots of patience and I truly appreciate it.
I would like to express my gratitude to Dr. Francoise Sauriol, Dr. Jiaxi Wang, Dr. Kim Munro for their assistants in NMR, ESI-MS and CD spectrometry. Dr. Wang and Kim always taught me hands-on tricks on instruments operations which helped me to find the first job in the chemistry I would like to thank my committee members for their valuable suggestions throughout my studies. Finally, I would like to thank Queen’s University, Department of Chemistry, Natural Sciences and Engineering Research Council (NSERC) for providing funding for my research.
# Table of Contents

Abstract .................................................................................................................................................. i

Acknowledgement ................................................................................................................................. ii

List of Figures ......................................................................................................................................... vi

List of Tables .......................................................................................................................................... viii

List of Symbols and Abbreviations ......................................................................................................... ix

Chapter 1 ................................................................................................................................................ 1

Introduction ............................................................................................................................................ 1

1.1 A Brief Historical Review of Biological NMR Spectroscopy ......................................................... 1

1.2 Stable Isotopic Labeling Methods for Protein NMR Spectroscopy ................................................ 5

1.2.1 Biosynthetic Uniform $^{15}$N- and $^{13}$C-labeling ................................................................... 5

1.2.2 Segmental Labeling .................................................................................................................. 9

1.2.3 Amino Acid-Type Selective Isotopic Labeling ....................................................................... 9

1.2.4 Reverse Labeling ..................................................................................................................... 11

1.2.5 Cell-free Expression Systems ................................................................................................. 11

1.2.6 Site-Specific Labeling ............................................................................................................ 12

1.2.7 $^{17}$O Labeling of Proteins ...................................................................................................... 13

1.3 Background of Solid-State $^{17}$O NMR Spectroscopy ................................................................... 14

1.4 Thesis Objectives ............................................................................................................................. 18

1.5 Organization of the Thesis ............................................................................................................ 19

Chapter 2 ................................................................................................................................................ 20

Overview of NMR Theory ..................................................................................................................... 20

2.1 Basic NMR Principles ..................................................................................................................... 20

2.2 Nuclear Spin Interactions .............................................................................................................. 23

2.2.1 Chemical Shifts ....................................................................................................................... 23

2.2.2 Dipolar Coupling and NOE .................................................................................................... 25

2.2.3 Quadrupolar Coupling ........................................................................................................... 27

2.3 Multidimensionality in Protein NMR Spectroscopy ....................................................................... 29

2.3.1 One-Dimensional NMR ....................................................................................................... 29

2.3.2 Two-Dimensional Heteronuclear NMR ............................................................................... 31

Chapter 3 ................................................................................................................................................ 33

Synthesis of Amino-Acid Selectively Labelled Yeast Ubiquitin ............................................................ 33

3.1 Overview .......................................................................................................................................... 33

3.2 Experimental .................................................................................................................................... 36

3.2.1 Synthesis of $^{17}$O Isotopically Labelled Glycine Hydrochloride ............................................. 37

3.2.2 Conversion of Glycine-$^{17}$O Hydrochloride to Glycine-$^{17}$O .............................................. 37

3.2.3 Expression and Purification of Yeast Ubiquitin ....................................................................... 38
3.2.4 Uniform Labeling of Yeast Ubiquitin ................................................................. 39
3.2.5 Selective $^{17}$O-Labeling of Yeast Ubiquitin by Residue Type ................................. 40
Table 1. Selective labeling of yeast ubiquitin with $^{17}$O-labelled amino acids ................. 42
3.2.6 Salting-Out of the Selectively $^{17}$O-Labelled Yeast Ubiquitin ...................................... 42
3.2.7 NMR-Monitored pH Titration .................................................................................. 42
3.2.8 NMR Spectroscopy .................................................................................................. 43

3.3 Results and Discussion ............................................................................................... 44
  3.3.1 Determination of Oxygen Isotopic Distribution in Synthesized Glycine-$^{18}$O and
      Glycine-$^{17}$O ............................................................................................................. 44
  3.3.2 One Bond Isotopic Shift in $^{13}$C NMR Spectrum of Glycine-$^{18}$O and
      $^{13}$C-$^{17}$O Phenylalanine .......................................................................................... 48
  3.3.3 $^{17}$O NMR Spectrum of Glycine-$^{17}$O .................................................................... 50
  3.3.4 Expression and Purification of Ubiquitin ................................................................. 51
  3.3.5 Isotopic Labeling of Yeast Ubiquitin ...................................................................... 54
  3.3.6 Ubiquitin $^{17}$O NMR pH-Titration .......................................................................... 65
  3.3.7 Solid-State NMR Spectra of Residue-Specifically Labelled Ubiquitin ......................... 71

3.4 Conclusion .................................................................................................................... 75

Chapter 4 Conclusions and Future Work ........................................................................... 77
References .......................................................................................................................... 81
List of Figures

Figure 1. The first $^1$H NMR spectrum of ethanol. The peaks from left to right represent OH, CH$_2$, and CH$_3$ protons. ................................. 2

Figure 2. Amino acid biosynthesis in E. coli. Reproduced from reference 15.............................. 11

Figure 3. Solid-state $^{17}$O NMR spectra at 77K of (A) $^{[17]O_2}$-hemoglobin, frozen solution; (B) $^{[17]O_2}$-hemoglobin, polyethylene glycol microcrystals; (C) $^{[17]O_2}$-myoglobin, polyethylene glycol microcrystals. The overall spectral covers ~ 4000 ppm and the position of the major singularity is positioned at ~ 850 ± 70 ppm. Reproduced from Reference 85.............................. 17

Figure 4. (A) Schematic representation of Larmour precession of a nuclear spin in magnetic field $B_0$. The dash circles represent the precession of the spin I about the applied magnetic field $B_0$. (B) Energy level diagram for nucleus with spin number of $\frac{1}{2}$................................................. 22

Figure 5. Schematic representation of an 1D NMR experiment.......................... 23

Figure 6. Representation of the dipolar coupling between two nuclei. R represent the internuclear distance between the two active nuclei. $\theta$ is the angle between two dipoles. ............. 25

Figure 7. Demonstration of the detection of nuclear Overhauser effects in 1D NMR spectrum. The reference spectrum contains four signals labelled “1” to “4” from left to right. Reproduced from reference 15 .......................................................... 26

Figure 8. Nuclear charge distribution of a spin 12 nucleus (in orange) and a quadrupolar nucleus (in blue).......................................................... 27

Figure 9. First (HQ1) and second (HQ2) order perturbations on energy levels of spin -5/2 nuclei.......................... 29

Figure 10. 1D $^1$H NMR spectrum of human ubiquitin. The chemical shift ranges in various types of protons are indicated and labelled accordingly. 89 .................................................. 30

Figure 11. 1D NMR spectrum of sucrose (A) with normal pulse sequence, and (B) with excitation sculpting technique. Reproduced from reference 91................................. 31

Figure 12. 3D Crystal Structure of Ubiquitin (PDB ID: 1UBQ) with secondary structures indicated ($\alpha$1 including residues 23-34, and $\alpha$2 including residues 57-60) and $\beta$-strands ($\beta$1 including residues 1-7, $\beta$2 including residues 11-16, $\beta$3 including residues 41-45, $\beta$4 including residues 48,49, and $\beta$5 including residues 66-70). The six glycine, two alanine, two phenylalanine and one tyrosine residues are labelled with the three-letter amino acid abbreviation with position specified in pink, green, purple and yellow, respectively. 112 .................................................. 35

Figure 13. Comparisons of calculated isotopic distribution of oxygen isotopes to the positive ESI-MS data of (A) glycine-$^{18}$O, and (B) glycine-$^{17}$O. In (A), the three peaks represent the positive ion fragment of $[^{16}O_2]$ Gly, $[^{18}O,^{16}O]$ Gly and $[^{18}O_2]$ Gly, respectively. In (B), the third peaks represent $[^{17}O_2]$ Gly and/or $[^{16}O,^{18}O]$ Gly, while the other four peaks represent $[^{16}O,^{16}O]$ Gly, $[^{16}O,^{17}O]$ Gly, $[^{17}O,^{18}O]$ Gly, and $[^{18}O,^{18}O]$ Gly, respectively, from left to the right. ..... 46

Figure 14. Comparisons of calculated isotopic distribution of oxygen isotopes to the positive ESI-MS data of (A) $[^{13}C,^{17}O]$-glycine, (B) $[^{13}C,^{17}O]$-alanine, (C) $[^{13}C,^{17}O]$-phenylalanine, and (D) $[^{13}C,^{17}O]$-tyrosine. The bars in black represent the experimental relative abundance, whereas the grey bars represent the simulated isotopic distribution of the oxygen isotopes.............................. 48

Figure 15. Carboxylic acid region from the 125 MHz-$^{13}$C NMR spectrum ($B_0 = 11.7$ T) of 10 mg glycine-$^{18}$O in 450 $\mu$L D$_2$O, pH 4, with a data collection time of 36 min. Peaks at 172.29 ppm (green), 172.26 ppm (purple), and 172.23 ppm (dark green) represent $^{13}C$-$^{16}O_2$Gly, $^{13}C$-$^{16}O_2$O$_2$Gly, and $^{13}C$-$^{16}O_2$Gly, respectively. The red trace is the sum of green, purple, and blue signals.......................... 49
**Figure 16.** Carboxylic acids region from the 125 MHz-13C NMR spectrum of 9.6 mg [13C17O2]-phenylalanine in 450 μL D2O. Peaks at 171.68 ppm (light green), 171.67 ppm (dark green), 171.65 ppm (purple), and 171.62 ppm (grey) represent 13C-16O-Phe, 13C-16/17O2-Phe, and 13C-16/17O-Phe, respectively. The red trace is the sum of light green, dark green, purple and grey signals.

**Figure 17.** 68 MHz-17O NMR spectrum of 8.5 mg glycine-17O in 500 μL D2O, with a data collection time of 11 min and reference to water (0 ppm).

**Figure 18.** SDS-PAGE analysis of (A) uniformly 15N-labelled and (B) [13C, 17O]-Glycine selectively labelled ubiquitin. Shown above is the Ni2+ purification product migration along SDS-PAGE after expression (Lanes 1-7 in A, Lanes 1-6 in B) and TEV-cleavage. Lane 1, flow-through containing unbounded material; lane 2, wash fraction in 25 mM Tris, 250 mM NaCl, pH 8.0; Lanes 3-7 (A) and Lane 3-6 (B), elution fractions containing His-TEV-ubiquitin; Lane 8 (A) and Lane (7), fraction containing cleaved ubiquitin after TEV cleavage, in 25 mM Tris, 250 mM NaCl, pH 8.0, and 250 mM imidazole; Molecular weight standards are shown to the left of the figure, and each well was labelled by numbers from left to right. The gel was stain with Coomassie Blue.

**Figure 19.** One-dimensional 1H NMR spectrum of 225 μM uniformly 15N-labelled yeast ubiquitin in 25 mM Tris, 250 mM NaCl, pH 8.0. The H2O peak at 4.8 ppm is suppressed by excitation sculpting technique to optimize the visualization of the remaining peaks. The sample was referenced to TMS externally at 0 ppm.

**Figure 20.** Annotated 600-MHz 1H,15N HSQC NMR spectrum of (A) uniformly 15N-labelled Ubiquitin, 700 MHz 1H,15N HSQC NMR spectra of (B) 15N–glycine, (C) 15N–phenylalanine, (D) 15N–alanine labelled, and (E) 15N–tyrosine labelled Ubiquitin in 95% H2O/5% D2O, 25 mM Tris, pH 8.0, and 3000 mM NaCl. Peaks are labelled with the one-letter amino acid abbreviation with position specified for which the protein was selectively 15N labelled. The peaks are assigned according to reference 23.

**Figure 21.** Amino acid sequence of His-TEV yeast ubiquitin. The N-terminal hexahistidine tag is underlined and TEV cleavage site is shown in red.

**Figure 22.** 125-MHz 13C NMR spectra of (A) 808 μM [13C, 17O2]-glycine labelled, (B) 808 μM [13C, 17O2]-alanine, (C) 801 μM [13C, 17O2]-phenylalanine and (D) 786 μM [13C, 17O2]-tyrosine labelled ubiquitin in 95% H2O/5% D2O, 25 mM Tris, pH 8.0, and 300 mM NaCl. The peaks labelled with * are from the TEV sequence.

**Figure 23.** Generalized presentation of (a) glycine, (b) alanine, (c) phenylalanine, and (d) tyrosine amino acid biosynthetic pathways in Escherichia coli. The amino acids specifically 15N-labelled are capitalized and are boxed and in bold. The defective genes are identified in italic and correspond to the follow enzymes: glyA, serine hydroxymethyl transferase; pheA, chorismate mutase-prephenate dehydratase; tyrB, tyrosine aminotransferase; aspC, aspartate aminotransferase; ilvE, transaminase B; avtA, transaminase C. This is a modified version of a figure taken from Waugh, 1996.

**Figure 24.** Deprotonation of the phenol group of 4-[17O]-L-tyrosine (a) and the α-carboxyl group of [13C, 17O2]-glycine labelled ubiquitin (b).

**Figure 25.** 17O NMR spectra of for (a) 4-[17O]-L-tyrosine labelled and (b) [13C, 17O]-glycine labelled Ubiquitin in 25 mM Tris, 250 mM NaCl, as a function of pH. In (a), the small peak marked by * was due to the presence of SO42- ions.

**Figure 26.** Experimental (in dots) and simulated (in lines) 17O NMR pH-Titration curves for 4-[17O]-L-tyrosine labelled Ubiquitin (Blue) and [13C, 17O2]-Glycine labelled Ubiquitin.
Lines are best non-linear least squares fits to a Henderson–Hasselbalch type equation and pK_a values of 10.31 and 2.32, respectively. The parameters used to obtain the titration curves are summarized on the right. \( \Delta obs \) represents the observed chemical shift at each pH; \( \delta p \) represents the chemical shift of the fully protonated state with respect to the titration, \( \Delta \delta 1 \) is the chemical shift change associated with the titration governed by pK_a.

Figure 27. \(^{17}\)O MAS NMR spectra of His-TEV-ubiquitin where the (a) glycine and (b) phenylalanine residues are doubly labelled by \(^{13}\)C and \(^{17}\)O. The spectra were obtained at 16.4 T with a sample spinning frequency of 30 kHz, at 263 K. The total experiment time was 24 hrs.

Figure 28. \(^{13}\)C CP/MAS NMR spectra of yeast ubiquitin where F4 and F45 residues are \([^{13}\text{C},^{17}\text{O}]\)-doubly labelled. The spectra were obtained at 16.4 T with a sample spinning frequency of 8 kHz.

List of Tables

Table 1. Selective labeling of yeast ubiquitin with \(^{17}\)O-labelled amino acids. ........................................ 42
Table 2. Escherichia coli Amino Acid Metabolism ................................................................. 62
### List of Symbols and Abbreviations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>CP</td>
<td>Cross Polarization</td>
</tr>
<tr>
<td>$C_Q$</td>
<td>Quadrupolar Coupling Constant</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>DAS</td>
<td>Dynamic-Angle Spinning</td>
</tr>
<tr>
<td>DOR</td>
<td>Double-Angle Rotation</td>
</tr>
<tr>
<td>EFG</td>
<td>Electric Field Gradient</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Extinction Coefficient</td>
</tr>
<tr>
<td>FID</td>
<td>Free Induction Decay</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
</tr>
<tr>
<td>Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>His</td>
<td>Histidine</td>
</tr>
<tr>
<td>HMQC</td>
<td>Heteronuclear Multiple Quantum Coherence</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single Quantum Coherence</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny Broth</td>
</tr>
<tr>
<td>Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>MAS</td>
<td>Magic-Angle Spinning</td>
</tr>
<tr>
<td>Met</td>
<td>Methionine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MPD</td>
<td>2-Methyl-2,4-Pentanediol</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular Weight Cut-off</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance Spectroscopy</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser Effect Spectroscopy</td>
</tr>
<tr>
<td>oNBTyr</td>
<td>o-Nitrobenzyl-Tyrosine</td>
</tr>
<tr>
<td>PAIN</td>
<td>Heteronuclear Proton-Assisted Insensitive Nuclei</td>
</tr>
<tr>
<td>PAR</td>
<td>Proton-Assisted Recoupling</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>p-OMePHe</td>
<td>p-Methoxy-Phenylalanine</td>
</tr>
<tr>
<td>Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>SAIL</td>
<td>Stereo Array Isotopic Labeling</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SSI</td>
<td>Streptomyces Subtilisin Inhibitor</td>
</tr>
<tr>
<td>ssNMR</td>
<td>Solid-State Nuclear Magnetic Resonance Spectroscopy</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>TMS</td>
<td>Tetramethysilane</td>
</tr>
<tr>
<td>Tyr</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Ub</td>
<td>Ubiquitin</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 A Brief Historical Review of Biological NMR Spectroscopy

Since the 1950’s, Nuclear Magnetic Resonance Spectroscopy (NMR) has evolved into an indispensable analytical method in a variety of scientific fields, including chemistry, biology, physics, and medicine. Over the past 70 years, continuous improvements in NMR instruments and developments of new techniques are broadening the horizons of NMR applications. In the first part of the introduction, the overall evolution of NMR from physics to chemistry and finally to biology will be highlighted and some of the milestones in its history will be discussed.

NMR was simultaneously discovered by two independent groups of physicists led by Felix Bloch et al.\(^1\) and Edward Purcell et al.\(^2\) at the end of World War II. They used two completely different approaches to describe the NMR phenomenon. Purcell’s approach is inherently quantum mechanical that measured energy absorption by the proton magnetic moments in a block of paraffin, whereas Bloch’s approach recorded the current induced in a receiver copper coil from rotation of a magnetisation vector. Both approaches manifested the NMR phenomenon and the first NMR studies of solution (obtained by Bloch et al. for H\(_2\)O) and solids (obtained by Purcell et al. for paraffin) opened the road to new insights into the world of nuclear physics.

In the early 1960s, NMR was a rather insensitive method and was limited to studies of small molecules. The low-resolution magnets constructed in the early days of NMR were sufficiently homogenous and stable to allow observation of broad NMR signals from nuclei other than proton in liquids. Major effort was put mainly into the elucidation of new techniques and construction of electronic circuits to detect weak NMR signals in the presence of inevitable, thermally generated electrical noise.
Only a couple years after the initial discovery of NMR in the late 1940s and early 1950s, a few chemists accidentally discovered that proton NMR could be a powerful tool for charactering molecular structures. Proctor and Yu\textsuperscript{4,5} observed two distinct signals from the two nitrogens in NH\sub{4}NO\sub{3}, and this swiftly switched the focus of NMR towards chemistry field. The first observation of chemical shift was confirmed one year later by detecting three separate resonance lines for the OH, CH\sub{2}, and CH\sub{3} protons in ethanol, as shown in Figure 1.\textsuperscript{3} This exciting discovery demonstrates that NMR spectroscopy could be a promising analytical method for chemists. It encouraged scientists to further improve the homogeneity and stability of magnetic fields. Russell Varian made great contributions on providing a significantly easier access to this emerging field for chemists by offering commercial homogenous magnetic systems. A further gain in sensitivity was provided by the emergence of superconducting magnets in early 1960s. The introduction of Fourier Transformed (FT) NMR in 1966 by Ernst and Anderson was a big leap in the development of NMR methodologies.\textsuperscript{6} The ability to excite and combine the signals from the repeated short and intense radio-frequency pulses can significantly improve the signal-to-noise ratio. The FT NMR approach allows the detection of weak signals from NMR insensitive elements, such as \textsuperscript{13}C, \textsuperscript{15}N and \textsuperscript{17}O. Ernst opened the door to the development of other pulse sequences and took NMR to a new dimension where it becomes the most powerful tool in chemical analysis.
Meanwhile, basic NMR principles were elucidated, and new techniques were explored along with the advancement in instrument construction. The “Nuclear Overhauser Effect” (NOE) was observed in 1953 as an evidence that spins can communicate through space by direct spin-spin interactions. The signal intensity of one nucleus can change as a result of a neighbouring nucleus being subject to radio frequency (RF) irradiation. This double resonance method is important in the assignment of NMR resonances, providing spectral information about spins which then leads to structural determination of organic, and later biological molecules.

In 1957, exploratory studies were undertaken on small biological molecules. At the beginning of this journey, $^1$H NMR spectra of macromolecules, such as hemocyanin and ovalbumin, in solution did not exhibit resolved peaks. Another small protein, bovine pancreatic ribonuclease, studied by the same group, yielded first $^1$H NMR spectrum of protein by increasing the concentration and using D$_2$O instead of H$_2$O at 40 MHz. The reported $^1$H NMR spectrum exhibits four lines representing the aromatic and aliphatic protons. However, the exchangeable protons of ribonuclease were removed along with the removal of protons from water, when only using D$_2$O as the solvent. Additionally, many studies in the 1960s on other proteins such as cytochrome C and myoglobin had some resonances out of the standard chemical shift range. The water peak appeared to be the main hurdle to be surmounted for biological $^1$H NMR, as it is several orders of magnitude larger than the signals of interest.

NMR studies of large biomolecules and complexes are also complicated by the spectral overcrowding from a large number of signals. The number of signals increases with the molecular weight. This complexity leads to resonance overlap and increases the difficulty to obtain site-specific information for biomolecules like a protein, which may contain thousands of atoms. The
introduction of two-dimensional (2D) NMR by Ernst et al. in 1976 was the next evolutionary step for biological NMR spectroscopy. The additional frequency axis created 2D correlation maps between spins via nuclear spin interactions, most frequently J-coupling (between H-H, C-H or C-C bond), NOE, rates of chemical exchange or relative diffusion rates. Resonance assignment and structure determination thus could be achieved based on the information provided by the 2D NMR spectra, and this led to NMR becoming an alternative structural tool to X-ray crystallography.

The use of this new method was limited initially in structural biology compared to X-ray diffraction, for several reasons. Very high protein concentrations were required to obtain good-quality NMR spectra. Obtaining a pure, highly concentration (mM) protein sample is a major bottleneck for biochemists. Recombinant protein production using E. coli is the common method of choice when large yield of protein product is needed, but this is not always successful. Many eukaryotic proteins do not express well in the prokaryotic host. It needs other modifications of conditions to optimize the yield and to purify the inclusion bodies, which is a common insoluble form of proteins. A decrease in yield and increase in overall cost is very common after all purifications and manipulations. Besides, the high concentration of protein sample may lead to unwanted aggregation and folding artefacts. Apart from that, the experimental implementation and maintenance are high, such as periodically filling with liquid nitrogen and helium, with comparison to X-ray crystallography.

However, the use of NMR as a structural tool did strengthen over the years as a result of numerous technological advancements. New designs of electronics in NMR spectrometers have been improved and spectrometers are equipped with cryoprobes. This significantly enhance experimental long-term stability and reduces the noise by maintaining the receiver coil at low temperature. The magnetic field can reach > 950 MHz and the power of computers keeps
improving at the same time promoting the development of new processing methods. Although solving structure by X-ray crystallography is fast and the analysis is more straightforward once the single crystals are obtained, NMR spectroscopy has gain popularity in protein studies as it can provide structural information for proteins in aqueous solution, a more native state of proteins.

1.2 Stable Isotopic Labeling Methods for Protein NMR Spectroscopy

As mentioned in the previous section, the main challenge of studying large proteins is spectral overcrowding, especially when the biomolecules of interest are greater than 100 kDa.\textsuperscript{16,17} This cannot be resolved even at the highest magnetic field available. Stable isotopic labeling strategies are a major avenue to alleviate this resonance overlap by either diluting the NMR active nuclei or allowing better separation of the resonances in multiple. A selective labeling method is to introduce amino acids enriched with one or more stable isotopes, such as \textsuperscript{13}C, \textsuperscript{15}N or \textsuperscript{2}H, into proteins. Since 1960s, various improvements and novel stable isotopic labeling schemes and production techniques have continuously been developed to navigate the possibilities to enhance NMR spectroscopic resolution of larger proteins and protein complexes greater than 100 kDa.\textsuperscript{18} In the following section, we will outline the stable-isotopic labeling strategies most commonly used to obtain high-quality solution and solid-state NMR spectra of biomolecules.

1.2.1 Biosynthetic Uniform \textsuperscript{15}N- and \textsuperscript{13}C-labeling

The simplest and most-effective labeling method in biological NMR is to uniformly label all nitrogen and carbon atoms in proteins with \textsuperscript{15}N (0.37\% natural abundance) or \textsuperscript{13}C (1.109 \% natural abundance).\textsuperscript{19} Uniformly \textsuperscript{15}N and \textsuperscript{13}C labelled proteins are often overexpressed in \textit{E. coli} systems, where bacteria are grown in minimal media supplemented with \textsuperscript{15}N-labelled ammonium
chloride or ammonium sulfate, and \(^{13}\)C-labled glucose or glycerol as sole source of nitrogen or carbon, respectively.\(^{19}\) Uniform \(^{15}\)N, \(^{13}\)C-labeling has been the most widespread application for full structure determination and dynamics investigations of biomolecules. Multidimensional (2D, 3D, and 4D) correlation techniques have been developed to resolve the signal congestion in uniformly \(^{15}\)N, \(^{13}\)C-labelled proteins. In principle, all the structural constraints, dihedral angles and distances, of a single protein can be extracted from its NMR spectra through the uniformly labeling method.\(^{20}\) However, not all the proteins are amenable to this strategy. Uniform isotopic labeling, particularly uniform \(^{13}\)C-labeling, requires high structural homogeneity to give high-resolution spectra. For proteins with a lack of spectral order, uniform \(^{13}\)C-labeling is not recommended due to considerable signal overlapping. There are three spectroscopic challenges in uniform \(^{13}\)C-labeling that could explain this phenomenon. Firstly, the dispersion of \(^{13}\)C isotropic chemical shifts is limited, leading to inhomogeneous linewidths from the protein.\(^{21}\) There is also line broadening resulted from \(^{13}\)C-\(^{13}\)C scalar couplings. Moreover, it is difficult to measure long-range \(^{13}\)C-\(^{13}\)C distances in the presence of strong one-bond \(^{13}\)C-\(^{13}\)C dipolar coupling due to dipolar truncation effect.\(^{20}\) Therefore, uniform \(^{15}\)N-labeing has seen more applications as it possesses less complications.

Uniform \(^{15}\)N-labeling provides better spectral dispersion than \(^{13}\)C because \(^{15}\)N has a much larger chemical shift range. In addition, the \(^{15}\)N isotopes can be incorporated into the backbone and side-chain amide, amine, imidazole, indole, and guanidino groups of proteins.\(^{22,23}\) Every \(^{15}\)N atom in amino acids, whether it is positioned in the backbone or side chain, is at least two bonds away from another \(^{15}\)N atom. This simplifies spectral interpretation since no \(^{1}J_{NN}\) couplings would be observed, preventing the multi-exponential relaxation behavior that is difficult to be accurately measured.\(^{19}\) In the late 1980’s, Fesik and Zuiderweg proposed a 3D NMR pulse scheme where a
\(^{1}\text{H} - ^{1}\text{H}\) homonuclear NOE 2D experiment is inserted into the detection period of \(^{1}\text{H} - ^{15}\text{N}\) HMQC.\(^ {24}\) It provides better separation of the unresolved signals in the \(^{1}\text{H} - ^{1}\text{H}\) 2D-NMR spectra by moving them into separate planes along the third axis (the \(^{15}\text{N}\) chemical shift axis), resulting in reduced spectral complexity. However, some limitations should still be noted although \(^{15}\text{N}\) labelled proteins play an important role in current protein dynamic studies. Specifically, the amide \(^{15}\text{N}\) relaxations are insensitive to motion when residues are located at the hydrophobic core and not all the motions of the backbones can be monitored by amide relaxation.\(^ {20}\) Additionally, only one third of the protein backbone atoms and 6 out of 20 of the amino acids’ side chains are composed of nitrogen. Therefore, \(^{15}\text{N}\) relaxation experiments could not provide a full picture of backbone and side chain motions a protein undergoes.\(^ {23}\) Nevertheless, the simplicity of the sample preparation for \(^{15}\text{N}, ^{13}\text{C}\) protein labeling along with the isolated spin system (\(^{1}\text{H} - ^{15}\text{N}\) and \(^{1}\text{H} - ^{13}\text{C}\)) makes uniform \(^{15}\text{N}, ^{13}\text{C}\) labeling an essential method in structural and dynamic studies.\(^ {24}\)

On the other hand, it is desirable to obtain information about \(^{13}\text{C}\) chemical shifts. Uniform \(^{13}\text{C}\)- and \(^{15}\text{N}\)-double labeling method is an extension of uniform \(^{15}\text{N}\) labeling of proteins. Since both backbone and side chains of amino acids contain carbon atoms, \(^{13}\text{C}\) chemical shifts provide more information on protein folding and stabilities than amide nitrogens.\(^ {21}\) Especially, the addition of \(^{13}\text{C}\) chemical shifts enables the study of methyl groups.\(^ {25}\) Methyl groups in amino acids, such as leucine, isoleucine and valine, are often buried in the hydrophobic core of protein tertiary structures and play an important part in protein folding. Moreover, chemical shift changes resulted from specific spin-relaxation experiment are easier to be understood by studying \(\text{C} \alpha\) carbon in addition to amide nitrogens.\(^ {25}\) For uniform\(^ {13}\text{C}\) - and \(^{15}\text{N}\)-double labeling, all of the sites in proteins are visible on NMR. As a result, heteronuclear through-bond correlations can be monitored by multidimensional triple resonance NMR experiments. The \(^{13}\text{C}\) and \(^{15}\text{N}\) chemical shifts provide an
addition plane from \(^1\)H, leading to separation of \(^1\)H–\(^1\)H planes. Nowadays, HNCA, CBCA(CO)NH, HNCO and HNCACB (or CBCANH) experiments are strategies routinely employed in structural determination of proteins.\(^{25}\) Uniform \(^{13}\)C- and \(^{15}\)N- double labeling of proteins combined with 3D NMR spectroscopy provide an insight of the chemical shifts of \(^1\)H, \(^{13}\)C and \(^{15}\)N in the backbone and thus enabling the determination of the connectivities of the backbone resonances in relatively small proteins (<25 kDa).\(^{26}\)

For larger proteins, the strong \(^1\)H–\(^1\)H dipolar and \(^1\)H–\(^{13}\)C/\(^1\)H–\(^{15}\)N heteronuclear relaxation pathways are the sources of sensitivity loss when carbon atoms and nitrogen atoms are uniformly labelled. To overcome this problem, another isotope, deuterium (\(^2\)H) has been introduced. \(^2\)H (I = 1) has a quadrupolar nucleus with a significantly smaller gyromagnetic ratio compared to \(^1\)H. The listed pathways are largely eliminated when the proteins are partially or completely deuterated. Triple uniformly labelled \(^2\)H–\(^{13}\)C–\(^{15}\)N proteins are now routinely produced and used for resonance assignment in protein NMR spectroscopy.\(^{27}\) Uniform deuteration is achieved by growing cells in minimal media containing deuterated water and deuterated carbon sources.\(^{28}\) However, complete deuteration appears to have inconveniences. Most pulse sequences end with the detection of proton resonance as an effective mean to enhance sensitivity. The absence of \(^1\)H lowers the quality of the spectra, but more importantly, it inhibits the detection of the structurally important \(^1\)H–\(^1\)H NOE connectivities. Luckily, for most soluble proteins, amide deuterons exchange with water protons during the purification step. For larger protein, long correlation times often lead to line broadening problem. Uniform deuteration is more challenging as the amide deuterons in the interior hydrophobic core of proteins are usually involved in strong hydrogen bonding which leads to reduced accessibility. Therefore, partial deuteration is demonstrated to improve resolution and sensitivity to observe short-range NOE contacts in large proteins.\(^{29}\)
1.2.2 Segmental Labeling

Uniform $^{15}$N-labeling and $^{13}$C- and $^{15}$N-double labeling is preferentially used in NMR studies if small proteins. The molecular weight limitation of biological NMR can be overcome using the segmental isotopic labeling method. Large proteins usually consist of multiple domains having different functions in protein-ligand interactions, which may complicate spectral assignment of individual signals. Stable isotopes can also be incorporated into sample proteins by post-translational modification. Expressed protein ligation and protein trans-splicing are the two main approaches used in segmental labeling of proteins, performed by inteins.\(^{30}\) The unique segmental isotopic labeling technique allows only specific segment within the protein to be selectively studied by NMR spectroscopy, while the rest of the proteins remain unlabelled [9]. It largely reduces the spectral complexity for large proteins and allows the elucidation of protein-ligand interaction as well as protein dynamics.\(^{30}\) Despite the clear advantages to using segmental labeling for large proteins, successful implementations of this method in NMR are still limited, possibly because it is technically demanding and requires more time and reagents than conventional methods.

1.2.3 Amino Acid-Type Selective Isotopic Labeling

A selective isotopic labeling strategy is based on incorporation of isotopes at specific sites along the polypeptide sequence. This simplifies spectral complexity for assignment purposes. The easiest and the most widely used method is the isotopic labeling of specific residue type using $^{13}$C and/or $^{15}$N-labelled amino acids. The $^{13}$C and/or $^{15}$N-labelled amino acids are implemented in the minimal media for the bacterial expression system along with all the other unlabelled amino acids. It was first explored by Kainosho and Tsuji in 1980’s.\(^{31}\) They developed this amino acid-type
selective double labeling method to facilitate signal assignment. They tried to assign the three $^{13}$CO signals from three methionine residues located at position 70, 73, and 103 in the amino acid sequence of a protein, SSI (Strepto-myces subtilisin inhibitor). The succeeding residues of the three Met residues were Cys71, Val74 and Asn104, respectively. Amino acid-type selective double labeling refers to the introduction of two different isotopes into the same amino acid sequence at two different residues. In this experiment, Kainosho and Tsuji used $^{13}$C isotope to label the Met residues and $^{15}$N isotope to label either Cys or Val residues. The $^{13}$CO signal that displays splitting is the [1-$^{13}$C] Met connected to the adjacent $^{15}$N labelled Cys or Val, due to $^{13}$C–$^{15}$N heteronuclear one-bond scalar coupling. This is a straightforward method for resonance assignment and therefore can be applied to NMR studies of large proteins.

While this technique is widely used in current research, it still has its disadvantages, which limits the application of this methodology. The main limitation is the mis-incorporation of $^{15}$N isotopes in undesired amino acids. The metabolic inter-conversion of amino acids from one to the other in the biosynthesis pathway of the cell is the main contributor of this problem, referred as “isotopic scrambling”. There are two ways to circumvent this problem. The first is to use mutated bacterial strains, called auxotrophs. These expression strains are engineered so that the metabolic pathways leading to the synthesis of unwanted amino acids are disrupted. The severity of isotopic scrambling descends from amino acids higher up in the metabolic pathway to the lower end products as shown in Figure 2. For example, only a single lesion is needed to minimize isotopic scrambling when selectively labeling amino acids cysteine (Cys), proline (Pro), glutamine (Gln), arginine (Arg), glycine (Gly), isoleucine (Ile), lysine (Lys), histidine (His), methionine (Met) and threonine (Thr), as they are located at the end of the metabolic pathway and are not the precursors for other amino acids.
1.2.4 Reverse Labeling

An alternative solution for isotopic scrambling is reverse labeling which does not require auxotrophic bacterial strains. With this approach, a labelled general carbon or nitrogen precursor with unlabelled amino acids are included in the growth medium. The cells will use the unlabelled amino acids for protein synthesis during bacterial growth, but they will use the $^{15}$N and/or $^{13}$C precursors ($^{15}$N-ammonium chloride and/or [2-$^{13}$C]-glycerol) to make up the missing amino acid(s). Compared to the traditional approach, reverse labeling minimizes isotopic scrambling while not bring any disadvantages to the protein expression.32

1.2.5 Cell-free Expression Systems

Besides cellular expression system, cell-free expression systems has been accepted as one of the most practical production methods for isotope-labelled proteins. A cell-free expression method employs in-vitro transcriptional/translational systems extracted from an organism
(Escherichia coli, wheat germs or a mixture of recombinant proteins) reconstituted in the reaction vessel along with all the 20 amino acids.\textsuperscript{33-35} Enzymes other than those necessary for transcription and translation are absence in the system. Because of the reduced amino acid metabolic activity in vitro, isotopic scrambling is nearly eliminated for most amino acids. It should be noted that another problem frequency encountered in the standard cell-free expression system and cellular expression system is isotopic dilution by the incorporation of the endogenous amino acids present in the E. coli extract into the target protein, resulting in lower labeling efficiency. Hence, an extra gel filtration step is usually added into the protocol to ensure the overall isotopic enrichment.

Cell-free systems not only provide opportunities for advanced isotopic labeling of proteins, they are also implemented in recent ingenious technologies. For example, Stereo Array Isotopic Labeling (SAIL) has demonstrated to be a potent technique to elucidate the structures of large proteins with advantages seen in NOESY spectra. The SAIL-labelled proteins are prepared by the cell-free method. The non-exchangeable side-chain protons, methyl and methylene groups specifically, are stereo-selectively replaced by $^2$H atoms, allowing the investigation of side-chain motions.\textsuperscript{20}

1.2.6 Site-Specific Labeling

Another promising approach to study large proteins is known as Site-Specific $^{13}$C/$^{15}$N labeling. Unnatural amino acids, such as $^{13}$C/$^{15}$N-labelled p-methoxy-phenylalanine (p-OMePHe) and $^{15}$N- labelled o-nitrobenzyl-tyrosine (oNBTyr), are incorporated into a specific position along the primary sequence of a protein to produce isotopically labelled amino acids \textit{in vivo} by mutagenesis.\textsuperscript{36} In this process, novel orthogonal amber tRNA/tRNA synthetase pair was evolved to generate tRNA charged with the desired isotopically-labelled unnatural amino acids, and
incorporates it into a protein at the nonsense codon UAG. In principle, any amino acid in the target protein can be substituted *in vivo* by unnatural amino acids. Monitoring the chemical shift of this specific unnatural amino acid at the desired position of the target protein is utilized as a novel tool in drug development. It provides more insights on the binding within the active site, site-directed screening for binders, and structures of protein-ligand complexes. The unnatural amino acid site-specific labeling technique is a promising strategy to expand NMR studies to more proteins that are not currently amenable because of their large sizes and complexity.

1.2.7 $^{17}$O Labeling of Proteins

The widespread occurrence of oxygen in biomolecules and the ubiquity of oxygen throughout living systems suggest that NMR studies of this element could provide much useful information about molecular conformation and bonding in a variety of biological molecules. However, NMR studies of $^{17}$O, the only NMR-active oxygen isotope, have progress very slowly. This is due to its natural abundance of only 0.0037 %, and its small gyromagnetic ratio which leads to inherently low sensitivity. Additionally, $^{17}$O has a spin number of $I = 5/2$ with a corresponding quadrupole interaction which often results in significantly broadened signals in NMR spectra. The low sensitivity and large line widths are even worse in solids making $^{17}$O NMR studies relatively uncommon. However, $^{17}$O isotropic chemical shift ($\delta_{iso}$) covers a wide range (~1000 ppm) in organic molecules. Furthermore, quadrupolar coupling constant ($C_Q$), and asymmetry parameter ($\eta$) can provide additional information about molecular structure. These parameters are sensitive to the electron distribution about the nucleus, which could provide information about the protonation state and its involvement in hydrogen bonds. Moreover, several methods have been developed to determine the intermolecular distances between oxygen and other nuclei and
other structural information. These suggest that oxygen could play an important role in biological NMR studies. The advent of higher magnetic fields and techniques for improving resolutions, such as faster magic-angle spinning (MAS), double-angle rotation (DOR) and dynamic-angle spinning (DAS), has led an increase in $^{17}$O NMR studies of biological molecules in recent years. However, to date, the synthesis of $^{17}$O-labeled proteins has not been reported in the literature.

1.3 Background of Solid-State $^{17}$O NMR Spectroscopy

The main experimental challenge for biomolecular chemistry is to provide detailed insight into the molecular bonding arrangement, enzyme catalysis, and changes that occur upon ligand-receptor interaction at atomic-level with high quality. NMR is an ideal biophysical experimental technique to study protein motion with timescales from picoseconds to seconds. Currently, while protein NMR studies have been widely applied to study solutions, methods are also being developed to work with solid samples. Tremendous developments in solid-state nuclear magnetic resonance spectroscopy (ssNMR) have shown its capability of providing information on the structure, dynamics, and interactions of biomolecules at atomic resolution.

Biomolecular ssNMR is increasingly integrated with other traditional structural biological techniques in recent applications, such as X-ray crystallography, electron microscopy, and solution-state NMR. It is complementary to the data accessible by other methods, as ssNMR enables the study of samples lacking a high-order symmetry or a crystalline lattice, under both static as well as dynamic condition. Significant advances in instrumentation, sample preparations, computational methods, and the developments of new methodology have made it possible to investigate selectively isotopically labelled proteins and protein complexes, to determining the
complete structure of large protein and probing the intermolecular interactions.\textsuperscript{42,43,46-48} It is also possible to study lipid bilayer membrane and membrane proteins in the lipid bilayer environment, as well as on amyloid fibre aggregates over a wide ranging timescale.\textsuperscript{49,50}

A solution NMR spectrum of a protein is typically a collection of sharp lines. The effect of anisotropic interactions, such as chemical shift anisotropy and homo- and heteronuclear dipolar couplings, are averaged out by fast molecular tumbling.\textsuperscript{51} While solution-state NMR is a sensitive probe to the chemical environment, ssNMR is also sensitive to the orientation of a molecule with respect to the applied magnetic field as the system is significantly immobilized. The major obstacle of structural determination by ssNMR is the line broadening due to anisotropic interactions. The magnitude of the anisotropic interactions correlates with the orientation within the molecule, which results in anisotropic broadening of the ssNMR signals. Meanwhile, the coherence lifetimes are shortened by these interactions, which lead to homogenous line broadening in ssNMR spectra.\textsuperscript{52} On the other hand, anisotropic interactions are a two-edge sword which can provide numerous structural information. For instance, the internuclear distance is a key reporter on structure and can be determined by dipolar interactions. Orienting the molecules within the samples with respect to the applied magnetic field, either by static or rotational alignment, can alleviate the line broadening problem which has yielded many important structures of membrane proteins.\textsuperscript{52-54}

Magic-Angle Spinning (MAS), cross polarization (CP), and proton dipolar decoupling techniques are effective ways to enhance the signals and minimize line broadening by averaging out the anisotropic interactions. Structure determination of proteins can be achieved by multidimensional CP MAS NMR spectroscopy, combined with stable isotopic labeling method to further enhance spectral resolution. For instance, Igumenova \textit{et al}.\textsuperscript{55} was able to correlate carbon
backbone and side-chain chemical shifts for uniformly $^{13}$C, $^{15}$N-enriched microcrystalline ubiquitin at high applied magnetic field strengths (800 MHz) and high magic angle sample spinning frequency (20 kHz). Applying the $^{13}$C, $^{15}$N and $^2$H isotopes labeling schemes to the folded globular ubiquitin, the spectral overcrowding is significantly reduced, and spectral resolution is increased in ssNMR. Thus, using these techniques has enabled them to obtain the optimal spectral resolution, while simultaneously preserving the structural information, such as isotropic chemical shifts.\(^{55}\)

Oxygen is directly involved and plays a key role in intra- and intermolecular interaction, specifically, in one of the most important interactions in biological processes, hydrogen-bonding. $^{17}$O NMR can provide information about dynamics and structure of amino acids both in solids and solutions. Recently, solid-state $^{17}$O NMR has drawn considerable attention in studies of hydrogen bonding and metal oxygen interactions, where crystallization is not a requirement and molecular size is greater. $^{17}$O, a nucleus with a moderate quadrupolar moment and very low natural abundance as mentioned in the previous section, is even more challenging in ssNMR than $^{13}$C and $^{15}$N. The quadrupolar interaction further contributes to the broadened line shape which cannot be completely removed by the MAS technique. In the past two decades, significant progress has been made in the field of $^{17}$O NMR studies of organic and biological molecules in aqueous solution and in the solid state.\(^{56-84}\) However, $^{17}$O NMR studies of proteins are very rare. In 1991, Oldfield et al.\(^{85}\) reported the first solid-state $^{17}$O NMR spectra of proteins, $[^{17}$O$_2]$-hemoglobin and $[^{17}$O$_2]$-myoglobin, under the stationary condition in a moderate magnetic field of 8.45 T. However, essentially all the information on chemical shift and electric field gradient tensors are lost due to the poor signal-to-noise ratio (Fig.3). They concluded that the chemical shift anisotropy dominates the observed $^{17}$O powder line shape spectra, instead of second-order quadrupolar interaction. Later
in 1999, the Oldfield group successfully obtained solid-state $^{17}$O NMR spectra for [C,$^{17}$O]-myoglobin (16.7 kDa per ligand) using MAS at a higher magnetic field, 11.7 T. The isotropic chemical shifts were compared with the calculated values from the density functional theory (DFT) method.\textsuperscript{86}

![Figure 3](image)

**Figure 3.** Solid-state $^{17}$O NMR spectra at 77K of (A) $[^{17}$O$_2]$-hemoglobin, frozen solution; (B) $[^{17}$O$_2]^{-}$-hemoglobin, polyethylene glycol microcrystals; (C) $[^{17}$O$_2]$-myoglobin, polyethylene glycol microcrystals. The overall spectral covers ~ 4000 ppm and the position of the major singularity is positioned at ~ 850 ± 70 ppm. Reproduced from Reference 85.

More recently, Zhu \textit{et al.}\textsuperscript{63,66} from our group reported the first comprehensive solid-state $^{17}$O MAS NMR spectroscopic study of large protein-ligand complexes, egg-white avidin-$[^{17}$O$_2]$biotin complex (64 kDa) and ovo-transfererin-Al\textsuperscript{III}-[$^{17}$O$_4$]oxalate (80 kDa). This study gives great encouragement for further NMR study of large proteins and protein complexes. The OTf-Al-oxalate sample (25 mg) used at 21.1 T in the study only contained 30 μg of [$^{17}$O$_4$, 50% $^{17}$O] oxalate, which is the most dilute protein system investigated by solid-state $^{17}$O NMR. Another exciting discovery was that the $^{17}$O spin-lattice relaxation time, $T_1$, was found to be very short, on the order of milliseconds, for the solid protein samples. More interestingly, the $T_1$ decreases as the
applied magnetic field increases from 11.71 T to 21.1 T. Thus, rapid data collection and a sensitivity gain by a factor of 2.5 could be achieved. Furthermore, they found the solid and solution phase of the protein samples exhibit the same structure, as the solid-state $^{17}$O NMR parameters, including quadrupole-coupling-tensor and chemical shift-tensor parameters, obtained from the $^{17}$O MAS NMR spectra, showed good agreement with those found in solution.$^{63}$ It suggests that solid-state $^{17}$O NMR spectra can be analyzed directly in the absence of solution $^{17}$O NMR spectra. It was a benchmark case to test the detection limit at high magnetic field of 21.1 T. They concluded that one should be able to obtain high quality $^{17}$O MAS NMR spectra for protein or protein-ligand complexes as large as 300 kDa per ligand with 90% enrichment and a data acquisition time of 48 hr. It encouraged more researchers to used solid-state $^{17}$O NMR as a new probe to study many other important proteins and protein-complex interactions.

1.4 Thesis Objectives

Generating 3D structures from $^{15}$N- and $^{13}$C-labelled proteins is already well established by various stable isotopically labeling methods. Recent studies have shown that $^{17}$O can be used as a new probe to provide structural restraints, structural information for large biomolecules.$^{45}$ In this thesis, we hypothesized that $^{17}$O labeling of proteins can provide additional structural information about specific amino acids within the structure and in the function of the proteins. We used yeast ubiquitin (8.6 kDa) as a model system to investigate general strategies of incorporating $^{17}$O-labelled amino acids into proteins.
1.5 Organization of the Thesis

In this chapter, we provide brief background information of biological NMR spectroscopy. This includes the historical development of protein NMR, a summary of various stable isotopic labeling methods, and more importantly, a few examples of the applications of solid-state $^{17}$O NMR in proteins. Chapter 2 is meant to provide sufficient descriptions about the fundamental theory of NMR. Essential knowledge and techniques, such as spin systems, nuclear interactions, and multidimensionality in NMR spectroscopy are discussed in this chapter. In Chapter 3, we present synthesis of amino acid-type selective $^{17}$O-labelled yeast ubiquitin. The efficient expression, labeling, and purification protocol of residue-specifically yeast ubiquitin (uniformly $^{15}$N-, selectively $^{15}$N-, and $^{13}$C, $^{17}$O- doubly labelled) are described in detail. The completion of $^{15}$N, $^{13}$C resonance assignments of the residue-specifically labelled ubiquitin will justify that the incorporation of $^{17}$O-labelled amino acids into ubiquitin following the same protocol is not a concern. We also report preliminary solution and solid-state $^{17}$O NMR results for the residue-specifically labelled ubiquitin. Chapter 4 will provide a summary and outline potential future work.
Chapter 2
Overview of NMR Theory

NMR is a physical phenomenon of resonance transition between magnetic energy levels when atomic nuclei are placed in an applied magnetic field \( (B_0) \) and an electromagnetic radiation with specific frequency is applied. The energy transferred takes place at a wavelength in the radio frequency range. The energy is emitted at the same frequency when the spins return to the ground state. The signal that matches the energy transferred is measured and further processed mathematically to yield an NMR spectrum. The position, intensity and pattern of the peaks are widely used as an analytical tool in chemistry for structural determination. It is also used to study chemical environment of atoms within a molecule as well as the physical properties at the molecular level, such as conformational change and high-pressure diffusion.

2.1 Basic NMR Principles

The basic principle behind NMR involves an intrinsic property of atomic nuclei called Spin \( (I) \). Only the nuclei with a non-zero spin number are active in NMR. Nuclei are composed of neutrons and protons which are further comprised of subatomic particles known as quarks and gluons, which have charge and spin. Nuclei with even mass and even charge number are not active in NMR. The spin of individual proton and neutron are paired with each other and thus result in \( I = 0 \) overall. When the number of protons and neutrons are both odd, the nucleus have an integer spin quantum number \( (I = 1, 2, 3,…). \) Nuclei with odd mass number have fractional spin (e.g. \( I = \frac{1}{2}, 3/2, 5/2,… \)). The common isotopes commonly used in organic chemistry and biological NMR are \( ^1\text{H}, \ ^{13}\text{C}, \ ^{15}\text{N}, \ ^{19}\text{F} \) and \( ^{31}\text{P} \), all of which have \( I = 1/2 \).
Initially, the nucleus is in the ground state when there is no external magnetic field applied. The NMR active nucleus has charge and spin, and it behaves like a bar magnet with magnetic moment $\mu$. When the nucleus is placed in an applied magnetic field ($B_0$), the angular momentum, $J$, of the spinning nucleus makes the nucleus align itself with the external applied field. The angular momentum $J$ is influenced by the strength of the applied field $B_0$, causing the nucleus to process about the $B_0$ at a frequency termed Larmour frequency ($\omega_0$) as shown in Figure 4 (A). The Larmour frequency is proportional to the applied magnetic field by a fundamental nuclear constant called gyromagnetic ratio ($\gamma$). The gyromagnetic ratio is different for different types of nuclei.

$$\omega_0 = -\gamma B_0$$ (1)

In the presence of an external magnetic field, nuclear spins orient themselves in the direction of the static magnetic field creating distinct nuclear energy levels based on the spin angular momentum of the nucleus. The number of orientation and thus the number of diverged energy level is $(2I+1)$. This phenomenon is called Zeeman splitting. Therefore, the nuclei commonly used in biological NMR studies, such as $^1$H, $^{13}$C, and $^{15}$N nuclei, all have 2 orientations as shown in Figure 4(B). The spins which align with the magnetic field are at lower in energy (-1/2), whereas those spins that align against the magnetic field are at higher energy level (1/2). The energy between the two levels is:

$$\Delta E = \frac{\gamma h B_0}{2\pi}$$ (2)

where $h$ is Plank’s constant. The magnitude of the splitting thus depends on the strength of the magnetic field. Each energy level is called Zeeman energy level and is populated by the spins which have the same angular momentum. The number of nuclei in each spin state can be represented by Boltzmann equation:
\[
\frac{N_{\frac{1}{2}}}{N_{-\frac{1}{2}}} = e^{-\Delta E / kT}
\]

Where \(N_{\frac{1}{2}}\) and \(N_{-\frac{1}{2}}\) represent the population of nuclei in the two energy states, \(\Delta E\) is the energy difference between the spin states, \(k\) is the Boltzmann constant \((1.3805 \times 10^{-23} \text{ J/K})\), \(T\) is the temperature. The population of each spin state depends on the energy differences, thus depends on the strength of the magnetic field and the nucleus type.

**Figure 4.** (A) Schematic representation of Larmour precession of a nuclear spin in magnetic field \(B_0\). The dash circles represent the precession of the spin \(I\) about the applied magnetic field \(B_0\). (B) Energy level diagram for nucleus with spin number of \(\frac{1}{2}\).

During the NMR experiment, the nucleus of interest is excited by the radiofrequency pulse at the Larmour frequency, resulting in a transition of spins from the low energy state to the high energy state. Excess population of spins at the high energy state will then relax back to the low energy state and finally reach the thermal equilibrium state. The excess population of spins produces a bulk magnetization. It exhibits an exponential decay over time after the initial radiofrequency pulse until the nuclei reach the equilibrium state, which is termed a free induction
decay (FID). The FID in the time domain is then Fourier transformed to a frequency domain with the peak positioned at the nucleus’ Larmour frequency, this give rise to the 1D NMR spectrum.

![Figure 5. Schematic representation of an 1D NMR experiment.](image)

**2.2 Nuclear Spin Interactions**

As mentioned in the previous section, the energy level of a nucleus in an applied magnetic field shows Zeeman splitting. The Zeeman interaction contributes the most in order of magnitude in the generation of energy differences. However, there are other types of interactions between the nuclear and electron spins. For the quadrupolar nucleus, there are four primary NMR interactions that make up the total NMR Hamiltonian:

\[
\hat{H} = \hat{H}_Z + \hat{H}_Q + \hat{H}_D + \hat{H}_{CS}\]

(4)

where \(\hat{H}_Z\) is the Zeeman interaction, \(\hat{H}_Q\) is the quadrupole interaction, \(\hat{H}_D\) is the dipolar interaction, and \(\hat{H}_{CS}\) is the chemical shift interaction. In general, we have \(\hat{H}_Z > \hat{H}_Q > \hat{H}_D > \hat{H}_{CS}\).

**2.2.1 Chemical Shifts**

An NMR spectrum is a plot of the energy absorption (referred to as resonances) against frequency as introduced in section 2.1. Based on Equation (1), all atoms of the same type are expected to have the same frequency. This is not the case in reality. When a nucleus is immersed
in the applied field, the electrons surrounding the nucleus also possess magnetic moment. The electrons generate their own local magnetic field around the nucleus, altering the $B_0$ field at the nuclear site. The actual magnetic field that the nucleus experiences is the combination of the external applied magnetic field and the sum of the magnetic field generated by other factors, expressed as:

$$B_{local} = B_0 - \sigma B_0$$  \hspace{1cm} (5)

where $\sigma$ is the quantitative representation of this electronic modulation, the so-called magnetic shielding effect. Magnetic Shielding is defined as “a nucleus whose chemical shift has been decreased due to addition of electron density, magnetic induction, or other effects.” In contrast, deshielding means that chemical shift of a nucleus increases due to the removal of electron density. A variety of factors can cause chemical shift, such as hydrogen bonding, magnetic anisotropy of $\pi$-systems, and inductive effects by electronegative groups. For example, protons in the fluoromethane appear to be at higher resonance frequency than those in methane. Fluorine is an electron-withdrawing group reducing the electron density around the hydrogen nuclei, so that there is less shielding and higher resonance frequency. The actual frequency of the nucleus of interest being measured referred to as chemical shift ($\delta$), expressed in parts per million (ppm). It is convenient to describe the position of the resonance in an NMR spectrum with reference to a standard compound which is defined to be at 0 ppm:

$$\Delta = \frac{\nu_{\text{sample}} - \nu_{\text{reference}}}{\nu_{\text{reference}}} \times 10^6$$  \hspace{1cm} (6)

The reference compound used for $^1$H and $^{13}$C in studies for this thesis is tetramethysilane (Si(CH$_3$)$_4$, abbreviated as TMS) and urea for $^{15}$N chemical shifts. Chemical shifts are very sensitive to steric and electronic effects and thus to secondary and tertiary structures in protein NMR spectroscopy.
By investigating the chemical shifts of the proteins, more insights can be gained on torsion angles, aromatic rings, solvent accessibility, ionic strength, and the effect of pH and temperature.

### 2.2.2 Dipolar Coupling and NOE

Dipolar coupling is also known as dipole-dipole interaction, which is the interaction between the dipoles of two molecules. In the context of NMR, it is the through-space interaction of the magnetic moments of two nuclei. It is the magnetic effect on nucleus 1 due to the magnetic field generated by nucleus 2, as illustrated in Figure 6. The dipolar coupling can happen between two of the same nuclei (homonuclear) or between two different nuclei (heteronuclear).

![Figure 6](image)

*Figure 6.* Representation of the dipolar coupling between two nuclei. R represent the internuclear distance between the two active nuclei. $\theta$ is the angle between two dipoles.

The dipolar coupling is inversely proportional to the sixth power of the distance between the two nuclei ($\frac{1}{r^6}$). The strength of dipolar coupling depends on the angle between the two spins by $(3\cos^2\theta-1)$ and relative motion relative to one another. The effect of dipolar interactions on the appearance of the NMR spectra usually is averaged by the fast tumbling of molecules in an isotropic solution and viscosity is low. Solids or liquid with high viscosity do not allow sufficiently
fast motion of the molecules, and thus the dipolar interactions tend to contribute to the shape of the spectra.

The dipolar interaction acts on the spin system through a mechanism called relaxation, which is the process that nucleus regained the thermal equilibrium after being perturbated by radiofrequencies. On the basis of dipolar interaction, nuclear Overhauser effect (NOE) was introduced to detect short-range through-space interaction. It refers to a phenomenon that the amplitude of a spin resonance alters when other nuclei are irradiated. As the example given in Figure 7, the resonance of spin 3 is saturated leading to an increase of the intensity of spin 1 compared to the reference spectrum. Spin 3 disappears whereas the multiplicity and the fine structure of the resonance of spin 1 remained unchanged. This effect is called cross-relaxation and is the evidence of the NOE between spin 1 and 3.

Figure 7. Demonstration of the detection of nuclear Overhauser effects in 1D NMR spectrum. The reference spectrum contains four signals labelled “1” to “4” from left to right. Reproduced from reference 15.
NOE is pivotal for application of NMR in structural biology. For example, not only the nature of the β sheets (parallel or anti-parallel) can be determined by the detection of NOE, but also the register (in-register or out-of-register) of the strands can also be indicated. More importantly, NOE provides a clear correlation map of the chemical shifts of the interacting nuclei when combined with multi-dimensional NMR (2D or 3D). This NOE-type of experiments are refereed as NOESY experiments. The cross peaks in the correlation maps indicate the which two nuclei are close in space, thus is a chief way for three-dimensional structure determination.

2.2.3 Quadrupolar Coupling

Most of the nuclei in the periodic table have a spin $I$ larger than $\frac{1}{2}$, which are called quadrupole spins. Unlike nuclei with spin $\frac{1}{2}$, the positive charge distribution in the nucleus is not spherical and is thus orientational sensitive to the electric field gradient (EFG) generated by the asymmetric distribution of electron density around the nucleus (Fig. 8). This unique intrinsic property of the nucleus is called electric quadrupole moment ($eQ$), and the coupling of $Q$ with the EFG is called quadrupole interaction (1-3). Like dipolar interaction, quadrupolar interaction is a ground state interaction, but it dependent on the distribution of charges in the molecule and the resulting EFGs from its surrounding.

![Figure 8](image.jpg)

**Figure 8.** Nuclear charge distribution of a spin $\frac{1}{2}$ nucleus (in orange) and a quadrupolar nucleus (in blue).
The EFGs at the quadrupolar nucleus is described as traceless tensors, therefore the quadrupolar interactions are always averaged to zero by isotopic tumbling in solution. In solid-state NMR, the e\(Q\) interact with EFG at the nuclear site to broaden the spectra significantly. The asymmetry of the quadrupolar interaction is given by asymmetry parameter (\(\eta\)), and the magnitude of the quadrupolar interaction is described by quadrupolar coupling constant (\(C_Q\)). Both \(C_Q\) and \(\eta\) contribute to the unique line shape of the solid-state NMR spectra and could be determined experimentally.

The quadrupole interaction causes the degeneracy of the nuclear energy levels to be lifted and results in \((2I + 1)\) energy levels. The size of the splitting depends on the size of EFG and the magnitude of the quadrupole moment. The quadrupolar interaction can be written as a sum of first and second-order interactions:

\[
H_Q = H_Q^1 + H_Q^2
\]  

(7)

The first-order quadrupolar interaction is proportional to \(C_Q\), whereas the second-order quadrupolar interaction is much smaller, which is inversely proportional to \(\frac{c_Q^2}{\mathcal{V}_0}\). When the quadrupolar interaction is strong enough, both two expansion terms of this interaction need to be taken into consideration. For example, the effects of first- and second-order interactions on the energy levels of a spin \(\frac{5}{2}\) are shown in Figure 9. As can be seen, the central transition (CT) is not affected by the first-order quadrupolar interaction.
Figure 9. First \( (H_Q^1) \) and second \( (H_Q^2) \) order perturbations on energy levels of spin -5/2 nuclei.

2.3 Multidimensionality in Protein NMR Spectroscopy

2.3.1 One-Dimensional NMR

In one-dimensional \(^1\)H NMR spectrum of a protein, each individual proton from the protein of interest gives rise to a single resonance for a single nucleus in the frequency domain, in the way as outlined in section 2.1. Protons within the same amino acid experience different environment at the molecular level and thus fall into different ranges as illustrated in Figure 10. For instance, the aliphatic protons are in upper field compared to the amide protons. However, protons of a particular type do not necessarily have identical chemical shifts, as the chemical shift of each \(^1\)H from the protein largely depends on its local environment in this molecule. This includes the interaction with other nuclei in the same amino acid residue, the interaction with nuclei from the neighbouring amino acid residues, the conformation of the protein and thus its exposure in contact with the solvent.
Figure 10. 1D $^1$H NMR spectrum of human ubiquitin. The chemical shift ranges in various types of protons are indicated and labelled accordingly.$^{89}$

In 1D protein NMR studies, water is of utmost importance that the natural condition of the biological samples is maintained. The enormous concentration of water, in comparison with the protein concentration, gives rise to a gigantic solvent signal dominating the 1D $^1$H NMR spectra. This makes it extremely difficult to observe and assign the spatially close $^1$H resonances in the protein. Thus, the water resonance is routinely suppressed by multiple available methods before recording of the spectrum of $^1$H in proteins. “Water suppression by Excitation Sculpting”, the so-called “Shaka sequence”, shows superior suppression quality compared to traditional presatuation or 1D NOESY methods. Its building block utilizes arbitrary wave forms and two subsequence $B_0$ field gradient echoes with binomial pulses between the gradients.$^{90}$ As seen in Figure 11, the proton NMR of sucrose without any water suppression technique has a big water peak at 4.7 ppm.
dominating the spectrum, whereas the water signal is much smaller by excitation sculpting method, allowing the observation of the sucrose signals.

![Figure 11. 1D NMR spectrum of sucrose (A) with normal pulse sequence, and (B) with excitation sculpting technique. Reproduced from reference 91.](image)

### 2.3.2 Two-Dimensional Heteronuclear NMR

It is obvious that the full assignment of each resonance in 1D $^1$H NMR spectrum of a protein is nearly impossible, especially as the protein size increases. The resonance congestion can be alleviated by increasing the dimensionally of the NMR experiments. The 2D $^1$H-$^{15}$N Heteronuclear Single Quantum Correlation (HSQC) experiment is the type of 2D experiment commonly used in this thesis to determine the success of the incorporation of the $^{15}$N-labelled amino acids into the target protein. Similar to 1D NMR experiment, the 2D $^1$H-$^{15}$N HSQC experiment is composed of preparation, evolution, mixing, and detection periods. The amide
proton (H\textsuperscript{N}) is initially excited by an applied radiofrequency pulse during preparation period. Its magnetization is then transferred to its covalently-linked amide \textsuperscript{15}N-labelled nitrogen, which is followed by the returning of the magnetisation to the H\textsuperscript{N} for detection of the spin frequencies. In this way, H\textsuperscript{N} is correlated with its adjacent amide \textsuperscript{15}N-labelled nitrogen linked by the covalent bond. The 2D \textsuperscript{1}H-\textsuperscript{15}N HSQC spectrum is a contour plot with two axes presenting two different nuclei, while the peaks representing the backbone H\textsuperscript{N} -amide \textsuperscript{15}N of a particular residue if there are no side-chain amides. Thus, isotopic enrichment of proteins with \textsuperscript{15}N allows the observation of each amino acid residue in the sequence. However, it should be noted time required to obtain high-quality 2D spectra is much longer for an isotopically labelled protein.
Chapter 3
Synthesis of Selectively $^{17}$O-Labelled Yeast Ubiquitin

3.1 Overview

When considering protein dynamics and interactions, it is important to understand the process of protein folding, which is also related to the functional properties of proteins. It is the amino acid sequence and the inter-residue interactions that influence the overall conformation of a secondary protein structure. The most common nuclei in NMR studies of proteins are $^{13}$C, $^{15}$N and $^1$H. $^{17}$O would be another alternative as its potential importance in biological NMR was recognized more than 20 years ago. One of the main reasons why we are interested in oxygen is its ubiquity in biology. Indeed, oxygen occupies a key position both at structural and functional levels. Oxygen plays a major role in the molecular conformation observed (secondary, tertiary and quaternary structure) for biological macromolecules (DNA and RNA, peptides, and proteins). $^{17}$O NMR can provide additional information about protein structure, dynamics, and functions, including hydrogen bonding environment within the structure, sites of post-translational modifications such as phosphorylation or glycosylation, and protein-ligand interactions. In proteins, structures are stabilized through a large number of hydrogen bonds. The breaking and forming of these hydrogen bonds have effect on its overall conformation, but more importantly, on enzyme catalysis. Taking into account the significance important of hydrogen bonding in biological systems, $^{17}$O NMR has great potential to be a valuable probe of both structure and function of proteins. However, because $^{17}$O has very low natural abundance as mentioned in Chapter 1, $^{17}$O-isotopic labeling is a prerequisite to any $^{17}$O NMR study.
Fiat and co-workers have presented detailed protocols on the synthesis of $[^{17}\text{O}]$-carbonyl-enriched amino acids.$^{94-96}$ The solution $^{17}\text{O}$ NMR spectra of these amino acids with $\alpha$-carboxyl group $^{17}\text{O}$-enriched gave single peaks between 249-265 ppm. The main objective of this thesis is to selectively label the $\alpha$-COOH site of glycine, alanine, phenylalanine and tyrosine through acid-catalyzed oxygen exchange experiment (Scheme 1), as well as to study the characteristics of their solution $^{17}\text{O}$ NMR. Then we will incorporate of these synthesized $^{17}\text{O}$-labelled amino acids into yeast ubiquitin.

![Scheme 1](image)

**Scheme 1.** Synthesis of $^{17}\text{O}_2$-Glycine through acid-catalyzed water-exchange experiment.

Ubiquitin is a relatively small regulatory protein that is only composed of 76 amino acids. It has a compact globular fold with a combination of several secondary structures, including five $\beta$ strands and two helices, as shown in Figure 12. Ubiquitin has found to play pivotal roles in eukaryotic cells, such as vesicular trafficking,$^{97}$ DNA repair,$^{98}$ signalling,$^{99}$ endocytosis,$^{100}$ protein proteasomal degradation$^{101}$ through ubiquitylation process. The compact globular shape of ubiquitin, with its mixed secondary structures, together with its inherent thermostability have made it an ideal test system for the development of labeling strategies and pulse sequence generation.$^{102-105}$ Many biological NMR studies have used stable isotopically labelled ubiquitin samples to test the efficiency of their NMR experiments. For example, Neilson *et al.*$^{106}$ used $^{13}\text{C},^{15}\text{N}$-labelled ubiquitin samples to demonstrate the efficiency of the pulse sequence for the proton-assisted
recoupling (PAR)/heteronuclear proton-assisted insensitive nuclei (PAIN) experiments. Furthermore, studies have been done previously using ubiquitin as a model protein to navigate the possibility of understanding protein dynamics by NMR spectroscopy. For instance, Huang et al. investigated the dynamical properties of ubiquitin’s hydration shell using deuterium NMR spectra. Extensive information about the structure and dynamics of ubiquitin is available from X-ray crystallography, solution NMR studies and a few solid-state NMR studies. We have chosen this protein to illustrate the feasibility of incorporating $^{17}$O labelled amino acids into proteins for $^{17}$O NMR studies.

![Figure 12. 3D Crystal Structure of Ubiquitin (PDB ID: 1UBQ) with secondary structures indicated (α1 including residues 23-34, and α2 including residues 57-60) and β-strands (β1 including residues 1-7, β2 including residues 11-16, β3 including residues 41-45, β4 including residues 48,49, and β5 including residues 66-70). The six glycine, two alanine, two phenylalanine and one tyrosine residues are labelled with the three-letter amino acid abbreviation with position specified in pink, green, purple and yellow, respectively.](image)

High-resolution NMR spectroscopy requires pure proteins in the millimolar range. Advances in molecular biology have made recombinant protein expression a routine technique for
introducing $^{15}\text{N}$ and $^{13}\text{C}$ isotopes into proteins. Numerous studies have reported the $^1\text{H}$, $^{15}\text{N}$ and $^{13}\text{C}$ resonances assignments for ubiquitin by multidimensional NMR techniques, either in solution-state or microcrystalline-state. Ubiquitin is widely used as a test system for optimizing structural assignment protocols. Since there is a lack of $^{17}\text{O}$ NMR studies of ubiquitin, the four target amino acids are doubly labelled with $^{13}\text{C}$ and $^{17}\text{O}$, so that the well-characterized $^{13}\text{C}$ resonances of ubiquitin can act as a probe to confirm the incorporation of $^{17}\text{O}$-labelled amino acid residues at the desired positions in the protein sequence. Moreover, before utilizing the synthesized $^{17}\text{O}$ isotopically labelled amino acids, $^{15}\text{N}$-glycine, $^{15}\text{N}$-alanine, $^{15}\text{N}$-phenylalanine and $^{15}\text{N}$-tyrosine were used to test the efficiency of the expression and purification protocol for producing a large amount of pure isotopically labelled ubiquitin for NMR studies, as the $^{17}\text{O}$-labelled amino acids are precious. Uniform and selective $^{13}\text{C}$,$^{15}\text{N}$- enriched ubiquitin were prepared by overexpression in *Escherichia coli* followed by purification procedures described by Muchmore *et al.* We will present the solution-state NMR spectra of $^{15}\text{N}$ uniformly-, $^{15}\text{N}$ residue-specifically, $^{13}\text{C}$ and $^{17}\text{O}$ residue-specifically doubly labelled yeast ubiquitin. The experimental chemical shifts are compared with the values reported in previous studies as ubiquitin is a well-characterized protein.

### 3.2 Experimental

**Materials:** All chemicals were purchased from Sigma-Aldrich unless otherwise stated. These materials were used in the condition they were obtained, without additional purification. The list of materials includes: $^{15}\text{N}$ labelled glycine (glycine-$^{15}\text{N}$, 98% $^{15}\text{N}$), $^{15}\text{N}$-labelled alanine (alanine-$^{15}\text{N}$, 98% $^{15}\text{N}$), $^{15}\text{N}$-labelled phenylalanine (phenylalanine-$^{15}\text{N}$, 98% $^{15}\text{N}$), $^{15}\text{N}$-labelled tyrosine (tyrosine-$^{15}\text{N}$, 98% $^{15}\text{N}$), $^{13}\text{C}$ labelled glycine (glycine-$^{13}\text{C}$, 99% $^{13}\text{C}$), $^{13}\text{C}$ labelled alanine
(alanine-1-\textsuperscript{13}C, 99\% \textsuperscript{13}C), \textsuperscript{13}C labelled phenylalanine (phenylalanine-1-\textsuperscript{13}C, 99\% \textsuperscript{13}C), \textsuperscript{13}C labelled tyrosine (tyrosine-1-\textsuperscript{13}C, 99\% \textsuperscript{13}C), and side-chain labelled tyrosine (tyrosine, phenol-\textsuperscript{17}O, 35-40\% \textsuperscript{17}O), and \textsuperscript{18}O-labelled water (H\textsubscript{2}\textsuperscript{18}O, 98.0\% \textsuperscript{18}O, 2\% \textsuperscript{16}O, all purchased from Cambridge Isotope Laboratories, Inc.). \textsuperscript{17}O-labelled water (H\textsubscript{2}\textsuperscript{17}O, 41.1\% \textsuperscript{17}O, purchased from CortecNet), hydrogen chloride solution (4.0 M in dioxane) and ion-exchange resin [poly-4-(vinylpyridine)]. The plasmid containing a gene coding for yeast ubiquitin (K48R mutant) with an N-terminal hexahistidine tag and a TEV (tobacco-etch virus) cleavage site, in the pET3 vector was provide by Gary Shaw’s lab at Western University. The auxotrophic bacterial strain DL39 GlyA λDE3 (aspC\textsuperscript{-} tyrB\textsuperscript{-} ilvE\textsuperscript{-} glyA\textsuperscript{-} λDE3) used for specific amino acid labeling was kindly supplied by L. McIntosh at the University of British Columbia. All other chemicals used were of the highest purity commercially available.

### 3.2.1 Synthesis of \textsuperscript{17}O Isotopically Labelled Glycine Hydrochloride

To a 2 mL solution of HCl (4 M, in dioxane) was added glycine (500 mg, 1 mmol) and H\textsubscript{2}\textsuperscript{17}O (1 g, 53 mmol) in a pressure vessel, at 70 °C for 24 hr with magnetic stirring. The reaction mixture was then evaporated to dryness by distillation, and the final yield was 96-98 \%. \textsuperscript{1}H NMR (400 MHz, dioxane): \(\delta\)\textsubscript{H}: 4.64 ppm (s, \textsuperscript{1}H). \textsuperscript{17}O NMR (67 MHz, dioxane): \(\delta\) 262.9ppm.

### 3.2.2 Conversion of Glycine-\textsuperscript{17}O Hydrochloride to Glycine-\textsuperscript{17}O

Four hundred milligrams of poly-4-vinylpyridine (3.4 meq) were packed in a Pasteur pipette (6 × 30-mm bed) above a plug of glass wool. The column was saturated with water and allowed to equilibrate for a few hours. The pipette was immersed in a room-temperature water bath and 6 mL water was passed through the resin by gentle air pressure. Glycine-\textsuperscript{17}O-HCl, 1 meq (112.5 mg), dissolved in 4 mL water was applied to the column, followed by washing with (2 × 1
mL) water. Gentle nitrogen pressure was applied such that the flow rate was about 1 mL/min. The effluent was evaporated to dryness on a rotary evaporator and further dried for analysis in a vacuum over P$_2$O$_5$. The yield was 110.2 mg (98%).

It should be noted that $^{18}$O-glycine was synthesized as a test prior to $^{17}$O-glycine following the same protocol as described above using H$_2^{18}$O. [3-$^{15}$N, 1, 1-$^{17}$O$_2$]-glycine, [$^{13}$C-$^{17}$O$_2$]-glycine, [$^{13}$C-$^{17}$O$_2$]-alanine, [$^{13}$C-$^{17}$O$_2$]-phenylalanine, and [$^{13}$C-$^{17}$O$_2$]-tyrosine were also synthesized followed the same procedure, starting with $^{15}$N labelled glycine, 1-$^{13}$C-glycine, 1-$^{13}$C-alanine, 1-$^{13}$C-phenylalanine, and 1-$^{13}$C-tyrosine, respectively.

### 3.2.3 Expression and Purification of Yeast Ubiquitin

The plasmid encoding for yeast ubiquitin (K48R mutant) in vector pET3 transformed into the E. coli DL39 GlyA λDE3 (aspC~ tyrB~ ilvE~ glyA~ λDE3) was grown in 1 L of LB media supplemented with 100 mg/mL carbenicillin and 50 mg/L kanamycin in a 1 L Fernbach flask. The plasmid together with the auxotrophic strain carry antibiotic resistance genes for carbenicillin and kanamycin. The inclusion of the two antibiotics in the cultivating culture is to isolate the bacteria containing the plasmid and the strain by artificial selection. Bacteria were grown at 37 °C with vigorous shaking at 300 rpm. When the optical density at 600 nm (A$_{600}$) reaches 0.6-0.8, 1mM IPTG was added to induce protein expression, and cell growth was continued for 3-4 h at 37 °C with shaking in the incubator.

The cells were harvested by centrifugation and resuspended in 20 mL of 25 mM Tris-HCl, pH 8, 300 mM NaCl, and 200 µL of protease inhibitors cocktail (EDTA-free) from Milipore Sigma. The cell suspension was rocked on ice for 10 min and then sonicated at 25% amplitude for 2 min with 15 sec on, 15 sec off in a Qsonica Sonicator Q500. Cell debris was removed by
ultracentrifugation at 4 °C for 40 min at 40000 rpm in a Beckman 70 Ti rotor. The supernatant was collected and applied to Ni^{2+} affinity sepharose column. The flowthrough fraction was collected and then the column was washed with a premixed buffer containing 4 mL of the elution buffer (300 mM NaCl, 250 mM imidazole and 25 mM Tris-HCl, pH 8.0) and 36 mL of the binding buffer (300 mM NaCl, 10 mM imidazole and 25 mM Tris-HCl, pH 8.0). After further eluted with 3×5 mL of elution buffer, the fractions were collected and checked by 20% SDS-polyacrylamide gel electrophoresis.

The eluted His-TEV-ubiquitin was pooled down to 300 µL and dialyzed against 25 mM Tris-HCl pH 8.0, 150 mM NaCl at room temperature overnight. 1 mL of 50 µM TEV protease was added to ubiquitin within dialysis bag to remove the hexahistidine tag. The dialyzed sample was applied to Ni^{2+} affinity sepharose column and washed with the dialysis buffer. The ubiquitin was collected in the flow-through and wash fractions after checking on 20% SDS-polyacrylamide gel electrophoresis. The two fractions containing ubiquitin were then concentrated down to 7 mL and applied to Superdex×30 PG size exclusion column. The column was equilibrated in 25 mM Tris-HCl, pH 8.0 and 150 mM NaCl). Elution was done using the same buffer and was monitored at 280 nm with a flow rate of 1mL/min. The void volume was 46.2 mL and fractions containing ubiquitin were collected between the elution volumes of 69.55 mL to 102.61 mL.

### 3.2.4 Uniform Labeling of Yeast Ubiquitin

Uniform {superscript}15N-labeling of yeast ubiquitin was achieved by growing E. coli BL21 (DE3) containing ubiquitin coding gene in M9 minimal media using {superscript}15NH₄Cl (0.75 g/L) as the sole nitrogen source.¹⁹ 25 mL of M9 minimal media² supplemented with 0.1% (v/v) of micronutrients,³ 0.01% (v/v) freshly made 10 mM FeSO₄, 0.2% (v/v) of 1 M MgSO₄, 0.01% of 1 M CaCl₂, 0.5%
(v/v) of $^{15}$NH$_4$Cl (0.75 g/L), 1% (v/v) of 40% (w/v) [$^{12}$C$_6$]-glucose, and 50 µL of 0.112 mM carbenicillin was inoculated with a single colony of freshly streaked *E. coli* BL21 (DE3) and grown overnight at 37 °C with shaking. 10 mM of the BL21 (DE3) culture was used to inoculate 1 L of M9 media containing the corresponding amounts of supplements described above. The amount of filter-sterilized carbenicillin added was increased to 400 mg for 1 L cultures. Bacteria were grown at 37 °C with vigorous shaking at 300 rpm until A$_{600}$ reaches 0.6-0.8. IPTG was added to 1 mM. Incubation at 37 °C with aeration and vigorous shaking was continued at 300 rpm for 14 hr, and then the cells were harvested. The purification and TEV cleavage of the uniformly $^{15}$N labelled yeast ubiquitin followed the same protocol as described above in the previous section.

### 3.2.5 Selective $^{17}$O-Labeling of Yeast Ubiquitin by Residue Type

The backbone and side-chain functional groups in a protein often can be distinguished by their $^{15}$N, $^{13}$C and $^{17}$O chemical shifts. $^{31}$ Selective $^{17}$O-labeling of yeast ubiquitin was achieved by growing *E. coli* DL39 GlyA λDE3 (aspC tyrB ilvE glyA λDE3) containing the plasmid encoding for yeast ubiquitin (K48R mutant) in pET3 vector on the minimal salt media supplemented with either synthesized $^{13}$C-$^{17}$O$_2$-glycine, $^{13}$C-$^{17}$O$_2$-alanine, $^{13}$C-$^{17}$O$_2$-phenylalanine, $^{13}$C-$^{17}$O$_2$-tyrosine, or $^{13}$C-4-$^{17}$O-L-tyrosine. Take glycine as an example, to specifically incorporate $^{13}$C and $^{17}$O double-labelled glycine into yeast ubiquitin, a mixture of amino acids (0.50 g glycine, 0.50 g alanine, 0.40 g arginine, 0.40 g aspartic acid, 0.05 g cysteine, 0.40 g glutamine, 0.65 g glutamic acid, 0.10 g histidine, 0.23 g isoleucine, 0.23 g leucine, 0.42 g lysine hydrochloride, 0.25 g methionine, 0.13 g phenylalanine, 0.10 g proline, 2.10 g serine, 0.23 g threonine, 0.17 g tyrosine, and 0.23 g valine) are used to make up a 1 Litre minimal salt media. It should be noted that 250 mg synthesized $^{13}$C-$^{17}$O$_2$-glycine in place of 0.50 g unlabelled glycine, and 3.1 g of serine instead
of 2 g from the list shown above are used to selectively label the glycine residues in ubiquitin. The quantity of the labelled amino acid used is different once it becomes the target of labeling. Table 1 gives a summary of the amount of amino acids used in our synthesis. A list of nucleic acid and salts (0.50 g adenine, 0.65 g guanosine, 0.20 g thymine, 0.50 g uracil, 0.20 g cytosine, 1.50 g sodium acetate, 1.50 g succinic acid, 0.50 g NH₄Cl, 0.85 g NaOH, and 10.50 g KH₂PO₄) were also added. Water was added to make it up to 1 litre. After autoclaving, 50 mL of 20% glucose, 2 mL of 1 M MgSO₄, 1.0 mL of 0.01 M FeCl₃, and 10 mL of a filter-sterilized solution containing 2 mg CaCl₂ • 2H₂O, 2 mg ZnSO₄ • 7H₂O, 2 mg MnSO₄ • H₂O, 50 mg L-tryptophan, 50 mg thiamine, 50 mg niacin, 1 mg biotin, and 100 mg ampicillin, 50 mg kanamycin are added under sterile conditions. The final solution is near pH 7.2. 50 mL of this minimal bacterial culture with a single colony of freshly grown E. coli DL39 GlyA λDE3 (aspC· tyrB· ilvE· glyA· λDE3) was used to inoculate 1 L of selective labeling media. 1 mM IPTG was added when A₆₀₀ near 0.6, and cell growth was continued at 37 °C with aeration and vigorous shaking for an additional 4 hours. The purification and TEV cleavage of the selectively labelled yeast ubiquitin followed the same protocol as described above.
Table 1. Selective labeling of yeast ubiquitin with $^{17}$O-labelled amino acids.

<table>
<thead>
<tr>
<th>Residue Type</th>
<th>$^{17}$O-Amino Acid</th>
<th>Quantity of amino acid/liter</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>L-Alanine</td>
<td>600 mg</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
<td>250-375 mg</td>
<td></td>
</tr>
<tr>
<td>Gly, Ser</td>
<td>Glycine</td>
<td>500 mg</td>
<td>No serine added</td>
</tr>
<tr>
<td>Phe</td>
<td>L-Phenylalanine</td>
<td>75 mg</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>L-Tyrosine</td>
<td>35 mg</td>
<td></td>
</tr>
</tbody>
</table>

*From Reference 19

3.2.6 Salting-Out of the Selectively $^{17}$O-Labelled Yeast Ubiquitin

5 mg of $[^{13}$C,$^{17}$O]-Phe-Ubiquitin was dissolved in 200 $\mu$L of buffer (250 mM NaCl, 25 mM Tris, pH 8) to make up the concentration to 2.2 mM. ($\text{NH}_4\text{SO}_4$ was added to reach 35% saturation initially and white needle-like microcrystals started to form. The addition of ($\text{NH}_4\text{SO}_4$ continued to reach 65% saturation until no more microcrystals formed. The microcrystals were collected by Amicon Ultra milipore concentrator (MWCO = 3 kDa) and then transferred to the 2.5-mm Bruker MAS probe by centrifugation at 10 G for 5 min.

3.2.7 NMR-Monitored pH Titrations

The pH-dependent chemical shifts of the sidechain phenol group of the 4-$^{17}$O-L-tyrosine labelled ubiquitin (2 mM) were measured by solution-state $^{17}$O experiments recorded at 37 °C with a Brucker 700 MHz NMR spectrometer. The samples, initially at pH ~ 7.07, were titrated by addition of 0.1 M HCl in microlitre aliquots. The pH-dependent chemical shifts of the mainchain carboxyl group of the $^{13}$C, $^{17}$O-glycine labelled ubiquitin were measured by $^{17}$O experiments.
recorded at 25 °C with a Brucker 500 MHz NMR spectrometer. The samples, initially at pH~ 4.24, were titrated by addition of 0.1 M NaOH in microlitre aliquots. Sample pH values were measured at room temperature (~21 °C), and not corrected for temperature or isotope effects. Each spectrum was acquired and processed by backward linear prediction with the first 16 points replaced by the predicted points (TDoff = 16) using Topspin 4.0.6. Experimental and theoretical titration curves were simulated fitting to the equations outlined in Equation 8. The precision of each fitted was adjusted and estimated by Excel Solver.

\[ \delta_{obs} = \frac{\Delta \delta_1}{1 + 10^{(pK_a - pH)}} + \delta_p \]  

(8)

where \( \delta_{obs} \) represents the observed chemical shift at each pH; \( \delta_p \) represents the chemical shift of the fully protonated state with respect to the titration, \( \Delta \delta_1 \) is the chemical shift change associated with the titration governed by pK_a

### 3.2.8 NMR Spectroscopy

All solution NMR samples contained either lyophilized unlabelled, uniformly \(^{15}\)N-labelled or selectively \(^{17}\)O-labelled yeast ubiquitin dissolved in 500 µL 95 % H₂O / 5 % D₂O were adjusted to pH 8.0 -8.1 by the addition of 1 M HCl, except for the titration section. Protein concentrations of uniformly and selectively labelled ubiquitin were determined by \( A_{280} \) and the extinction coefficient of yeast ubiquitin (\( \varepsilon = 1490 \text{ M}^{-1}\text{cm}^{-1} \)).

The \(^1\)H and \(^{13}\)C NMR spectra were acquired on Bruker Avance-500, -600, and -700 MHz spectrometers. Suppression of the H₂O resonance was accomplished with the use of excitation sculpting technique. The \(^1\)H and \(^{13}\)C NMR NMR spectra were referenced to the tetramethysilane (TMS) at 0.0 ppm. All \(^{17}\)O experiments were performed on Bruker Avance-500 and -700, and the chemical shifts were referenced to liquid water at -4.6 ppm. Solution samples were prepared in 5-
mm NMR tubes and Bruker broadband solution probes were used to record the NMR spectra. $^1$H-$^{15}$N HSQC spectra of uniformly labelled ubiquitin were collected on an Oxford AS600 spectrometer equipped with a pulsed field gradient (PEG) triple resonance probe at 25 °C and on a Bruker Avance-600 MHz spectrometer, using the following acquisition parameters: spectral width: 9615 Hz ($^1$H); 10647 Hz ($^{15}$N), acquisition time: 7.43 hr, carrier frequencies: 6.5 ppm ($^1$H); 125.0 ppm ($^{15}$N). The nitrogen resonance was referenced to the urea at 7.71 ppm.

3.3 Results and Discussion

3.3.1 Determination of Oxygen Isotopic Distribution in Synthetized Glycine-$^{18}$O and Glycine-$^{17}$O

Scheme 2 shows the reaction mechanism of the oxygen exchange experiment. To test the experiment, we typically started with $^{18}$O-labelled water, because it is cheaper than $^{17}$O-labelled water. Synthesized $^{18}$O-labelled glycine contains a mixture of $^{18}$O and $^{16}$O oxygen isotopes, as the purchased $^{18}$O-labelled water contains 98% $^{18}$O and 2% of $^{16}$O. Synthesis of $^{17}$O-labelled glycine utilized $^{17}$O-labelled water as the starting material, which has the following isotopic composition: $\text{H}_2^{17}\text{O}$, 41.1% $^{17}$O, 52.7% $^{18}$O, and 6.2% $^{16}$O. Therefore, the $^{17}$O-labelled product is in a mixture of $^{16}$O, $^{17}$O and $^{18}$O, as either or both oxygen atoms are likely to be labelled with $^{18}$O and $^{17}$O, respectively.
Scheme 2. The reaction scheme of acid-catalyzed $^{18}$O-exchange experiment. The $^{18}$O from $^{18}$O-labelled water is colored in red.

The oxygen isotopes have the same number of protons and electrons, but a different number of neutrons resulting in different masses. Thus, the isotopic pattern of $^{18}$O, $^{17}$O and $^{16}$O in glycine can be determined by positive ion mode ESI-MS spectrometry. The molecular ion, [M+H]$^+$, of glycine-$^{18}$O displays three adjacent peaks attributable to ions of different isotopic compositions as seen in Figure 13A. The amount of each constituent element can be calculated, shown in Figure 13A, taking into consideration the isotopic abundance for C, N, H, $^{18}$O, and $^{16}$O. As mentioned earlier, the water used for the synthesis of glycine-$^{18}$O only contains 98% $^{18}$O and 2% of $^{16}$O. Thus, the $^{17}$O peak was absent in the MS data. In contrast, three oxygen isotopes were involved in the $^{17}$O-exchange experiment from the purchased $^{17}$O water (H$_2$O$^{17}$O, 41.1% $^{17}$O, 52.7% $^{18}$O, and 6.2% $^{16}$O). As a result, more possible combinations of oxygen isotopes led to the observation of more peaks in the MS, as seen in Figure 13B.
Figure 13. Comparisons of calculated isotopic distribution of oxygen isotopes to the positive ESI-MS data of (A) glycine-$^{18}$O, and (B) glycine-$^{17}$O. In (A), the three peaks represent the positive ion fragment of [${^{16}O_2}$]Gly, [${^{18}O,^{16}O}$]Gly and [${^{18}O_2}$]Gly, respectively. In (B), the third peaks represent [${^{17}O_2}$]Gly and/or [${^{16}O,^{18}O}$]Gly, while the other four peaks represent [${^{16}O,^{16}O}$]Gly, [${^{16}O,^{17}O}$]Gly, [${^{17}O,^{18}O}$]Gly, and [${^{18}O,^{18}O}$]Gly, respectively, from left to the right.

The relative intensity of each peak in the positive ESI-MS spectrum depends on the chemical formula of the ionic fragment and is proportional to the probability of occurrence of each mass. This probability is the sum of probabilities of all combinations resulting in the same nominal mass. For instance, the peak at m/z =78.04 (Figure 13B) can be attributed to the sum of [${^{17}O_2}$]-Gly and [${^{16}O,^{18}O}$]-Gly, as they both have the same nominal mass of 78 g/mol. Therefore, the relative intensity of isotope peaks (and their overall pattern as well) provide a useful means for the identification of isotopic composition. By creating the best match of the simulated isotopic distribution with the experimental isotopic pattern (shown in Figure 13), the enrichment level of $^{18}$O was determined to be 54.8 %. Since the $^{18}$O level is 98.0% in the starting material ($^{18}$O-H$_2$O), the overall efficiency of the $^{18}$O-exchange experiment was 56.0 %. The percentage of $^{17}$O in the glycine-$^{17}$O was determined in the same fashion, which is 20% (the $^{17}$O-labelled water used has
$^{17}$O, 41.1% $^{18}$O, 17.8% $^{16}$O). The efficiency of $^{17}$O-exchange experiment is 50%, which is consistent to the $^{18}$O experiment, indicating the water exchange protocol is highly reproducible.

Therefore, the same protocol was applied to synthesize $^{13}$C and $^{17}$O double-labelled glycine, alanine, phenylalanine and tyrosine. The $^{13}$C contributions to the weighting of the m/z intensities are also taken into consideration when creating the simulated spectrum. By adjusting the ratios of the oxygen isotopes until the simulated spectrum resembled the experimentally obtained spectrum, the percentage of $^{17}$O in the $[^{13}$C, $^{17}$O ]-glycine, $[^{13}$C, $^{17}$O ]-alanine, $[^{13}$C, $^{17}$O ]-phenylalanine and $[^{13}$C, $^{17}$O ]-tyrosine was determined to be 21.6%, 16.7%, 41% and 29.3%, respectively by ESI-MS. The detailed ESI-MS results are shown in Figure 14.
Figure 14. Comparisons of calculated isotopic distribution of oxygen isotopes to the positive ESI-MS data of (A) $[^{13}\text{C}, ^{17}\text{O}]$-glycine, (B) $[^{13}\text{C}, ^{17}\text{O}]$-alanine, (C) $[^{13}\text{C}, ^{17}\text{O}]$-phenylalanine, and (D) $[^{13}\text{C}, ^{17}\text{O}]$-tyrosine. The bars in black represent the experimental relative abundance, whereas the grey bars represent the simulated isotopic distribution of the oxygen isotopes.

3.3.2 One Bond Isotope Shift in $^{13}$C NMR Spectrum of Glycine-$^{18}$O and $^{13}$C,$^{17}$O-Phenylalanine

Each peak in a $^{13}$C NMR spectrum identifies a carbon atom in a different environment within the molecule. A small change in the $^{13}$C chemical shift was observed when the $^{13}$C atom is bonded to different isotopes of the same element. For example, $^{13}$C = $^{16}$O versus $^{13}$C = $^{18}$O. One-bond isotope shifts arising from different oxygen isotopes are usually very small, which is only observable at higher magnetic fields. Figure 15 shows the $^{13}$C NMR signal for glycine-$^{18}$O at 125
MHz ($B_0 = 11.7$ T). It can be observed in Figure 15 that the one bond isotope shift of the $^{13}$C atom attached to two $^{16}$O versus $^{16}$O/$^{18}$O is $0.028 \pm 0.002$ ppm. The substitution of $^{16}$O with $^{18}$O resulted in the peak shifted upfield. The presence of the isotope shift in the $^{13}$C NMR spectrum of glycine-$^{18}$O further confirmed that the oxygen exchange did take place.

Figure 15. Carboxylic acid region from the 125 MHz-$^{13}$C NMR spectrum ($B_0 = 11.7$ T) of 10 mg glycine-$^{18}$O in 450 μL D$_2$O, pH 4, with a data collection time of 36 min. Peaks at 172.29 ppm (green), 172.26 ppm (purple), and 172.23 ppm (dark green) represent $^{13}$C-$^{16}$O$_2$-Gly, $^{13}$C-$^{16/18}$O$_2$-Gly, and $^{13}$C-$^{18}$O$_2$-Gly, respectively. The red trace is the sum of green, purple, and blue signals.

In the case of [$^{13}$C,$^{17}$O$_2$]-Phenylalanine, the overall shape of the peak is as a result of one-bond isotopic shifts arising from $^{16}$O, $^{18}$O and $^{17}$O. As seen in Figure 16, one-bond isotope shift of $^{13}$C attached to two $^{16}$O (light green) versus $^{16}$O/$^{18}$O (purple) is determined be to $0.027 \pm 0.002$ ppm, which is consistent with result obtained from glycine-$^{18}$O. However, one bond isotope shift arising from the $^{13}$C atom attached to $^{17}$O was difficult to observe, because of the line broadening from the $^{17}$O relaxation, as seen in Figure 16 in the dark green trace.
50

**Figure 16.** Carboxylic acids region from the 125 MHz $^{13}$C NMR spectrum of 9.6 mg $[^{13}$C$^{17}$O$_2]$-phenylalanine in 450 µL D$_2$O. Peaks at 171.68 ppm (light green), 171.67 ppm (dark green), 171.65 ppm (purple), and 171.62 ppm (grey) represent $^{13}$C-$^{16}$O$_2$-Phe, $^{13}$C-$^{16/17}$O$_2$-Phe, and $^{15}$C-$^{16/18}$O-Phe, $^{13}$C-$^{18}$O$_2$-Phe, respectively. The red trace is the sum of light green, dark green, purple and grey signals.

### 3.3.3 $^{17}$O NMR Spectrum of Glycine-$^{17}$O

Another method of determining the level of $^{17}$O in the $^{17}$O-labelled compound is by $^{17}$O NMR. As shown in Figure 17, the $^{17}$O NMR spectrum of glycine-$^{17}$O (8.5 mg in 500 µL D$_2$O) displays the glycine-$^{17}$O signal at 260.02 ppm. The integration ratio of this peak and the water signal is 1.8 : 1. The percentage of $^{17}$O in the glycine-$^{17}$O and in the water signal are proportional to their integrations, and thus the level of $^{17}$O in glycine-$^{17}$O was determined to be 16.3 % from the $^{17}$O NMR spectrum shown in Figure 17. Moreover, the position of the chemical resonance of glycine-$^{17}$O can be used for further $^{17}$O NMR studies.
Figure 17. 68 MHz-\textsuperscript{17}O NMR spectrum of 8.5 mg glycine-\textsuperscript{17}O in 500 µL D\textsubscript{2}O, with a data collection time of 11 min and reference to water (0 ppm).

3.3.4 Expression and Purification of Ubiquitin

The plasmid encoding for yeast ubiquitin coding gene (K48R mutant), a TEV cleavage site, and a N-terminal hexahistidine tag, in a pET3 vector was transformed into the auxotrophic DL39 GlyA λDE3 (asp\textsuperscript{C} tyrB\textsuperscript{-} ilvE\textsuperscript{-} glyA\textsuperscript{-} λDE3) cells\textsuperscript{116}. The overexpression of yeast ubiquitin is followed by Ni\textsuperscript{2+} affinity chromatography to yield large quantity of uniformly \textsuperscript{15}N-labelled and residue-specifically labelled yeast ubiquitin, as described in Section 3.2.3. Elution fractions containing His-TEV-ubiquitin were pooled and concentrated to about 30 mM, then dialyzed overnight with TEV protease. Subsequent Ni\textsuperscript{2+} affinity chromatography was used to further eliminate any trace of uncleaved ubiquitin. Both purification results were monitored by SDS-PAGE (Fig. 18). All the elution fractions from purification of \textsuperscript{15}N uniformly labelled glycine (Fig. 18A, lane 3-7) His-TEV ubiquitin contain intense bands, which migrated similarity to His-TEV ubiquitin (MW\textsubscript{obs} 11.3). The wash fraction collected from Ni\textsuperscript{2+} affinity chromatography after
cleavage of $^{15}$N uniformly labelled His-TEV-ubiquitin (Figure 18A, lane 8) only shows cleaved ubiquitin ($MW_{obs}$ 8.5) indicating the completion of TEV cleavage.

![Figure 18](image)

**Figure 18.** SDS-PAGE analysis of (A) uniformly $^{15}$N-labelled and (B) $[^{13}$C, $^{17}$O]-Glycine selectively labelled ubiquitin. Shown above is the Ni$^{2+}$ purification product migration along SDS-PAGE after expression (Lanes 1-7 in A, Lanes 1-6 in B) and TEV-cleavage. Lane 1, flow-through containing unbounded material; lane 2, wash fraction in 25 mM Tris, 250 mM NaCl, pH 8.0; Lanes 3-7 (A) and Lane 3-6 (B), elution fractions containing His-TEV-ubiquitin; Lane 8 (A) and Lane (7), fraction containing cleaved ubiquitin after TEV cleavage, in 25 mM Tris, 250 mM NaCl, pH 8.0, and 250 mM imidazole; Molecular weight standards are shown to the left of the figure, and each well was labelled by numbers from left to right. The gel was stain with Coomassie Blue.
Based on the SDS-PAGE analysis, loss of protein during purification appeared to be minimal. Bands migrating at the molecular weight of yeast ubiquitin were very faint in the lanes containing discarded fractions such as pellets and other impurities. The expression protocol therefore was employed for the incorporation of selectively $^{15}$N-, phenol-$^{17}$O tyrosine, and the four synthesized $^{13}$C, $^{17}$O-labelled amino acids (glycine, alanine, phenylalanine, and tyrosine) into yeast ubiquitin. As can be seen in Figure 18, $[^{13}$C, $^{17}$O]-Gly-His-TEV-ubiquitin was obtained after protein expression and purification (Lanes 3-6). Cleaved $[^{13}$C, $^{17}$O]-Gly-ubiquitin (MW$_{obs}$ 8.5) is shown in Lane 8, indicating the success of the protein preparation. After dialysis and lyophilization, dry weight yields of 28-32 mg of uniformly $^{15}$N-labelled and ~23-25 mg of residue-selectively $^{13}$C, $^{17}$O-labelled yeast ubiquitin were routinely purified from a one litre culture.

Preliminary one-dimensional $^1$H NMR experiments were obtained on both uniformly/selectively $^{15}$N- and residue-specifically $^{13}$C, $^{17}$O-labelled ubiquitin in solution. The isotopically labelled yeast ubiquitin at 1.5 mM concentration was stable at physiological pH and room temperature for weeks. The one dimensional $^1$H NMR spectrum shown in Figure 19 contains peaks with line width for an 8.5 kDa protein ($\Delta \nu_1 \approx 7$ Hz) with good dispersion, indicating the protein is well behaved at this condition. Thus, 2D $^1$H-$^{15}$N HSQC, $^{13}$C and $^{17}$O NMR experiments could be carried for the purpose of isotopically labelled amino acids assignments.
**Figure 19.** One-dimensional $^1$H NMR spectrum of 225 μM uniformly $^{15}$N-labelled yeast ubiquitin in 25 mM Tris, 250 mM NaCl, pH 8.0. The H$_2$O peak at 4.8 ppm is suppressed by excitation sculpting technique to optimize the visualization of the remaining peaks. The sample was referenced to TMS externally at 0 ppm.

### 3.3.5 Isotopic Labeling of Yeast Ubiquitin

Labeling either uniformly or selectively with $^{15}$N was used as a general starting point prior to $^{13}$C and $^{17}$O selectively labeling. The expression protocol developed for uniformly labeling ubiquitin with $^{15}$N was very similar to that described by Muchmore *et al.* $^{19}$ Uniform $^{15}$N-labeling of yeast ubiquitin in *E. coli* BL21 (DE3) cells grown in M9 minimal media using $^{15}$NH$_4$Cl (0.75 g/L) as the sole nitrogen source. $^{116}$ As expected, due to growth on supplemented minimal media rather than normal LB media, the yields of uniformly isotope-labelled protein were lower than
those obtained for unlabelled protein. However, one litre of induction culture provided enough uniformly $^{15}$N-labeled protein necessary for NMR analysis.

To ensure that the incorporation of the $^{15}$N isotope or the $^{15}$N-labelled amino acids into ubiquitin was effective, a $^{1}$H-$^{15}$N HSQC spectrum was collected for each type of $^{15}$N-labelled protein. The $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N uniformly (Figure 20A) and $^{15}$N-selectively labelled ubiquitin (Figures 20B-E) reveal good dispersions of peaks along the amide proton dimension with uniform intensity. Figure 20 (A) shows a $^{1}$H-$^{15}$N HSQC spectrum of uniformly $^{15}$N-labelled ubiquitin in which one would expect to find the same number of peaks as there are backbone and side chain amide protons in the protein. It clearly indicated that incomplete incorporation of the $^{15}$N isotope into ubiquitin was not a concern. The $^{1}$H-$^{15}$N HSQC spectrum for the $^{15}$N-labelled glycine (Figure 20B) clearly displayed six distinct glycine resonances on top of the spectrum and were identified based on the unique backbone amide nitrogen chemical shift of each residue (G10, G35, G53, G47, G75, G76). Figures 20D and 20E illustrate the $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N–phenylalanine labelled and $^{15}$N–tyrosine labelled ubiquitin. Since there are two phenylalanine (F4 and F45) and one tyrosine (Y59) residues in ubiquitin one would expect finding the same number of peaks in their corresponding spectra. While this was the case for $^{15}$N–labelled phenylalanine (Figure 20C) and $^{15}$N–labelled tyrosine ubiquitin (Figure 20E), the $^{1}$H-$^{15}$N HSQC spectra of $^{15}$N–labelled alanine ubiquitin contained two resonances (Figure 20D). In addition to the one alanine residue in the ubiquitin sequence (A46), another alanine residue is from the SNA sequence, between the TEV-cleavage sequence and ubiquitin sequence (Figure 21).
Figure 20. Annotated 600-MHz ¹H-¹⁵N HSQC NMR spectrum of (A) uniformly ¹⁵N-labelled Ubiquitin, 700 MHZ ¹H-¹⁵N HSQC NMR spectra of (B) ¹⁵N–glycine, (C) ¹⁵N–phenylalanine, (D) ¹⁵N–alanine labelled, and (E) ¹⁵N–tyrosine labelled Ubiquitin in 95% H₂O/5% D₂O, 25 mM Tris, pH 8.0, and 3000 mM NaCl. Peaks are labelled with the one-letter amino acid abbreviation with position specified for which the protein was selectively ¹⁵N labelled. The peaks are assigned according to reference 23.
Figure 21. Amino acid sequence of His-TEV yeast ubiquitin. The N-terminal hexahistidine tag is underlined and TEV cleavage site is shown in red.

The auxotrophic strains showed success in labeling ubiquitin with $^{15}$N-glycine, $^{15}$N-alanine $^{15}$N-phenylalanine and $^{15}$N-tyrosine with no unwanted $^{15}$N-incorporation into other amino acids. Thus, it was then employed for $^{13}$C, $^{17}$O- labelled glycine, $^{13}$C, $^{17}$O- labelled alanine, $^{13}$C, $^{17}$O- labelled phenylalanine and $^{13}$C, $^{17}$O- labelled tyrosine labeling of ubiquitin with same results (Figure 22). The $^{13}$C NMR spectrum of the cleaved ubiquitin-[$^{13}$C, $^{17}$O$_2$]-Gly shows the expected number of glycine residues (G10, G35, G53, G47, G75, G76) from the ubiquitin sequence. Similarly, unique resonances were clearly displayed for two phenylalanine (F4, F45) and one tyrosine residue (Y59) in ubiquitin in the $^{13}$C spectrum (Figure 22 C-D). In addition to the alanine residue in ubiquitin (A46), the alanine from SNA sequence was displayed in the $^{13}$C NMR spectrum of [$^{13}$C, $^{17}$O$_2$]-Gly-ubiquitin (Figure 22B).
When attempting to selectively label a specific residue type, two concerns must be taken into account. The first is the isotope dilution due to the endogenous host biosynthetic pathway. The second and the greater concern is the unwanted scrambling of the label to other types of residues due to isotope transfer of $^{15}$N via a variety of aminotransaminases found throughout the
biosynthetic pathway. It is the interplay between the pathways that poses a challenge for selective labeling. Some amino acids are substrates for one or more of the general aminotransferases involved in the biosynthetic pathways, while others are direct metabolic precursors of others. The key precursor of the amino acids this project is interested in is outlined in Table 2. In either case, scrambling of the label is a potential threat to the outcome of the experiment. In contrast, isotope dilution is easier to minimize as the amino acid biosynthetic pathway is regulated by negative feedback inhibition. It mitigates the concerns to some degree by keeping the biosynthetic pathways quiescent, but the ability to label a specific amino acid type in a protein is not necessarily simply a case of supplementing minimal media with high concentration of the \(^{15}\text{N}\)-labelled amino acid of interest along with the remaining nineteen unlabelled amino acid. To address these potential labeling pitfalls in a most reliable and efficient way, a collection of genetic tools are created to manipulate the amino acid metabolism in \textit{E.coli}. It is now well documented that a number of auxotrophic strains of \textit{E.coli} have been constructed with ideal genotypes for residue-specific, selective labeling of proteins with nearly any \(^{15}\text{N}\)-amino acid. These strains of \textit{E. coli} have been modified that each contains defects in genes coding for particular enzymes closely linked to the biosynthetic pathway. Hence, the isotopic dilution and incorporation of label into undesired sites can be avoided or prevented by repression of bacterial amino acid metabolism using defined media supplemented with essentially all amino acids and by use of bacterial hosts with lesions blocking the appropriate steps of amino acid synthesis and degradation.
Table 2. *Escherichia coli* Amino Acid Metabolism

<table>
<thead>
<tr>
<th>Residue Type</th>
<th>$^{17}$O-Amino Acid</th>
<th>Key precursor</th>
<th>Relevant host genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>L-Alanine</td>
<td>Glu, Val</td>
<td>avtA, aspC, ilvE, tyrB</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
<td>Ser</td>
<td>glyA</td>
</tr>
<tr>
<td>Gly, Ser</td>
<td>Glycine</td>
<td>3-Phosphoglycerate</td>
<td>serA</td>
</tr>
<tr>
<td>Phe</td>
<td>L-Phenylalanine</td>
<td>Chorismate</td>
<td>pheA; aspC, ilvE, tyrB&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Tyr</td>
<td>L-Tyrosine</td>
<td>Chorismate</td>
<td>aspC, tyrB&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*From Reference 116

In the case of glycine, loss or misincorporation of an isotopic label is a serious concern when growing bacteria with $^{15}$N-labelled glycine resulting from transaminase activity. Serine is the precursor to glycine (Figure 23) and is involved in the one-carbon metabolism of the cell. Interconversion of serine and glycine can be repressed or prevented with mutations on *glyA*, as outlined in Fig. 23A. The auxotrophic strain *E. coli* DL39 GlyA λDE3 (*aspC*, *tyrB*, *ilvE*, *glyA*, λDE3),<sup>116</sup> used for the selective $^{15}$N-labeling of glycine in ubiquitin contains a number of genetic lesions including mutations in the gene encoding serine hydroxymethyltransferase (*glyA*).<sup>12</sup> Specific labeling requires only a lesion in the *glyA* gene since it is the end of a pathway and not a precursor for other amino acids. This lesion appears to be leaky since exclusion of serine form the growth media still result in the undesired labeling on serine residues in the ubiquitin as seen in Figure 23A.<sup>13</sup> However, when serine is present at a large amount (3.10 g/L) in the media, negative feedback inhibition prevents this shuffle of the $^{15}$N isotope. Unlike glycine, labeling of phenylalanine and tyrosine requires a combination of genetic lesions to guard against dilution and scrambling of the isotope because of the overlapping of the specificities. As shown in Figure 23C
and D, gene products of tyrB and aspC convert phenylalanine and tyrosine to glutamate resulted in the scrambling of the isotopic label by transamination, and thus necessitates the mutations. Additionally, selective labeling with phenylalanine also requires genetic lesions in pheA to prevent the interconversion of phenylpyruvate back to chorismate as outlined in Figure 23C, along with mutation on ilvE. The DL39 GlyA λDE3 (aspC− tyrB− ilvE− glyA− λDE3) used was constructed with all the genetic lesion mentioned above to serve the purpose of selective labeling of multiple amino acids.

Unfortunately, there are no true alanine auxotrophs available. Alanine can be synthesized in several pathways in E. coli. One of the major routes for alanine biosynthesis is the transamination between valine and alanine by alanine-valine aminotransferase, the gene product of avtA gene (Figure 23B)\(^{120}\) It also appears to be the product of transamination of pyruvate and glutamate, which involves other general aminotransferase.\(^{19}\) The best strain to label alanine is with defects in all four general aminotransferases (ilvE, avtA, aspC, and tyrB). Enrichment of ubiquitin with \(^{15}\)N-alanine and \(^{13}\)C,\(^{17}\)O-alanine were performed in DL39 GlyA λDE3 (aspC− tyrB− ilvE− glyA− λDE3) and CT19 AvtA::Tn5 (aspC− tyrB− ilvE− glyA− λDE3), both resulted in no scrambling of isotopes. However, DL39 GlyA λDE3 (aspC− tyrB− ilvE− glyA− λDE3) appears to have higher a yield of \(^{15}\)N-alanine and \(^{13}\)C,\(^{17}\)O-alanine and thus was used for incorporation of the labelled alanine in this project.
Figure 23. Generalized presentation of (a) glycine, (b) alanine, (c) phenylalanine, and (d) tyrosine amino acid biosynthetic pathways in *Escherichia coli*. The amino acids specifically $^{15}$N-labelled are capitalized and are boxed and in bold. The defective genes are identified in italic and correspond to the following enzymes: $\text{glyA}$, serine hydroxymethyl transferase; $\text{pheA}$, chorismate mutase-prephenate dehydratase; $\text{tyrB}$, tyrosine aminotransferase; $\text{aspC}$, aspartate aminotransferase; $\text{ilvE}$, transaminase B; $\text{avtA}$, transaminase C. This is a modified version of a figure taken from Waugh, 1996.$^{120}$
3.3.6 Ubiquitin $^{17}$O NMR pH-Titration

Despite the large number of NMR studies performed on ubiquitin, there has not been any solution-state $^{17}$O NMR study on ubiquitin, nor has there been any investigation on how $^{17}$O NMR spectroscopy can be used to study electrostatic interactions among proteins. $^1$H, $^{13}$C, and $^{15}$N NMR has been used in acid-base chemistry as a tool to determine the pKa values. Considering that none of these NMR active nuclei is a quadrupole nucleus, more information on the electronic structure and binding of proteins can be learnt through $^{17}$O NMR. The next goal of this chapter is to explore the possibility of using $^{17}$O NMR spectroscopy for acid-base chemistry through the synthesized [$^{13}$C,$^{17}$O]-Gly-Ubiquitin and 4-[$^{17}$O]-L-Tyrosine-Ubiquitin. To our knowledge, this is the first discussion of the possibility that $^{17}$O NMR could be used to determine the pKa value, and to potentially study the electrostatic interactions of biomolecules.

Electrostatic interactions have a paramount importance in the structures, dynamics and functions of proteins.$^{121}$ These interactions are primarily established by the pH-dependent ionizable side chains and termini. The acid dissociation equilibrium constant (pKa values) and the charge states of ionizable residues in functionally important proteins provide more insights into the electrostatic properties of proteins and their complexes.$^{122}$ These are the essential characteristics to study in drug discovery and development for the prediction of pharmacokinetic and pharmacodynamic properties.$^{123}$ Various straightforward methods are available to determine the pKa values, this includes UV-spectroscopic, scanning potentiometric, and capillary electrophoretic techniques.$^{124}$ Determining the pKa values of large proteins can be challenging with these traditional techniques. NMR spectroscopy appears to be the most direct method to study the residue-specific charge states and pKa values of protein or protein complexes of all such experimental approaches at a molecular level. The chemical shifts of NMR-active nuclei (e.g.
$^{13}\text{C}, \, ^{1}\text{H}, \, ^{15}\text{N}$) altered with respect to the protonation state of the adjacent ionizable sites as a function of pH. Thus, NMR spectroscopy offers a site-specific method to monitor the chemical shift changes from non-exchangeable $^{1}\text{H}, \, ^{13}\text{C}$ or $^{15}\text{N}$ nuclei within a given residue with response to a pH titration for protein or protein complexes. The pK$_a$ values and hence the protonation states of the ionizable residues in proteins are determined by the inflection point of the sigmodal curve created by plotting the chemical shifts against different pH values fitted to Henderson–Hasselbalch type equation (Equation 8).

In solution-state $^{17}\text{O}$ NMR, the oxygen resonance for each $^{13}\text{C}, ^{17}\text{O}$-labeled amino acid in ubiquitin is not well-resolved due to the unfavorable properties of $^{17}\text{O}$. As mentioned in Chapter 1, the strong quadrupolar interaction tends to cause severe line broadening in the NMR spectra of quadrupolar nuclei, especially when it is coupled with molecular tumbling in solution-state. Despite the advancement and expansion of $^{17}\text{O}$ NMR spectroscopy in the last decade, sensitivity and resolution still limit the its routine application to biomolecules, where motion and sample heterogeneity complicate the interpretation of data quantitively. One way of reducing the quadrupolar line broadening is to perform solid-state experiments. The last goal of this chapter is to obtain high-resolution solid-state NMR spectra of the residue-specifically $^{17}\text{O}$-labeled ubiquitin.

It is known that sample preparation has a profound effect on the line widths in solid-state NMR spectra of proteins and peptides. NMR lines of lyophilized proteins are usually inhomogeneously broadened. Addition of cryoprotectants or hydration of the freeze-dried samples can improve line widths of lyophilized proteins. An increasing body of NMR studies is finding that crystalline forms of proteins have substantive advantages over amorphous lyophilized powders for structural studies using solid state NMR. Having some degree of microscopic order are shown to give narrow NMR peaks in $^{13}\text{C}$ cross-polarization magic angle spinning (CP MAS)
spectra of natural abundance ubiquitin. In last part of this chapter, we present solid-state $^{17}$O MAS NMR spectra and $^{13}$C CP/MAS NMR spectra residual-specifically labelled microcrystalline yeast ubiquitin.

The charge states of the $\alpha$-carboxyl group of $[^{13}$C, $^{17}$O]-glycine (Figure 24a) and phenolic group of 4-$^{17}$O]-L-tyrosine (Figure 24b) in the residue-specifically labelled ubiquitin were monitored using the chemical shifts of the $^{17}$O nucleus as a reporter in solution. The chemical shifts of the oxygen-17 species within a range of pH values in the two titration experiments are summarized in in Table 3. The pH-dependent chemical shifts were fitted versus pH data to modified Henderson-Hasselbalch model (Eq7) for a chemical equilibrium in the fast exchange limits as shown in Figure 25C. The $pK_a$ values were determined for the two ionizable groups of interest that experienced complete deprotonation (Figure 25) over the pH range sampled.
Figure 24. Deprotonation of the phenol group of 4-[¹⁷O]-L-tyrosine (a) and the α-carboxyl group of [¹³C,¹⁷O₂]-glycine labelled ubiquitin (b).¹²⁶

The resonances observed in the solution ¹⁷O NMR spectrum the [¹³C,¹⁷O]-glycine-ubiquitin (Figure 24 b) was assigned to be due to the terminal G76 residue. Other five glycine residues (G10, G35, G53, G47, G75) were not observable due to quadrupolar line broadening for the backbone amide groups. It can be seen that the ¹⁷O signal from the G76 shifted downfield by ~11 pm upon deprotonation from pH = 1.86 to pH = 3.11 at 25 °C. The ¹⁷O NMR pH-titration curve of these glycine residues have a sigmoidal shape (Figure 25), typical for a single titrating site, and were fitted to the modified Henderson-Hasselbalch equation. The protonated ¹⁷O chemical shift and deprotonated ¹⁷O chemical shift of the backbone of [¹³C,¹⁷O]-Gly in ubiquitin were obtained from the pH = 1.86 and pH = 3.11 samples, respectively (Table 3). The fitted curve shows good agreement with the experimental data when the pKₐ value was simultaneous varied.
during the fitting. The pKa value of 2.32 extracted from the $^{17}$O NMR pH-titration curve correlates well with the pKa value of the $\alpha$-carboxyl group of glycine determined by electro-potentiometric method for free amino acids (pKa = 2.34).\textsuperscript{126}

Table 3. pH-Dependent $^{17}$O chemical shifts (ppm) of the $\alpha$-carboxyl group in $[^{13}C,^{17}O]$-Gly-Ub and phenolic group in $[^{Phenol-^{17}O}]-Tyr-Ub$

<table>
<thead>
<tr>
<th>pH</th>
<th>$\delta_{\text{Observed (ppm)}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^{13}C,^{17}O]$-Gly-Ub</td>
<td>258.76</td>
</tr>
<tr>
<td>1.861</td>
<td></td>
</tr>
<tr>
<td>2.494</td>
<td>264.53</td>
</tr>
<tr>
<td>3.112</td>
<td>270.30</td>
</tr>
<tr>
<td>4.213</td>
<td>271.63</td>
</tr>
<tr>
<td>$[^{Phenol-^{17}O}]$-Tyr-Ub</td>
<td>77.15</td>
</tr>
<tr>
<td>7.07</td>
<td></td>
</tr>
<tr>
<td>7.75</td>
<td>79.42</td>
</tr>
<tr>
<td>8.75</td>
<td>84.25</td>
</tr>
<tr>
<td>9.75</td>
<td>96.54</td>
</tr>
<tr>
<td>10.75</td>
<td>125.36</td>
</tr>
<tr>
<td>11.75</td>
<td>151.27</td>
</tr>
<tr>
<td>12.75</td>
<td>157.11</td>
</tr>
</tbody>
</table>
Figure 25. $^{17}$O NMR spectra of for (a) 4-$^{17}$O-L-tyrosine labelled and (b) $[^{13}$C, $^{17}$O]-glycine labelled Ubiquitin in 25 mM Tris, 250 mM NaCl, as a function of pH. In (a), the small peak marked by * was due to the presence of $\text{SO}_4^{2-}$ ions.

The $^{17}$O chemical shifts of the sidechain phenolic group of Y59 from the [Phenol-$^{17}$O]-Tyr-Ub were measured between pH = 7.07 to pH = 12.75. It can be seen that the $^{17}$O chemical shift was shifted downfield overall as the chemical environment becomes more basic. The $^{17}$O
chemical shift difference between protonated and deprotonated states is about 80 ppm. The blue sigmoidal curve in Figure 25 shows the side-chain chemical shifts versus pH together with the fitting from a Henderson-Hasselbalch equation (Eq7) and correlates well with the experimental data (in blue dots). The pKa value of the phenolic oxygen-17 of Tyr59 in ubiquitin was extracted from the simulated titration curve and was determined to be 10.44 ± 0.04.

![Figure 25](image)

**Figure 26.** Experimental (in dots) and simulated (in lines) $^{17}$O NMR pH-Titration curves for 4-$^{17}$O-L-tyrosine labelled Ubiquitin (Blue) and $[^{13}$C,$^{17}$O$_2$]-Glycine labelled Ubiquitin (Orange). Lines are best non-linear least squares fits to a Henderson–Hasselbalch type equation and pKa values of 10.31 and 2.32, respectively. The parameters used to obtain the titration curves are summarized on the right. $\Delta_{obs}$ represents the observed chemical shift at each pH; $\delta_p$ represents the chemical shift of the fully protonated state with respect to the titration, $\Delta\delta_1$ is the chemical shift change associated with the titration governed by pK$_a$

<table>
<thead>
<tr>
<th>Phenol-17O-Tyr-Ub</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_p$ (ppm)</td>
<td>80.0638</td>
</tr>
<tr>
<td>$\Delta\delta_1$ (ppm)</td>
<td>74.5232</td>
</tr>
<tr>
<td>pK$_a$</td>
<td>10.4404</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$[^{13}$C,$^{17}$O]-Gly-UB</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_p$ (ppm)</td>
<td>254.0237</td>
</tr>
<tr>
<td>$\Delta\delta_1$ (ppm)</td>
<td>18.1950</td>
</tr>
<tr>
<td>pK$_a$</td>
<td>2.3237</td>
</tr>
</tbody>
</table>

3.3.7 Solid-State NMR Spectra of Residue-Specifically Labelled Ubiquitin

1D $^{17}$O MAS NMR spectra of lyophilized His-TEV-ubiquitin at 16.4 T with glycine residues enriched with $^{13}$C and $^{17}$O are shown in Figure 27, along with that of His-TEV-ubiquitin where phenylalanine residues were enriched with $^{13}$C and $^{17}$O. By increasing the spinning speed, we managed to observe the quadrupole central transition of the backbone $^{17}$O resonances of the glycine and phenylalanine residues in $[^{13}$C,$^{17}$O]-Gly-Ubiquitin and $[^{13}$C,$^{17}$O]-Phe-Ubiquitin,
respectively, by the magic angle spinning (MAS) method. This immediately confirmed that the \(^{17}\text{O}\)-labeling of ubiquitin was successful. Unfortunately, the peaks are not well-resolved. We are not able to observe and identify each backbone oxygen from each glycine or phenylalanine residue in the residue-specifically labeled ubiquitin. The line broadening of central transition is primarily due to the second-order quadrupolar interaction, which is inversely proportional to the magnetic field. Thus, it can potentially be further reduced by performing \(^{17}\text{O}\) MAS NMR experiments at higher fields in the future.

\[ \text{[^{13}\text{C},^{17}\text{O}] - Gly His-TEV-ubiquitin} \]

\[ \text{[^{13}\text{C},^{17}\text{O}] - Phe His-TEV-ubiquitin} \]

- **17\text{O} MAS NMR spectra of His-TEV-ubiquitin where the (a) glycine and (b) phenylalanine residues are doubly labelled by \(^{13}\text{C}\) and \(^{17}\text{O}\). The spectra were obtained at 16.4 T with a sample spinning frequency of 30 kHz, at 263 K. The total experiment time was 24 hrs.**

Another important factor that is known to be important in protein NMR is sample preparation. Lyophilized powder sample, the most common solid form of most proteins, does not provide sufficient chemical shift resolution for large conformationally rich molecular systems. This is because the fact that chemical shifts are sensitive to molecular conformation. Although the freeze-dried proteins might be properly folded, the conformation of side chains could be variable.
Each chemically equivalent site is rendered magnetically inequivalent, thus giving rise to a slightly different chemical shift of the same molecule. Therefore, the linewidths for lyophilized proteins are typically not narrow enough to assign single sites by solid-state NMR spectroscopy. The structural homogeneity generally results in $^{13}$C linewidths of 1-2 ppm for samples that appear to be flaky. Significant efforts have been made on improving the resolution of ssNMR spectra for freeze-dried protein samples by adding hydration stabilizing additives prior to lyophilization, such as polyethylene glycol (PEG). For most proteins, the native structures are well-retained, and the retention activities are preserved by keeping protein hydrated using these protectants. The ssNMR spectra have higher resolution than spectra obtained without the cryoprotectants. However, the spectral resolution is still not high enough to assign high percentage of resolvable single sites. There have been some studies demonstrating that microcrystalline form of proteins can do a better job of providing high-resolution ssNMR spectra. Single site resolution has been compared for SH3 domain from $\alpha$-spectrin prepared by lyophilization with and without hydration with PEG, and precipitation with ammonium sulfate. The hydrated microcrystalline precipitate from “salting-out” method gave the best $^{13}$C spectral resolution, and the linewidths are comparable to small crystalline model compound.

In the last part of the project, we initially intended to try to obtain microcrystalline residue-specifically labeled ubiquitin by the batch crystallization method. We tried 60% (v/v) 2-methyl-2,4-pentanediol (MPD) in acidic condition (pH 4.02) using sodium citrate as the buffer as the precipitant. This protocol was adapted from published conditions using MPD and 20 mM citrate buffer at pH 4.0-4.2. However, we did not manage to achieve consistent yields using the same condition, and the time required for each preparation various from two days up to three weeks. It is a well-known fact that crystallization of macromolecules is a form of biochemical high art that
is difficult and unpredictable. Therefore, we have tried the alternative “sating-out” method by ammonium sulfate. Microcrystalline precipitates of \([^{13}\text{C},^{17}\text{O}_2]\)-Phe-Ubiquitin were produced readily when ammonium sulfate was 20% saturated completed at 65% of saturation. This approach has been applied to different proteins that have been previously characterized by single crystal X-ray diffraction for achieving high resolution ssNMR spectra. In the case of ubiquitin, it can be performed to obtain nanocrystals of limited protein as crystalline solid.

The \(^{13}\text{C}\) CP/MAS NMR spectra of the microcrystalline \([^{13}\text{C},^{17}\text{O}_2]\)-Phe-Ubiquitin precipitate and lyophilized powder hydrated with glycerol were compared in Figure 28. They both have two spinning side bands on each side of the \(^{13}\text{C}\) resonance of the backbone carbon of the\([^{13}\text{C},^{17}\text{O}_2]\)-Phe. While the narrow features in the ssNMR \(^{13}\text{C}\) spectra indicates that the sample is crystalline, the two \(^{13}\text{C},^{17}\text{O}\) doubly labeled phenylalanine residues at position 4 and 45 were identified in the microcrystalline form. This is in stark contrast to the hydrated lyophilized protein, where one peak with linewidths greater than 1 ppm is observed at around ~ 170 ppm. Two distinguishable well-resolved peaks with narrower linewidth representing two single carbon sites are present in the ssNMR \(^{13}\text{C}\) spectra of the microcrystalline ubiquitin. The observation at this degree of resolution demonstrates that the application of solid state \(^{17}\text{O}\) NMR methods to macromolecules is amenable by producing samples with crystalline local order.
Figure 28. $^{13}$C CP/MAS NMR spectra of yeast ubiquitin where F4 and F45 residues are $[^{13}\text{C},^{17}\text{O}]$-doubly labelled. The spectra were obtained at 16.4 T with a sample spinning frequency of 8 kHz.

3.4 Conclusion

Over the course of this investigation, four $^{13}$C,$^{17}$O-labelled amino acids (glycine, alanine, phenylalanine, and tyrosine) were successfully synthesized. One-bond isotope shifts from the oxygen isotopes were observed in the $^{13}$C NMR spectrum of glycine. Using a positive ESI-MS technique, the $^{13}$C,$^{17}$O-labelled amino acids were determined to be enriched with $\sim 20\%$ $^{17}$O, which is sufficient to be used for isotopically labeling of yeast ubiquitin.

The second part of this chapter demonstrates successful incorporation of the residue-specially $^{13}$C,$^{17}$O-labelled amino acids into ubiquitin. We have obtained consistent high yields of pure yeast ubiquitin with either backbone or sidechain $^{17}$O-labelled amino acids from the overexpression, isotopic labeling and purification protocols using an auxotrophic strain and minimal media. This method can become a general strategy for $^{17}$O labeling specific amino acid
types in proteins. This is the first time that syntheses of $^{17}$O-labelled proteins are reported. It provides the necessary material to carry out other $^{17}$O NMR experiments that will highlight the utility of this isotopic labeling protocol for evaluating the oxygen atoms of biological molecules by NMR spectroscopy.

The $[^{13}$C,$^{17}$O$_2$]-Gly-ubiquitin and [phenol-$^{17}$O]-Tyr-ubiquitin were then used to navigate the possibility of using solution $^{17}$O NMR as a probe to study electrostatic interactions within or between biological molecules. The pKa values of the carboxylic acid group of the terminal glycine residue and side-chain phenolic group of tyrosine (Y59), were determined from the sigmodal $^{17}$O NMR-titration curves, correlated well with the pKa values reported in the literature that are determined by other methods. $^{17}$O NMR studies as described herein will undoubtedly prove quite informative in this regard.

Investigation of large molecules by ssNMR methods are always impeded by difficulties in obtaining high resolution NMR spectra, especially for a quadrupolar nucleus. The result presented in part of this chapter demonstrate that nanocrystalline material prepared by precipitation using ammonium sulfate is suitable for structural or other biophysical studies by ssNMR. Single carbon sites are clearly seen in the microcrystalline $[^{13}$C,$^{17}$O$_2$]-Phe-Ubiquitin using $^{13}$C CP/MAS ssNMR spectra with good sensitivity. It is clear that nanocrystalline $[^{13}$C,$^{17}$O$_2$]-Phe-Ubiquitin provide higher resolution of $^{13}$C ssNMR spectra compared to the spectra of the conventional lyophilized powder.
Chapter 4 Conclusions and Future Work

Knowledge of protein structures is a significant factor for understanding protein functions. Protein functions are also related to protein dynamics, ranging from picoseconds to over seconds. Unlike X-ray crystallography which cannot examine samples in solution and requires protein crystals, NMR spectroscopy is the most powerful technique to determine molecular structure at atomic level in solution. However, X-ray crystallography and solution NMR are not useful in studies of biological systems which are insoluble or are resistant to crystallization, such as amyloid aggregated, membrane proteins, and large protein complexes. Recently, Solid-State NMR spectroscopy becomes a promising tool for structural, dynamical, and functional characterization of proteins. Especially, rapid methodological development of Magic Angle Spinning method allows us to progress from small peptides to large protein complexes as it utilizes isotropic and anisotropic interactions during measurements.\textsuperscript{125} The inherently quantitative nature of NMR spectroscopy is an invaluable property for its use in structure determination, particularly for materials that lack the long-range order required by other techniques. By applying the isotopic labeling schemes to proteins, a strong reduction of spectral overlapping and increase in spectral resolution in ssNMR spectra can be achieved. The correlation of isotropic backbone and sidechain $^{13}$C and $^{15}$N resonances using multidimensional solid-state NMR spectra has facilitated the chemical shifts assignments of all the $^1$H, $^{13}$C, $^{15}$N nuclei in proteins. To reach the goal of high spectral resolution with high-throughput productivity, novel stable isotopic labeling methods and techniques are continuously being developed, this include the development of $^{17}$O labeling methods and $^{17}$O NMR spectroscopy. Direct observation of hydrogen bonding through $^{17}$O isotopes and measurement of quadrupolar interaction may potentially provide new insights into protein structures, dynamics and functions. But $^{17}$O NMR of proteins remains to be a challenge.
In this thesis, yeast ubiquitin was selected as a test system for the purpose of optimizing $^{17}$O labeling protocols and for testing solution and solid-state $^{17}$O NMR spectroscopy. Ubiquitin is a relatively small protein composed only of 76 amino acids with good thermal stability. Extensive information about the structural chemical shift assignment and dynamics of ubiquitin is available from X-ray crystallography and solution NMR studies.

We have used this protein to illustrate that high-resolution spectra at high magnetic field could be attained and to explore the optimized condition for high quality solid-state $^{17}$O NMR of residue-specifically $^{17}$O-enriched materials. We have successfully synthesized $^{13}$C,$^{17}$O doubly labelled glycine, alanine, phenylalanine, and tyrosine (including sidechain $^{17}$O-labelled tyrosine), and incorporated them into yeast ubiquitin, without isotopic scrambling. We then applied $^{17}$O-labelled ubiquitin for solution and solid-state $^{17}$O NMR experiments. $^{17}$O NMR spectroscopy allows a fast, sensitive and reliable determination of pK$_a$ values over a broad pH-range and can be used for compounds with more than one ionizable center. The results from pK$_a$ determination of $\alpha$-carboxyl group of $[^{13}$C, $^{17}$O]-glycine and phenolic group of 4-$[^{17}$O]-L-tyrosine in the residue-specifically labelled ubiquitin by solution $^{17}$O NMR spectroscopy correlate well with data obtained with the commonly used potentiometric and UV-spectroscopic methods. In the future studies, protein binding capacitances could be analyzed and provides more insights on origins of the pKa values obtained through $^{17}$O NMR, and thus electrostatic interactions can also be studied suggested by our findings. Furthermore, it is applicable to compounds with low solubility, and only microgram amounts of an analyte are required. In conclusion, successful extraction of pKa values from $^{17}$O NMR pH-titration curves opens the door for $^{17}$O NMR spectroscopy to investigate protein dynamics and interactions.
We have also obtained preliminary solid-state $^{17}$O NMR spectra for the residue-specifically $^{17}$O-labelled ubiquitin. High-quality solid-state $^{17}$O NMR spectra confirmed the $^{17}$O-labeling of yeast ubiquitin. However, because $^{17}$O is quadrupolar, solid-state $^{17}$O NMR spectra display severe line broadening. It is known that samples preparation has a profound effect on the line widths in solid-state NMR spectra of proteins and peptides. NMR signals from lyophilized proteins are usually inhomogeneous broaden. Addition of cryoprotectants or hydration of the freeze-dried samples can improve line widths of proteins. Microcrystalline proteins have proven to provide even better spectral resolution in terms of line width in many studies. Most of the solid-state NMR works of microcrystalline ubiquitin samples used high concentration of 2-methyl-2,4-pentanediol (MPD) as precipitant at pH ~ 4.0-4.2 while the crystallization conditions available in the PDB use polyethylene glycol 400 (PEG 400) as precipitant at ~pH 5.6. Initially, we have obtained the microcrystals of $^{[13}$C,$^{17}$O]-Gly-Ubiquitin using 60% (v/v) MPD as the precipitant under the acidic condition. The microcrystals of isotopically labelled ubiquitin were formed in batch after 48 hours at 4 °C. However, the yield is not reproducible, and conditions applied vary for each trial. Thus, efforts need to be made in the future to propagate the factors causing to this problem, so that microcrystalline $^{17}$O-labelled ubiquitin could be routinely prepared by the batch method for ssNMR $^{17}$O studies. It is anticipated that resolution in solid-state $^{17}$O NMR spectra of microcrystalline proteins will improve.

On the other hand, we have managed to obtain bulk samples of $^{[13}$C, $^{17}$O$_2$]-Phe-Ubiquitin nanocrystals rapidly by “salting-out” method using ammonium sulfate which are long-term stable. As evaluated by $^{13}$C CP/MAS ssNMR spectra, nanocrystalline protein prepared in this fashion has higher degree of order and provides higher resolution $^{13}$C CP/MAS ssNMR spectra. The method developed in this thesis is rapid and should find use in sample preparation for structural studies.
using ssNMR. With fast and efficient crystallization conditions and well-characterized chemical shifts, ubiquitin is no doubt a good system for developing new $^{17}$O NMR methods for structural and dynamics studies of biological molecules.

The results reported in this thesis represent the first set of data in the emerging field of $^{17}$O NMR for proteins. Further development is needed in both $^{17}$O- labeling of proteins and new $^{17}$O NMR techniques especially at ultra-high magnetic fields. Solid-state $^{17}$O NMR will become a useful tool for studying proteins dynamics on a routine basis in the very near future.
References

(10) Denis, M. Biochimie 1986, 68, 459.
(15) Marion, D. Molecular & Cellular Proteomics 2013, 12, 3006.
(22) McIntosh, L. P.; Griffey, R. H.; Muchmore, D. C.; Nielson, C. P.; Redfield, A. G.; Dahlquist, F. W. Proceedings of the National Academy of Sciences of the United States of America 1987, 84, 1244.
(64) Zhu, J.; Geris, A.; Wu, G. *Physical Chemistry Chemical Physics* 2009, 11, 6972.
(69) Zhu, J.; Wu, G. *Journal of the American Chemical Society* 2011, 133, 920.
(102) Verardi, R.; Traaseth, N. J.; Masterson, L. R.; Vostrikov, V. V.; Veglia, G. Advances in Experimental Medicine and Biology 2012, 992, 35.
(110) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Journal of Molecular Biology 1987, 194, 531.
(113) LeMaster, D. M.; Richards, F. M. Biochemistry 1988, 27, 142.


(121) Bondar, A.-N.; White, S. H. In *Biochimica Biophysica Acta*; Elsevier 2012, 18, 942.

(122) Zhou, H.-X.; Pang, X. *Chemical reviews* 2018, 118, 1691.


