Characterizing labile protons by NMR spectroscopy

by

Jacob Arthur Brockerman

B.A., Western Washington University, 2010
M.S., Western Washington University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in
THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

February 2019

©Jacob Arthur Brockerman, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

**Characterizing labile protons by NMR spectroscopy**

submitted by **Jacob Arthur Brockerman** in partial fulfillment of the requirements for

the degree of **Doctor of Philosophy**

in **Biochemistry and Molecular Biology**

**Examinining Committee:**

*Lawrence P. McIntosh*  
Supervisor

*Filip L.A. Van Petegem*  
Supervisory Committee Member

*Stephen G. Withers*  
Supervisory Committee Member

*Suzana Straus*  
University Examiner

*Carl Michal*  
University Examiner

**Additional Supervisory Committee Members:**

Supervisory Committee Member

Supervisory Committee Member
Abstract

Labile hydrogens in proteins, including those on ionizable functional groups, undergo rapid exchange with water and thus are typically difficult to characterize by current techniques in structural biology. Their properties are often inferred from biophysical arguments rather than direct experimental determination. Here I present new methods for studying these hydrogens by NMR spectroscopy using two glycoside hydrolases as model protein systems.

Chapter 2 centres on characterizing the hydroxyl protons of serine and threonine residues in proteins by NMR spectroscopy. Using auxotrophic E. coli strains, I produced Bacillus circulans xylanase (BcX) selectively labeled with $^{13}\text{C}/^{15}\text{N}$-serine or $^{13}\text{C}/^{15}\text{N}$-threonine. Signals from two serine and three threonine hydroxyls in these protein samples were readily observed using long-range heteronuclear scalar correlation experiments. Their dihedral angle-dependent scalar couplings with adjacent sidechain protons were determined via a quantitative $^{13}\text{C}/^{15}\text{N}$-filtered spin-echo difference experiment. The hydrogen exchange kinetics of these hydroxyls were measured using a $^{13}\text{C}/^{15}\text{N}$-filtered CLEANEX-PM pulse sequence. Collectively, these experiments provided insights into the structural and dynamic properties of this model protein.

Chapter 3 characterizes a mutant of T4 phage lysozyme (T4L) with an altered catalytic mechanism due to the substitution of residue Thr26 with a histidine (T26H). It has been proposed that T26H-T4L hydrolyzes peptidoglycan via a double displacement mechanism with His26 serving the unusual role of a nucleophile. To gain further insights into this or alternative mechanisms, I used NMR spectroscopy to measure the acid dissociation constants ($pK_a$ values) and/or ionization states of all the Asp, Glu, His, and Arg residues in the T4L mutant. Most notably, the $pK_a$ value of the proposed nucleophile His26 is $6.8 \pm 0.1$, whereas that of the general
acid Glu11 is 4.7 ± 0.1. If the proposed mechanism holds true, then T26H-T4L follows a reverse protonation pathway where only a minor population of protein is in its catalytically competent ionization state (His26 deprotonated and Glu11 protonated). I also demonstrated that all arginines in T26H-T4L, including the active site Arg145, are positively charged under neutral pH. This stands in contrast to a recent neutron crystallographic study of T26H-T4L in which, perplexingly, Arg145 was proposed to have a deprotonated guanidine sidechain.
Lay Summary

Polar amino acids are important for both protein structure and function. However, some hydrogens in the sidechains of these amino acids typically undergo rapid exchange with solvent water. Such “labile" hydrogens often serve important roles in establishing the structure of a protein, as well as the catalytic mechanism of an enzyme. Because of solvent exchange, these hydrogens are challenging to study by many standard experimental techniques. My thesis focuses on overcoming this challenge using NMR spectroscopy. In one chapter, I present new methods to characterize the hydroxyls of serines and threonines. The utility of these methods are demonstrated through insights into the structure of a model xylanase protein. In a second chapter, I discuss how I studied the labile hydrogens from amino acids in the active site of an enzyme, called T4 lysozyme, in order to better understand the mechanism by which it cleaves the cell walls of bacteria.
Preface


For the biochemistry publication, all experiments regarding enzyme kinetics and crystallization were carried out by Ludwiczek, D’Angelo and Yalloway in the labs of Dr. Withers and Dr. Strynadka. Ludwiczek and I carried out NMR experiments in the lab of Dr. McIntosh with the assistance of Dr. Okon.

For the protein science publication, Bergeron, Vuckovic and Deng performed the computational modeling and secretion assays in the labs of Dr. Strynadka and Dr. Finlay. Bergeron and I carried out NMR experiments in the lab of Dr. McIntosh with the assistance of Dr. Okon.

Chapter 2 is a reformatted version of the paper: Brockerman, J.A., Okon, M., and McIntosh, L.P. “Detection and characterization of serine and threonine hydroxyl protons in *Bacillus circulans* xylanase by NMR spectroscopy” *J. Biomol. NMR* 58: 17-25 (2014). I performed all of the experiments and data analysis with NMR spectroscopy assistance by Dr. Okon. Dr. McIntosh and I wrote and edited the paper.
Chapter 3 is a reformatted version of the paper: Brockerman, J.A., Okon, M., Withers, S.G., and McIntosh, L.P. “Measuring the pK_a values of the catalytic residues in the retaining glycoside hydrolase T26H mutant of T4 lysozyme” Protein Sci. 28(3) 620-632 (2019). I performed all of the experiments and data analysis with NMR spectroscopy assistance by Dr. Okon. Dr. McIntosh and I wrote and edited the paper.
Table of Contents

Abstract........................................................................................................................................ iii
Lay Summary .................................................................................................................................. v
Preface ............................................................................................................................................ vi
Table of Contents ........................................................................................................................... viii
List of Tables .................................................................................................................................. xi
List of Figures ................................................................................................................................. xii
List of Abbreviations ..................................................................................................................... xiv
List of Amino Acid Abbreviations ............................................................................................... xvi
Acknowledgements ....................................................................................................................... xvii

Chapter 1: Introduction ..................................................................................................................... 1
1.1  Protein structure function and dynamics................................................................................ 1
1.2  Structural and functional roles of polar amino acids ............................................................. 1
1.3  Characterizing labile protons by NMR spectroscopy: hydrogen exchange ....................... 4
1.4  Characterizing labile protons by NMR spectroscopy: protonation states and acid

dissociation constants...................................................................................................................... 8
1.5  Characterizing pH-dependent conformational equilibria of proteins by NMR: a case

study of PrgK ................................................................................................................................. 9
1.6  Linking residue specific pKₐ values to catalysis: a case study of BcX ............................... 16
   1.6.1  Inverting mechanism of glycoside hydrolases................................................................. 16
   1.6.2  Retaining mechanism of glycoside hydrolases ............................................................... 17
   1.6.3  The role of protonation states in general acid/base and nucleophilic catalysis .......... 19
Chapter 2: Detection and characterization of serine and threonine hydroxyl protons in *Bacillus circulans* xylanase by NMR spectroscopy

2.1 Overview

2.2 Introduction

2.2 Materials and Methods

2.2.1 Protein expression and purification

2.2.2 Protein characterization

2.2.3 NMR spectroscopy

2.2.4 \(^3\)J\(_{\text{CH}}\) quantitation

2.2.5 Hydrogen exchange kinetics

2.3 Results and Discussion

2.3.1 Detection of serine and threonine hydroxyl signals

2.3.2 Coupling constant measurements

2.3.3 Hydrogen exchange kinetics

2.3.4 Structural insights

2.4 Summary

Chapter 3: The \(pK_a\) values of the catalytic residues in the retaining glycoside hydrolase

T26H mutant of T4 lysozyme
3.3 Materials and Methods........................................................................................................66
  3.3.1 Protein expression and purification ..............................................................................66
  3.3.2 NMR Spectroscopy ........................................................................................................67
  3.3.3 pK_a Determination of His, Asp, and Glu ................................................................. 68
3.4 Results..................................................................................................................................69
  3.4.1 Histidine pK_a and tautomer state determination .........................................................72
  3.4.2 Aspartic acid pK_a determination ..................................................................................76
  3.4.3 Glutamic acid pK_a determination ...............................................................................81
  3.4.4 All arginines are protonated at neutral pH ................................................................. 83
3.5 Discussion..............................................................................................................................89
  3.5.1 Inverting mechanism of wild-type T4L .................................................................... 89
  3.5.2 Retaining mechanism of T26H-T4L with His26 as a nucleophile .............................. 90
  3.5.3 Alternative mechanism for T26H-T4L ....................................................................91
  3.5.4 Arg145 is protonated with a high pKa value .............................................................. 93
  3.5.5 Ionization states determined from neutron crystallography ....................................93

Chapter 4: Conclusions and future studies .............................................................................95
  4.1 Characterizing hydroxyl protons by NMR spectroscopy: Summary, significance,
    limitations, and potential applications ..................................................................................95
  4.2 Characterizing T4L* T26H: Summary, significance, limitations, and potential
    applications ......................................................................................................................... 97

Bibliography .............................................................................................................................100

Appendices ...............................................................................................................................111

Appendix A Pulse sequences for the detection and characterization of hydroxyls .......... 111
List of Tables

Table 3.1 $\text{pK}_a$ values of T26H-T4L* ................................................................. 82
List of Figures

Figure 1.1 Polar amino acid sidechains ................................................................. 3
Figure 1.2 Approximate exchange rates for labile groups ...................................... 5
Figure 1.3 Analysis of D1 of PrgK ........................................................................ 10
Figure 1.4 Proline isomerization determined by $^{13}$C$^\beta$ and $^{13}$C$^\gamma$ chemical shifts .................. 12
Figure 1.5 A salt bridge between His42 and Tyr75 is only seen in conformer A ................ 13
Figure 1.6 Conformer specific pKas of His42 .......................................................... 14
Figure 1.7 Model of the PrgK pH dependent conformational equilibria .................... 15
Figure 1.8 Mechanisms of glycoside hydrolases ...................................................... 18
Figure 1.9 The catalytic site of BcX ....................................................................... 21
Figure 1.10 pH activity profile of BcX E172H ............................................................ 23
Figure 1.11 NMR monitored pH titrations of His in BcX E172H ................................. 27
Figure 1.12 Catalytically competent populations of general acid / base enzymes .......... 30
Figure 1.13 pH activity profile of BcX N35D ............................................................. 33
Figure 2.1 The $^1$H-NMR spectrum of uniformly $^{13}$C/$^{15}$N-labeled BcX ................. 38
Figure 2.2 NMR analysis of Auxotrophic E. coli JW0003-2 of selective Thr labeling ....... 41
Figure 2.3 NMR analysis of Auxotrophic E. coli JC158 of selective Ser labeling .......... 42
Figure 2.4 Comparison of auxotrophic and prototrophic metabolic interconversion ....... 43
Figure 2.5 Assignment strategy for isotopically labeled Thr & Ser ............................. 45
Figure 2.6 Detection and assignment of Thr hydroxyl protons ................................... 48
Figure 2.7 Detection and assignment of Ser hydroxyl protons ................................... 49
Figure 2.8 Quantitative spin-echo difference experiment ......................................... 52
Figure 2.9 Hydroxyl HX rates via $^{15}$N/$^{13}$C-filtered CLEANEX-PM ....................... 57
Figure 2.10 Environment of detected hydroxyls

Figure 3.1 Proposed T4L mechanisms

Figure 3.2 $^{15}$N HSQC assignment of T4L* T26H

Figure 3.3 Comparison of T4L* and T4L* T26H

Figure 3.4 NMR monitored pH titrations and pKa fit of T4L* T26H His

Figure 3.5 NMR monitored pH titrations of T4L* T26H His

Figure 3.6 NMR monitored pH titrations and pKa fit of T4L* T26H Asp and Glu

Figure 3.7 Asp and Glu protonation state determination by H/D exchange

Figure 3.8 Assignment strategy for Arg guanindo groups

Figure 3.9 Protonation state determination of Arg by pH monitored titrations

Figure A.1 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered experiment

Figure A.2 Pulse sequence for a 2D $^{13}$C-HSQC experiment

Figure A.3 Pulse sequence for a 2D long-range constant-time methyl $^{13}$C-HSQC experiment

Figure A.4 Pulse sequence for a 2D methyl-selective $^{13}$C$^\beta$-decoupled $^{13}$C-HSQC experiment

Figure A.5 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered quantitative spin-echo difference experiment to measure $^{13}$C couplings to hydroxyl protons

Figure A.6 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered CLEANEX-PM to measure the hydrogen exchange kinetics

Figure A.7 Pulse sequence for a 3D intra-residue C(C)TOCSY_NH

Figure A.8 Pulse sequence for a 3D intra-residue H(CC)TOCSY-NH
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>one-dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>ASA</td>
<td>accessible surface area</td>
</tr>
<tr>
<td>BcX</td>
<td>xylanase from <em>Bacillus circulans</em></td>
</tr>
<tr>
<td>CLEANEX</td>
<td>clean chemical exchange</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multiple quantum correlation</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum correlation</td>
</tr>
<tr>
<td>INEPT</td>
<td>insensitive nuclei enhanced by polarization transfer</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>$K_a$</td>
<td>acid dissociation equilibrium constant</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>enzyme kinetic rate constant or turnover number</td>
</tr>
<tr>
<td>kD</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis constant</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cutoff</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>ONPX2</td>
<td>2'-nitrophenyl β-xylobioside</td>
</tr>
<tr>
<td>OD_{600}</td>
<td>Optical density at wavelength 600 nanometers</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank (<a href="http://www.pdb.org">http://www.pdb.org</a>)</td>
</tr>
<tr>
<td>pH*</td>
<td>pH meter reading uncorrected for isotope effects</td>
</tr>
<tr>
<td>pK_a</td>
<td>-log(K_a)</td>
</tr>
<tr>
<td>PM</td>
<td>phase modulated</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>RBM</td>
<td>ring building motif</td>
</tr>
<tr>
<td>rmsd</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>T_1</td>
<td>longitudinal relaxation time</td>
</tr>
<tr>
<td>T3SS</td>
<td>type 3 secretion system</td>
</tr>
<tr>
<td>T4L</td>
<td>lysozyme from Enterobacteria phage T4</td>
</tr>
<tr>
<td>T4L*</td>
<td>cysteine free mutant of T4L</td>
</tr>
<tr>
<td>TOCSY</td>
<td>total correlation spectroscopy</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
### List of Amino Acid Abbreviations

<table>
<thead>
<tr>
<th>Letter</th>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ala</td>
<td>alanine</td>
</tr>
<tr>
<td>R</td>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>D</td>
<td>Asp</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>N</td>
<td>Asn</td>
<td>asparagine</td>
</tr>
<tr>
<td>C</td>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>E</td>
<td>Glu</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>Q</td>
<td>Gln</td>
<td>glutamine</td>
</tr>
<tr>
<td>G</td>
<td>Gly</td>
<td>glycine</td>
</tr>
<tr>
<td>H</td>
<td>His</td>
<td>histidine</td>
</tr>
<tr>
<td>I</td>
<td>Ile</td>
<td>isoleucine</td>
</tr>
<tr>
<td>L</td>
<td>Leu</td>
<td>leucine</td>
</tr>
<tr>
<td>K</td>
<td>Lys</td>
<td>lysine</td>
</tr>
<tr>
<td>M</td>
<td>Met</td>
<td>methionine</td>
</tr>
<tr>
<td>F</td>
<td>Phe</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>P</td>
<td>Pro</td>
<td>proline</td>
</tr>
<tr>
<td>S</td>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>T</td>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>W</td>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Y</td>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>V</td>
<td>Val</td>
<td>valine</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to express my great appreciation to my supervisor Dr. Lawrence P. McIntosh. His patient guidance and constructive criticism has been instrumental to my growth as a scientist. Thanks to my committee Dr. Stephen Withers and Dr. Filip Van Petegem for their insight and advice. It has been a pleasure to work with lab mates past and present. Their invaluable knowledge and friendships have been a source of gladness in the lab. Thanks to Florian, Cecilia, Ben, Desmond, Geneviève, Karl, Chloe, Miriam, Soumya, Laura, Adrienne, Helen, Stacy and Mark. I hope to have the opportunity of working with you again one day. I am eternally grateful to my wife Christel, who has helped me through the difficulties and frustrations of research. Lastly, thanks to UBC for providing financial support for the research presented.
Chapter 1: Introduction

1.1 Protein structure function and dynamics

A mechanistic understanding of protein function requires a comprehensive analysis of its inter-related structural, dynamic and electrostatic properties. These properties are controlled by a delicate interplay of enthalpic and entropic interactions between the amino acids that are covalently linked to form the polypeptide chain of a protein and its surrounding solvent. Each of the twenty proteogenic amino acids have distinct physicochemical properties that establish the sequence-dependent conformation and function of a given protein. For example, under neutral pH conditions, some amino acid sidechains are typically positively (Arg, His, Lys) or negatively (Asp, Glu) charged, whereas others can be classified as neutral polar (Asn, Cys, Gln, Ser, Thr) or predominantly non-polar (Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Tyr, Val). These can be further classified by size, aromaticity, conformational flexibility, chemical reactivity, and so forth. The overarching goal of my thesis is to develop and apply NMR spectroscopic methods to characterize the polar functional groups of the amino acids that undergo pH-dependent ionization or are typically difficult to detect due to rapid hydrogen exchange with the solvent.

1.2 Structural and functional roles of polar amino acids

Polar amino acids serve a multitude of roles in proteins. A key feature to these residues is that they contain hydrogens covalently bonded to nitrogen, oxygen, and sulfur atoms (Fig. 1.1). In several cases, these hydrogens (protons) are involved in acid-base equilibria between the charged and neutral forms of an amino acid sidechain.

Polar amino acids are routinely involved in intramolecular and intermolecular (e.g. solvent) dipolar interactions and hydrogen bonding. The neutral sidechains of Asn, Cys, Gln,
Ser, Thr, Trp, and Tyr can act as hydrogen bond donors and/or acceptors. The strengths of these interactions vary greatly based on donor/acceptor type, distance and geometry, but typically stabilize the native fold of a protein by a favorable net free energy change of ~ 0.5 – 1.5 kcal/mol (Sheu et al., 2003). Oppositely charged amino acids can also interact electrostatically. The resulting ionic bonds between charged residues are often co-localized with hydrogen bonding to form salt bridges. In these cases, the positively charged functional groups of arginine, lysine, and histidine (and the N-terminal residue) typically serve the role of hydrogen bond donor, whereas the negatively charged sidechains of aspartate and glutamate (and the C-terminal residue) are commonly found as hydrogen bond acceptors. Salt bridges can stabilize by proteins by as much as ~ 3 - 5 kcal/mol (Anderson et al., 1990).

The catalytic residues of enzymes are always polar. Depending upon their ionization state, the sidechains of Asp, Arg, Cys, Glu, His, Lys, Ser, Thr and Tyr can serve as general bases/ acids, and nucleophiles in various catalytic mechanisms (Fersht, 2017).

A common feature to all polar residues is the ability of hydrogens on their nitrogen, oxygen, and sulfur atoms to readily exchange with solvent hydrogens. This is a pH-dependent phenomenon due to both acid- and base-catalyzed exchange and the possibility that a given sidechain may have different protonation states depending upon its acid dissociation constant (pKₐ value). In the context of my thesis, such exchanging hydrogens are termed labile. The rate at which a labile hydrogen exchanges with the solvent, and the pH at which a polar functional group deprotonates, are two very useful parameters that can be measured in order to probe the structure, dynamics, and function of a protein.
A useful way to analyze polar amino acids is to divide them into three categories based on their sidechain attributes. Under physiological conditions, Lys, His and Arg typically have positively charged sidechains, Asp and Glu negatively charged sidechains, and Cys, Ser, Thr, Asn and Gln polar neutral sidechains. The $pK_a$ values correspond to the amino acids X in acetyl-Gly-X-Gly-amide peptides (Platzer et al., 2014).
1.3 Characterizing labile protons by NMR spectroscopy: hydrogen exchange

Characterizing labile hydrogens in proteins is a challenging task. Hydrogens are generally not observed by X-ray crystallography except with very high resolution diffraction data (Ho and Agard, 2008). Thus their presence or absence is typically inferred through physicochemical arguments. In principle, neutron diffraction can be used to directly detect protons or deuterons, but in practice this technique has many demanding experimental constraints (Kossiakoff et al., 1990). Also, as will be discussed in Chapter 3, this technique has yielded contradicting results with respect to ionization states determined by other methods (Yonezawa et al., 2017; Yoshimura et al., 2017). Of course, protons are observable by \(^1\)H-NMR spectroscopy provided that they do not undergo very rapid exchange with solvent water or are not obfuscated by chemical shift degeneracy.

Dating back to the pioneering work of Linderstrøm-Lang, exchange measurements with various combinations of the three hydrogen isotopes (\(^1\)H = H = protium, \(^2\)H = D = deuterium and \(^3\)H = tritium) have been used to probe protein structure and dynamics. Hydrogen exchange is acid- and/or base-catalyzed depending on the functional group. In the case of an amide, the intrinsic pseudo-first order exchange rate constant \(k_{\text{int}}\) is given by equation 1.1

\[
k_{\text{int}} = k_{\text{OH}}[\text{OH}^-] + k_{\text{H}}[\text{H}_3\text{O}^+] + k_{\text{H}_2\text{O}}
\]

(1.1)

where \(k_{\text{OH}}\) is the second order rate constant for hydroxide-catalyzed exchange, \(k_{\text{H}}\) for hydronium-catalyzed exchange, and \(k_{\text{H}_2\text{O}}\) for a minor contribution of water-catalyzed exchange. The exchange of more label sidechain hydrogens can also be catalyzed by general acids and bases, and is thus dependent upon sample buffer as well as pH (Wuthrich and Wagner, 1979). The typical pH-dependence of \(k_{\text{int}}\) for various main chain and sidechain functional groups is presented in Fig. 1.2 (adapted from Wuthrich and Wagner, 1979).
Figure 1.2 Log-log plots showing the observed or predicted exchange rate constants, $k_{\text{int}}$, for labile groups on amide backbone and polar sidechains of the proteogenic amino acids in a random coil polypeptide (Wuthrich and Wagner, 1979). The exchange rates are pH-dependent and need to be below $\sim 10 - 100$ s$^{-1}$ for detection by $^1$H-NMR.
Early $^1$H-NMR studies on peptides and small proteins, including basic pancreatic trypsin inhibitor (BPTI), set the foundation for measuring and interpreting residue-specific protium/deuterium exchange (Wuthrich and Wagner, 1979). The reaction below represents a labile amide proton (NH) exchanging with deuterated water.

$$\text{D}_2\text{O} + \text{NH} \rightleftharpoons \text{ND} + \text{DHO}$$

The signal from the NH can be measured by a variety of approaches, including 1D $^1$H-NMR as well as 2D $^1$H-$^1$H COSY and $^{15}$N-HSQC spectroscopy. The exchange rate constant ($k_{ex}$) can be determined by fitting the signal intensity as a function of time to an exponential decay. This approach typically requires that an observable proton exchanges on the timescale of minutes or longer in order to transfer a protein into D$_2$O buffer and record NMR spectra of sufficient resolution and signal-to-noise. Alternatively, rapid exchange on the seconds timescale can be measured using an approach such as the CLEANEX-PM (Hwang et al., 1998) that monitors magnetization transfer from water to the group of interest. By varying sample pH, the rate of exchange can often be "tuned" for one of these two detectable timescale windows.

Labile protons participating in hydrogen bonding within the core of the protein are certainly expected to exchange much more slowly than those exposed on its surface. However, two remarkable observations from hydrogen exchange studies are that, given sufficient time, all the amides within a protein will exchange with water, and that different amides often exchange with different rates. This clearly demonstrates that proteins are dynamic, undergoing fluctuation ranging from local conformational changes to global unfolding, that lead to water contact and exchange.

A classical model for interpreting amide hydrogen exchange describes the protein in an equilibrium between two states: a closed (NH$_{\text{closed}}$ and ND$_{\text{closed}}$) exchange incompetent state and
an open exchange competent state (NH\textsubscript{open} and ND\textsubscript{open}). This model can be summarized by the reaction scheme:

\[
\begin{align*}
\text{NH}_\text{closed} & \xrightleftharpoons[k_\text{cl}]{k_\text{op}} \text{NH}_\text{open} & \text{NH}_\text{open} & \xrightarrow{k_\text{int}} \text{ND}_\text{open} & \text{ND}_\text{closed} & \xleftarrow{k_\text{op}} k_\text{cl}
\end{align*}
\]

where \(k_\text{op}\) and \(k_\text{cl}\) are the opening and closing rate constants for fluctuations to the exchange competent state, and \(k_\text{int}\) is the intrinsic exchange rate in that state. The latter has been well parameterized for polypeptide sequence, as well as sample temperature, ionic strength and isotope effects (Bai, 1999; Bai et al., 1995; Bai et al., 1993).

The overall observed hydrogen exchange rate constant \(k_\text{ex}\) from NH to ND (i.e. in ~ 100 % D\textsubscript{2}O) depends on these individual rate constants as follows:

\[
k_\text{ex} = \frac{k_\text{op} \times k_\text{int}}{k_\text{op} + k_\text{cl} + k_\text{int}} \quad (1.2)
\]

There are three limiting cases for this mechanism. The first and the simplest is when \(k_\text{op} \gg (k_\text{cl} + k_\text{int})\) and hence \(k_\text{ex} \approx k_\text{int}\). In this case, the amide hydrogen is always in the open state and exchange is not impeded. This describes readily exchangeable hydrogens that are exposed on the surface of a protein or present in disordered polypeptide regions.

In the case where \(k_\text{cl} \gg k_\text{op}\), there are two limits commonly called EX1 and EX2. In the EX1 limit, \(k_\text{int} \gg k_\text{cl}\) and equation 1.2 simplifies to \(k_\text{ex} = k_\text{op}\). In other words, exchange is limited only by the opening of the protein and independent of additional parameters such as sample pH (hence "unimolecular"). EX1 exchange provides kinetic insights on protein unfolding, but is normally seen only under high pH conditions where \(k_\text{int}\) is very fast (Ferraro et al., 2004).

In the EX2 limit, \(k_\text{cl} \gg k_\text{int}\) and equation 1.2 simplifies to exchange modeled by \(k_\text{ex} = (k_\text{op}/k_\text{cl})k_\text{int}\). This is "bimolecular" exchange because \(k_\text{int}\), and hence \(k_\text{ex}\), is pH-dependent. The
ratio of \( \frac{k_{op}}{k_{cl}} = \frac{k_{ex}}{k_{int}} \) can be determined by comparing the measured \( k_{ex} \) with the \( k_{int} \) value parameterized with suitable reference compounds. This ratio corresponds to the equilibrium constant \( K_{eq} \) for the fluctuations between the exchange competent (open) and incompetent (closed) states of the protein. Typically, \( 1/K_{eq} \) is called the protection factor (PF) for a given exchangeable hydrogen, and the free energy for the associated conformational fluctuation is \( \Delta G^o = -RT \ln(K_{eq}) = RT \ln(PF) \). For the most slowly exchanging amides, \( \Delta G^o \) often corresponds to the free energy change for global unfolding of a protein. For less protected amides, it corresponds to sub-global and local conformational fluctuations (Ferraro et al., 2004).

### 1.4 Characterizing labile protons by NMR spectroscopy: protonation states and acid dissociation constants

Several amino acids, as well as the N- and C-termini of a polypeptide, are ionizable and can be characterized with an acid dissociation constant \( -\log(K_a) \) or \( pK_a \) value; Fig 1.1) measurable in a variety of ways. A very powerful tool for probing residue and conformation specific protonation states is NMR spectroscopy. NMR-monitored pH titrations generally occur in the fast exchange regime with the chemical shifts of non-labile reporter nuclei \( \delta_{obs} \) being a population-weighted average between those of conjugate acid (protonated, \( \delta_{HA} \)) and base (deprotonated, \( \delta_{A} \)) form of an ionizable moiety. In the case of a simple ionization equilibrium, the resulting chemical shift versus pH curves can be fit to an adapted Henderson-Hasselbalch (equation 1.3) in order to extract the associated \( pK_a \) values.

\[
\delta_{obs} = \frac{\delta_{HA} 10^{pH} + \delta_{A} 10^{pK_a}}{10^{pH} + 10^{pK_a}}
\]

(1.3)
When two sequential protonation states influence the chemical shift of the observed nuclei, the biphasic titration curves can be fit using equation 1.4

$$\delta_{\text{obs}} = (\delta_{\text{HAH}} 10^{-2pH} + \delta_{\text{HA}} 10^{p\text{H}-p\text{Ka}_1} + \delta_{\text{A}} 10^{p\text{Ka}_1-p\text{Ka}_2})^+ / (10^{-2pH} + (10^{p\text{H}-p\text{Ka}_1}) + 10^{p\text{Ka}_1-p\text{Ka}_2})$$

(1.4)

where $\delta_{\text{HAH}}$, $\delta_{\text{HA}}$, and $\delta_{\text{A}}$ are the chemical shifts of the reporter nuclei in the three indicated protonation states, linked by the two macroscopic $p\text{Ka}_1$ and $p\text{Ka}_2$ values, respectively (McIntosh et al., 2011a).

Changes in protonation states are often directly linked to changes of other biochemically important phenomena. The following sections are examples from my published research showing the coupling of protonation states/$p\text{Ka}$ values to protein conformational changes and to enzymatic catalysis.

1.5 Characterizing pH-dependent conformational equilibria of proteins by NMR: a case study of PrgK

PrgK is a protein that is part of the type III secretion system (T3SS) from the pathogenic bacterium *Salmonella enterica* serovar Typhimurium. The T3SS is used to inject effector proteins into host cells (Galan et al., 2014). The basal body of the T3SS is primarily made up of three proteins, InvG, PrgH, and PrgK (Deng et al., 2017). These proteins form ring structures through which effector proteins are secreted (Schraidt and Marlovits, 2011). InvG, PrgH, and PrgK contain small globular domains called Ring Building Motifs (RBMs) that assemble to form the ring structures of the T3SS (Worrall et al., 2016). PrgK contains two of these RBM domains, termed D1 and D2, connected by a linker polypeptide (Fig. 1.3a,b). The D1 domain adopts two conformations that differ whether a portion of the linker is bound or disordered. The linker
bound conformation is called conformer A and the linker free one is conformer B (Fig. 1.3c) (Bergeron et al., 2015).

The difference between conformer A and B results from the cis/trans isomerization of the linker Leu77-Pro78 amide. This conclusion follows from an analysis of the $^{13}$C chemical shifts of the proline residues in PrgK (Fig. 1.4). Further studies also showed that the relative population of conformer B decreased under alkaline conditions. A reasonable hypothesis is this pH-dependent conformation change arises from a change in protonation state in one or more residues in PrgK. Comparing the structures of conformer A determined by X-ray crystallography and conformer B by Rosetta modeling indicated that a salt-bridge between His42 and Tyr75 only occurs in the former (Fig. 1.5). With an intrinsic pK$_a$ value of 6.5 (intrinsic pK$_a$ value meaning the pK$_a$ of a His in a disordered tripeptide reference compound), the deprotonation of His42 is an

Figure 1.3 (a) Schematic representation of PrgK, with the boundaries of the D1 domain, linker, and D2 domain indicated. (b) The D1 domain has a RBM fold. (c) D1 adopts two conformations, differing whether the linker bound (conformer A) or not (conformer B).
attractive explanation for this pH-dependent conformation change.

In support of this explanation, conformation specific $pK_a$ values for His42 were determined by NMR-monitored pH titrations (Fig. 1.6). In conformer A, His42 has a $pK_a$ value of 7.0 ± 0.1, whereas in conformer B, it is elevated to 7.7 ± 0.1. This leads to a model for the pH-dependent conformational equilibrium of PrgK summarized in (Fig. 1.7). In conformer A, the linker, including residues Leu77 and Pro78, interacts with a hydrophobic surface of D1 (Fig. 1.5). This interaction is also coupled with the ordering of the C-terminal end of the second helix of D1, which includes Tyr75 (Fig. 1.3c). However, favorable packing of the linker along the hydrophobic surface only occurs with the less favorable cis isomer of the Leu77-Pro78 peptide bond. Below pH 7, His42 is primarily protonated and cannot serve as a hydrogen bond acceptor for the O$^{\eta}$H of Tyr75. This destabilizes the interaction between D1 and the linker, which results in an equilibrium with conformer B, which has linker unbound and a trans Leu77-Pro78 peptide. As pH increases, His42 deprotonates to the neutral N$^{\delta+1}$H tautomer, allowing N$^{\varepsilon}$ to form a salt bridge with Tyr75, and thereby stabilize conformer A. The lower $pK_a$ value of His42 in conformer A than B reflects these coupled interactions.
Figure 1.4 The cis/trans isomerization of an X-Pro peptide is easily determined by the difference between the proline $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ chemical shifts. Differences of ~3 ppm and ~9 ppm are indicative of the trans or cis isomer, respectively. Based on this criteria, in conformer A of PrgK, Leu77-Pro78 adopts the cis isomer, whereas in conformer B, it adopts the trans isomer. In contrast, all other X-Pro amides are found as the trans isomer in both conformers.
Figure 1.5 A salt bridge between His42 and Tyr75 is only seen in conformer A. Shown is the crystal structure of the PrkG D1 domain (4WM4.pdb) with Leu77-Pro78 in a *cis* conformation and His42 hydrogen bonded to Tyr75 (Bergeron et al., 2018).
Figure 1.6 (a) The PrgK D1 domain is in slow exchange between conformers A and B. These give rise to separate peaks from His42 in a $^{13}$C-HSQC spectrum of the $^{13}$C-labeled protein. Each $^{13}$C$^{\delta_2}$ signal is a doublet due to scalar coupling with the adjacent $^{13}$C$^{\gamma}$. (b, c & d) Each conformer gives a distinct pH titration curve with different fit pK$_a$ values for His42. It is also noteworthy that upon deprotonation, the $^{13}$C$^{\delta_2}$ signal of His42 in conformer A shifts downfield by 4.5 ppm to 123.7 ppm. Furthermore, its $^{13}$C$^{\gamma}$ signal shifts upfield by > 1 ppm (not shown). These highly diagnostic patterns of pH-dependent chemical shift changes demonstrate that His42 of conformer A predominantly adopts the neutral N$^{\delta_1}$H tautomer under alkaline conditions. (Sudmeier et al., 2003; Platzer et al., 2014). In contrast, the $^{13}$C$^{\delta_2}$ signal of His42 in conformer B shifts downfield by only ~2 ppm to an extrapolated chemical shift of 121.4 ppm. These spectral changes indicate that, at high pH, His42 in this conformer exists as an approximately equimolar mixture of both neutral tautomers.
Figure 1.7 Model of the PrgK pH-dependent conformational equilibria. The schematic summarizes the conformational equilibria of PrgK involving coupling of linker binding and the ordering of the C-terminal helix of protein with the cis/trans isomerization of Leu77/Pro78 and the pH-dependent interaction of Tyr75 with His42. At low pH, His42 is predominantly charged, disrupting its ability to act as a hydrogen bond acceptor for Tyr75. This results in a roughly equal distribution of conformers A and B. As pH increases, His42 deprotonates allowing it to form a salt bridge with Tyr75 and stabilizing conformer A.
16 Linking residue specific pKₐ values to catalysis: a case study of BcX

The 20 kDa xylanase from Bacillus circulans (BcX) is an extensively studied glycoside hydrolase that cleaves xylan, a principle component of hemicellulose. More generally, glycoside hydrolases catalyze the hydrolysis of the glycosidic linkages of glycosides and have long been studied due to their biological and industrial importance. As with most enzymes, they are commonly classified based on their substrate specificity. This includes a number of different parameters such as sites of action, their preference for α- or β-linked carbohydrates, and whether the enzyme acts on the end (exo-acting) or the interior (endo-acting) of the carbohydrate chain. Although describable with an enzyme commission number (EC 3.2.1.x), a much more powerful classification system is that of the Carbohydrate-Active Enzyme database (www.cazy.org). Developed by Dr. Bernard Henrissat, CAZy divides glycoside hydrolases, glycosyltransferases, and carbohydrate binding modules (CBMs) into types/clans and families based on sequence, structure, and specificity comparisons (Consortium, 2018). Although these enzymes can be extremely diverse, they generally only utilize a few common catalytic mechanisms. In particular, glycoside hydrolases are broadly classified as inverting or retaining in reference to the net stereochemical outcome for the anomeric centre at the site of hydrolysis. For example, BcX is a family GH-11 retaining endo-β-(1,4)-xylanase.

1.6.1 Inverting mechanism of glycoside hydrolases

Glycosyl hydrolases that invert the stereochemistry of the carbohydrate anomeric centre often do so through a single-step Sₙ₂ displacement mechanism (Rye and Withers, 2000) (Fig.
Typically two catalytic residues (usually Asp or Glu) are required with one serving as a general acid to donate a proton to the aglycone leaving group and the other as a general base to deprotonate a nucleophilic water for attack at the anomeric centre (McCarter and Withers, 1994; Yip and Withers, 2004). The reaction goes through an oxocarbenium ion-like transition state and results in the inversion of the anomeric stereochemistry of the glycone product.

### 1.6.2 Retaining mechanism of glycoside hydrolases

Glycoside hydrolases that retain the stereochemistry of the glycone carbohydrate anomeric centre typically do so through a Koshland double-displacement mechanism (Koshland, 1953) involving two key catalytic residues (Rye and Withers, 2000). In the glycosylation step, one residue (usually an Asp or Glu) acts as a nucleophile and attacks the anomeric centre. The other residue, again typically an Asp or Glu, acts as a general acid to protonate the aglycone leaving group (Fig. 1.8b) (McCarter and Withers, 1994). Proceeding through an oxocarbenium ion-like transition state, this yields a glycosyl-enzyme intermediate. In the subsequent deglycosylation step, the now-deprotonated general acid acts as a general base to deprotonate a nucleophilic water molecule that attacks the anomeric centre to release the glycone with a net retention of anomeric stereochemistry. BcX is a textbook example of retaining glycoside hydrolase that utilizes Glu78 as a nucleophile and Glu172 as a general acid/base for catalysis. Because xylan is a valuable industrial starting material, considerable effort has been made to engineer xylanases, such as BcX, for tailored activity under potentially harsh conditions of reaction pH value and temperature.
Figure 1.8 (a) Inverting mechanism of glycoside hydrolases with a postulated transition state shown. (b) Retaining mechanism of glycoside hydrolases.
1.6.3 The role of protonation states in general acid/base and nucleophilic catalysis

General acid/base catalysis is extremely common to enzymatic mechanisms, particularly given the limited number of amino acids capable of performing these tasks at physiological pH conditions. A base in these reactions is typically a residue with a $pK_a$ value one or two units below the pH optimum ($pH_{opt}$) of the enzyme (Fersht, 2017). This puts the catalytic residue predominantly in the required deprotonated state, while also allowing it to accept a proton over the course of the reaction (e.g., if its $pK_a$ value was very low, it would be the conjugate base of a strong acid thus a poor proton acceptor near $pH_{opt}$). A residue with a $pK_a$ value one unit below the reaction $pH_{opt}$ is $\sim 91\%$ in its deprotonated general base and $\sim 9\%$ in its conjugate general acid form state and thus can readily accept a proton for the reaction to proceed; two units below puts these population at $\sim 99\%$ and $\sim 1\%$, respectively. These values are readily obtained from the Henderson-Hasselbalch equation:

$$\frac{[A^-]}{[HA]} = 10^{pH-pK_a}$$

Nucleophilic catalysis shares many of the same properties as general base catalysis. Nucleophiles are Lewis bases and typically are either negatively charged or neutral with an unshared pair of electrons. This allows them to easily form covalent bonds with electron deficient atoms (Lewis acids). Also, like general bases, ionizable catalytic nucleophiles typically have a $pK_a$ value one or two units below the pH optimum of the reaction. This allows them to still readily covalently bond to Lewis acids.

By the same principles, a good general acid is typically a residue with a $pK_a$ value one to two units above the $pH_{opt}$ of the reaction. This puts most of its population in a catalytically competent protonated state that can readily donate a proton for the reaction to proceed.
In the case of BcX, the bell-shaped pH-dependent $k_{\text{cat}}/K_m$ activity profile of this model GH (Fig 1.10a dotted line) can be explained by the NMR-measured $pK_a$ values of its catalytic residues (Ludwiczek et al., 2013). Note that $k_{\text{cat}}/K_m$ is the second order rate constant for reaction of free enzyme and substrate. The substrates used for kinetic measurements of BcX are not ionizable, and thus the pH-dependence of $k_{\text{cat}}/K_m$ can be attributed to changes in the enzyme. At its pH$_{\text{opt}}$ 5.7, Glu78 ($pK_a = 4.6$) is deprotonated to serve as a nucleophile for the rate-limiting glycosylation step, whereas Glu172 ($pK_a = 6.7$) is protonated to act as a general acid. Subsequently, the $pK_a$ value of Glu172 drops to 4.2 in a long lived glycosyl-enzyme intermediate with Glu78 covalently bonded to a 2-deoxy-2-fluoro-β-xylobiosyl moiety. The dramatic reduction in the $pK_a$ value of Glu172, which is due primarily to the loss of charge repulsion from Glu78, enables this residue to act as a general base and thereby facilitate deglycosylation. Importantly, the "$pK_a$ cycling" of Gly172 results intrinsically from electrostatic changes that occur along the reaction pathway of BcX as Glu78 alternates between its negatively-charged nucleophilic and neutral covalently-modified states.
Figure 1.9 The catalytic site of BcX. Glu78 and 172 are the catalytic residues acting as nucleophile and general acid/base respectively. Residues in the first shell help catalysis by stabilizing substrate binding or perturbing the pKₐ of the catalytic residue (pdb 2bvv).
1.6.4 Engineering BcX for altered pH-dependent activity

In an effort to increase the industrial applications of BcX and to better understand the electrostatic basis of catalysis, numerous mutants of BcX have been produced (Reitinger et al., 2010; Ludwiczek et al., 2013; Joshi et al., 2001; Reitinger et al.). These can be broadly collected into three groups: mutations to the catalytic residues Glu78 and Glu172, mutations to non-catalytic residues in the active site ("first shell"), and mutations to residues adjacent to the active site ("second shell") (Fig. 1.9). To highlight the interplay between pKₐ values and catalysis, the following sections will focus on two variants of BcX, that I helped characterize (Ludwiczek et al., 2013). One has a mutation of the catalytic general acid/base, E172H, and the other a first shell mutation, N35H. Both of these substitutions alter the pH-dependent enzymatic activity of BcX, but do so through very different means.

Remarkably, despite its absolute conservation in naturally occurring GH-11 family members, mutation of Glu172 to a histidine resulted in an enzyme with 8% activity towards 2'-nitrophenyl β-xylobioside (ONPX2, Fig 1.10a) (Ludwiczek et al., 2013). The decreased activity (kₗ/Kₘ) is primarily due to reduction in kₗ, rather than Kₘ, value. Perhaps even more striking, the E172H variant has a pHₗ of 4.67, which is ~ 1 log unit less than that of WT-BcX. Fitting the pH-dependent kₗ/Kₘ profile yielded apparent pKₐ values of 3.78 and 5.56 for the acidic and basic limbs of the bell-shaped curve, respectively. The simplest interpretation of these data is that the two pKₐ values correspond to the Glu78 and His172, and that these residues serve as the catalytic nucleophile and general acid.
Figure 1.10 (a) pH-activity profile of E172H-BcX shows a pH optimum of 4.7 and two curves with inflection points at of 3.78 and 5.56. The fit activity curve of WT-BcX is included for comparison (dotted line, right y-axis; pK_aGlu78 = 4.39, pK_aGlu172 = 6.80, pH_opt = 5.60). (b) Fitting the pH-dependent $^{13}$C chemical shifts of Glu78 (spectra not shown) yielded a pK_a value of 3.39 for this residue. (c) pH titrations of the histidine residues in $^{13}$C-labeled E172H-BcX monitored with $^{13}$C-HSQC spectra (overlaid). His156 is distal to the active site and not considered further. (d) Fitting the pH-dependent $^1$H chemical shifts of E172H yielded a pK_a value of 5.22 for this residue. (e) X-ray crystallographic structure of E172H-BcX with the four monomers from the asymmetric unit superimposed. Selected residues are shown in stick format (carbon, green for mutant, grey for WT; oxygen, red; nitrogen, blue; fluorine, magenta) (data acquired by Matrin Ludwiczek).
To confirm the assignment of these kinetically-determined pK<sub>a</sub> values, derived from the pH-dependence of the second-order rate constant (k<sub>cat</sub>/K<sub>m</sub>) for the reaction of free enzyme and substrate, NMR spectroscopy was used to site-specifically probe the Glu and His residues in E172H-BcX. Using <sup>13</sup>C-NMR to monitor the pH-dependent spectra of the <sup>13</sup>C<sup>δ</sup>-glutamic acid labeled protein, a biphasic titration curve was observed for Glu78 (Fig 1.10b). The major transition, attributed to the deprotonation of Glu78, fit to a pK<sub>a</sub> value of 3.39, whereas the minor, attributed to perturbations from the titration of nearby His172, fit to an apparent pK<sub>a</sub> value of 4.96. Conversely, the <sup>1</sup>H<sup>ε1</sup> of His172 followed a major titration with a pK<sub>a</sub> value of 5.22, assigned to its own deprotonation, and a minor value of 3.46 that likely reflects the behaviour of Glu78 (Fig. 1.10c,d). The approximate agreement of these NMR-measured pK<sub>a</sub> values for Glu78 and His172 and their correspondence with the apparent pK<sub>a</sub> values governing the activity profile of E172H-BcX provides strong support for the proposed catalytic role of His172 as a general acid.

The X-ray crystallographic structure of E172H-BcX in its apo form supports these conclusions (Figure 1.10e). Overall, the structure superimposes closely with that of wild-type BcX, including the position of the partially buried nucleophile Glu78 (fractional sidechain accessible surface area ASA ~ 0.1, and hydrogen bonded to Tyr69 and Gln127 (Willard et al., 2003)). However, in contrast to the favorable hydrogen bonding interactions of Glu172 with Asn35 and Tyr80 observed in WT-BcX (not shown), the sidechain of His172 shows variable conformations in the four monomers of the asymmetric unit. In each conformation, His172 is unable to hydrogen bond with any neighboring residues. Based on the pH-dependence of its <sup>13</sup>C<sup>γ</sup> and <sup>13</sup>C<sup>δ2</sup> chemical shifts (Fig 1.11), His172 is most likely in the neutral N<sup>ε2</sup>H tautomeric form.
under the crystallization conditions of pH 7 to 8. However, at pH > 6, the $^{13}$C-δ-1Hδ2 and $^{13}$Cγ-1Hββ' NMR signals of the histidine sidechain were not detected. This is suggestive of signal broadening due to msec-μsec timescale conformational exchange and/or tautomerization of neutral His172. Also consistent with the NMR-monitored titration data for E172H-BcX, the close proximity of Glu78 and His172 (~ 6 Å) should result in favorable electrostatic interactions, thereby lowering the pKₐ value of the catalytic nucleophile and raising that of the general acid. However, its partially buried environment (fractional sidechain ASA ~ 0.1) and lack of hydrogen bonding partners will reduce the pKₐ value of His172. The balance of these opposing effects likely leads to the overall reduction in the pHopt of E172H-BcX relative to the WT parent.

Three key features of Fig 1.11 that helped determine H172’s pKₐ and tautomeric state are worth noting. First, the $^{15}$Nδ1 and $^{15}$Nε2 are highly sensitive to the ionization state of a histidine ring, and have chemical shifts of ~ 170 ppm when protonated and ~ 250 ppm when deprotonated. (Pelton et al., 1993) At pH 3.93, the unassigned imidazole $^{15}$N’s of His172 have chemical shifts of 174.0 and 179.7 ppm (Fig 1.11 f), and thus its sidechain must be predominantly charged under this experimental condition. This strongly supports the assignment of the pKₐ value of 5.22 extracted from the biphasic titration curve of 1Hε1 shown in Figure 1.9d to the deprotonation of His172 and that of 3.46 to nearby Glu78 (and not vice versa).
Figure 1.11 the overlaid spectra of selectively $^{13}\text{C}/^{15}\text{N}_3$-histidine-labeled E172H-BcX titrated from approximately pH 7.3 to 2.4 at 25 °C (arrows indicate increasing sample pH). Shown are portions of $^{15}\text{N}$-decoupled $^{13}\text{C}$-HSQC spectra corresponding the (a) $^{13}\text{C}^\beta - ^1\text{H}^{\beta\beta}$, (b) $^{13}\text{C}^{\delta 2} - ^1\text{H}^{\delta 2}$, and (c) $^{13}\text{C}^{e 1} - ^1\text{H}^{e 1}$ signals, respectively. Although the $^{13}\text{C}^\delta - ^1\text{H}^{\beta\beta}$ signals were aliased, they are presented with corrected chemical shifts. The split $^{13}\text{C}^{\delta 2}$ and broad $^{13}\text{C}^\beta$ peaks are due to $^{13}\text{C}-^{13}\text{C}$ couplings. An out-and-back HB(CB)CG(Lohr and Ruterjans, 1996) experiment was used detect the $^{13}\text{C}^\gamma - ^1\text{H}^{\beta\beta}$ peaks in (d). The histidine amide $^{15}\text{N} - ^1\text{H}^N$ signals observed in $^{13}\text{C}$-decoupled sensitivity-enhanced $^{15}\text{N}$-HSQC spectra are shown in (e). Finally, in (f), a portion of the long-range $^{13}\text{C}$-decoupled $^{15}\text{N}$-HSQC spectrum (using an INEPT "1/4J" delay of 10 msec to detect $^2\text{J}_{\text{NH}}$ and $^3\text{J}_{\text{NH}}$ couplings) recorded at pH 3.93 is presented. The $^{15}\text{N}^{\delta 1}$ and $^{15}\text{N}^{e 2}$ signals of protonated His156 were not resolved (Plesniak, Connelly, et al., 1996) and those of His172 were not specifically assigned. The * indicate the peaks corresponding to the $^1\text{H}^{e 1}$ for each histidine.(Pelton et al., 1993) In the case of His149, which adopts the neutral $^{15}\text{N}^{e 2}$ tautomeric form ($pK_a < 2.3$), the signal from the deprotonated $^{15}\text{N}^{\delta 1}$ is at 249 ppm (see (Plesniak, Connelly, et al., 1996))
Second, based on model compounds, deprotonation of histidine to its $\text{N}^{\epsilon 2}\text{H}$ tautomeric form leads to chemical shift changes of $\sim +7$ ppm (i.e., downfield) for the $^{13}\text{C}^\gamma$ nucleus and $\sim -2$ ppm for the $^{13}\text{C}^\delta^2$, whereas the opposite changes of $\sim -2$ ppm and $\sim +7$ ppm, respectively, result for formation of its $\text{N}^{\delta 1}\text{H}$ tautomer. (Reynolds et al., 1973; Goux and Allerhand, 1979) Also, a $^{13}\text{C}^\delta^2$ chemical shift $> 122$ ppm is diagnostic of the $\text{N}^{\delta 1}$ tautomer. (Sudmeier et al., 2003). The downfield change in $^{13}\text{C}^\gamma$ shift (Fig 1.11 d) and upfield change in $^{13}\text{C}^\delta^2$ shift (Fig 1.11 b) observed for the catalytic His172 with increasing pH indicate that the $\text{N}^{\epsilon 2}\text{H}$ tautomer is preferentially adopted. By comparison, exposed His156 showed $\sim +2$ ppm and $+1.4$ ppm downfield changes in both $^{13}\text{C}^\gamma$ and $^{13}\text{C}^\delta^2$ shifts, respectively with increasing pH, indicating approximately equal populations of both tautomers in rapid equilibrium.

Third, the $^{13}\text{C}^\delta^2-\text{H}^\delta^2$ signal (Fig 1.11 b) of His172 was only detected at sample pH values $< 6$ and the $^{13}\text{C}^\gamma-\text{H}^{\beta\beta'}$ signals (Fig 1.11 d) only at pH values $< 4.5$. Raising the sample temperature to 35 °C did not help. This precluded measurement at higher pH of these shifts for neutral His172 and use of the long-range $^{15}\text{N}$-HSQC experiment to confirm its tautomeric form. Importantly, the absence of these signals is suggestive of msec-μsec timescale conformational exchange and/or tautomerization of neutral His172 leading to signal broadening.
1.6.5 Reverse protonation

In the above discussed scenario, the $pK_a$ value of the nucleophile is less than that of the general acid, and thus the enzyme is predominantly in its catalytically competent protonation state at its $pH_{opt}$ (Figure 1.12a). However, in many enzymes (Mock, 1992; Ludwiczek et al., 2013; Frankel et al., 2005; Knuckley et al., 2007; Sims et al., 2003; Vocadlo et al., 2002) it has been found that the $pK_a$ value of the general acid is lower than that of the nucleophile or general base. In this so-called "reverse protonation mechanism," at its $pH_{opt}$, only a small fraction of the enzyme is in its catalytically competent protonation state (Fig 1.12b). This minor population can be surprisingly efficient for catalysis.

Previously, it was demonstrated that substitution of Asn35 (which directly contacts Glu172) with Asp significantly reduced the $pH_{opt}$ of BcX to a value of 4.6, along with a small increase in activity (Joshi et al., 2000). This arises from a reverse protonation mechanism in which the Asp35-Glu172 dicarboxyl pair serves as the general acid with a $pK_a$ value lower than that of the nucleophile Glu78. To further investigate this phenomenon, a His residue at position 35 was introduced. Remarkably, the resulting N35H-BcX variant exhibited ~ 20% increased activity towards ONPX2, and followed a bell-shaped $k_{cat}/K_m$ versus pH curve with apparent $pK_a$ values of 3.12 and 5.65 and a $pH_{opt}$ of 4.39 (Fig 1.13a). Thus, as with the initial N35D variant, either mutation shifted the activity profile of BcX by > 1 log unit towards more acidic conditions. This result is somewhat surprising as His35 can be a positively-charged residue near the catalytic Glu78 and Glu172, whereas Asp35 can be negatively-charged, yet both reduced the $pH_{opt}$ of BcX. To understand these seemingly contradictory effects, NMR spectroscopy was used to monitor the pH-titrations of the His and Glu residues in N35H-BcX.
Figure 1.12 Catalytically competent populations of general acid / base enzymes. (a) A general base with a pK\textsubscript{a} of 4 and a general acid with a pK\textsubscript{a} of 6 put ~ 99% of the enzymatic population in the correct protonation state at the pH\textsubscript{opt} of 5. (b) A general base with a pK\textsubscript{a} of 6 and a general acid with a pK\textsubscript{a} of 4 put ~ 1% of the enzymatic population in the correct population state at the same pH\textsubscript{opt}.
The ionization equilibria of Glu78, Glu172, and His35 in N35H-BcX were investigated by measuring their pH-dependent $^{13}\text{C}_{\delta}$ and $^{13}\text{C}_{\varepsilon1}$-1$^1\text{H}_{\varepsilon1}$ chemical shifts, respectively (Figure 1.13b-e). In all cases, multiphasic titration curves were observed, indicating interactions between the three residues. The $^{13}\text{C}_{\delta}$ of Glu78 exhibited a major titration with a $pK_a$ value of 5.47, assigned to its own ionization. This conclusion is supported by a similar $pK_a$ reported by Gln7, a residue at the edge of the BcX active site with a $^{13}\text{C}_{\delta}$ chemical shift that is particularly sensitive to the ionization state of Glu78. Glu78 also showed a minor titration with a $pK_a$ value of 2.96, attributable at least in part to perturbations by Glu172. Conversely, the $^{13}\text{C}_{\delta}$ of Glu172 followed a major titration with a $pK_a$ value of 3.39 that is predominantly due to its own ionization, as well as a minor titration with a $pK_a$ value of 7.31. The latter is attributed to His35. In a reciprocal fashion, the $^1\text{H}_{\varepsilon1}$ and $^{13}\text{C}_{\varepsilon1}$ of His35 showed its own titration at a $pK_a$ value of 7.48, as well as smaller pH-dependent perturbations at $pK_a$ values of 5.63 or 5.49, respectively, which likely arise from the ionization of Glu78. At lower sample pH values, the $^{13}\text{C}_{\varepsilon1}$-$^1\text{H}_{\varepsilon1}$ signals of His35 change in a highly non-linear fashion, such that the $^{13}\text{C}_{\varepsilon1}$ chemical shift follows a $pK_a$ value of 2.20 whereas the $^1\text{H}_{\varepsilon1}$ follows one at 3.80; the latter is likely due to Glu172. This perplexing behaviour indicates that, as with E172H-BcX, the directly-bonded $^1\text{H}_{\varepsilon1}$ and $^{13}\text{C}_{\varepsilon1}$ nuclei are sensitive to different pH-dependent changes in BcX.

Collectively, these measurements indicate that His35 has an elevated $pK_a$ value of 7.48 and thus is positively-charged under the more acidic conditions where N35H-BcX is active. As a result of this favorable electrostatic interaction, the $pK_a$ value of adjacent Glu172 is reduced dramatically to 3.39. Concomitantly, the $pK_a$ value of Glu78 is elevated to 5.47, suggesting that unfavorable interactions with Glu172 dominate over any favorable interactions with His35. The approximate correspondence of these latter two NMR-measured $pK_a$ values with those
determined kinetically indicates that the acidic limb of the bell-shaped activity profile of N35H-BcX is set by the general acid Glu172, whereas the basic limb is set by the nucleophile Glu78. This is opposite to the case of WT-BcX, and indicative of a reverse protonation mechanism in which the group with the lower $pK_a$ value is the general acid and that with the higher value is the nucleophile. Therefore, at the $pH_{opt}$ of N35H-BcX, only a small fraction is in the catalytically-competent ionization state with the nucleophile deprotonated and the general acid still protonated. Although only present with a low population at the $pH_{opt}$ of N35H-BcX, Glu172 is effectively a stronger general acid and therefore the net activity of the enzyme is comparable to that of WT-BcX. This is somewhat easier to understand than the case of the originally characterized N35D variant in which the Asp35-Glu172 dicarboxyl pair serves as the general acid to set the acidic limb of its pH-activity profile.
Figure 1.13 (a) pH-activity profile of N35H-BcX shows a pH_{opt} of 4.39 and inflection points at pH 3.12 and 5.65. The fit activity curve of WT-BcX is included for comparison (dotted line, right y-axis; pK_{aGlu78} = 4.39, pK_{aGlu172} = 6.80, pH_{opt} = 5.60). (b) Fitting the pH-dependent $^{13}\text{C}\delta$ chemical shifts of Glu78, Glu172 and the non-ionizable reporter Gln7 (which primarily senses the ionization state of Glu78; (McIntosh et al., 1996b, 1996a)) yielded the indicated pK_{a} values. (c) Overlaid $^{13}\text{C}$-HSQC spectra, showing the pH-dependent $^{13}\text{C}_{\epsilon^1}$-$^{1}\text{H}_{\epsilon^1}$ signals of His35 and His156 (arrows indicate increasing sample pH). His156 is distal to the active site and not considered further. Fitting the pH-dependent $^{1}\text{H}_{\epsilon^1}$ (d) and $^{13}\text{C}_{\epsilon^1}$ (e) chemical shifts of His35 yielded the indicated pK_{a} values (data acquired by Marin Ludwiczek).
1.7 Specific aims of the thesis

As exemplified by the above case studies, the overall aim of my thesis is the explicit characterization of the protonation states and pKₐ values of ionizable functional groups serving in biologically interesting contexts. Chapter 2 of my thesis presents NMR methods development for the direct detection of serine and threonine hydroxyl protons. This work was done in the very well characterized BcX. Chapter 3 of my thesis summarizes the characterization of a single threonine to histidine (T26H) point mutant in T4 lysozyme (a glycoside hydrolase family 24 member) that remarkably changes its catalytic mechanism from inverting to retaining. I determined the pKₐ values of all Asp, Glu, and His residues in the T26H variant and concluded that this enzyme also likely follows a reverse protonation mechanism. The final chapter briefly provides concluding remarks and a short discussion of possible future experiments.
Chapter 2: Detection and characterization of serine and threonine hydroxyl protons in *Bacillus circulans* xylanase by NMR spectroscopy

2.1 Overview

Hydroxyl protons on serine and threonine residues are not well characterized in protein structures determined by both NMR spectroscopy and X-ray crystallography. In the case of NMR spectroscopy, this is in large part because hydroxyl proton signals are usually hidden under crowded regions of $^1$H-NMR spectra and remain undetected by conventional heteronuclear correlation approaches that rely on strong one-bond $^1$H-$^{15}$N or $^1$H-$^{13}$C couplings. However, by filtering against protons directly bonded to $^{13}$C or $^{15}$N nuclei, signals from slowly-exchanging hydroxyls can be observed in the $^1$H-NMR spectrum of a uniformly $^{13}$C/$^{15}$N-labeled protein. Here I demonstrate the use of a simple selective labeling scheme in combination with long-range heteronuclear scalar correlation experiments as an easy and relatively inexpensive way to detect and assign these hydroxyl proton signals. Using auxotrophic *E. coli* strains, I produced *Bacillus circulans* xylanase (BcX) labeled with $^{13}$C/$^{15}$N-serine or $^{13}$C/$^{15}$N-threonine. Signals from two serine and three threonine hydroxyls in these protein samples were readily observed via $^3$J$_{CH}$ couplings in long-range $^{13}$C-HSQC spectra. These scalar couplings (~ 5 to 7 Hz) were measured in a sample of uniformly $^{13}$C/$^{15}$N-labeled BcX using a quantitative $^{13}$C/$^{15}$N-filtered spin-echo difference experiment. In a similar filtering approach, the threonine and serine hydroxyl hydrogen exchange kinetics were measured using a $^{13}$C/$^{15}$N-filtered CLEANEX-PM pulse sequence. Collectively, these experiments provide insights into the structural and dynamic properties of several serine and threonine hydroxyls within this model protein.
2.2 Introduction

Serine and threonine residues often play key structural and functional roles in proteins due to their sidechain hydroxyls serving as hydrogen bond donors/acceptors (Creighton, 2010). However, unless neutron (Kossiakoff et al., 1990) or very high resolution X-ray diffraction (Ho and Agard, 2008) data are available, the positions of hydroxyl protons are most often inferred, rather than directly observed, in crystallographically determined structures of proteins. Similarly, only ~ 1% of the NMR spectral assignments reported for serine and threonine residues in the BioMagResBank correspond to hydroxyl protons. Several factors have contributed to this under-representation, including the generally rapid exchange of these labile protons (Liepinsh et al., 1992; Liepinsh and Otting, 1996) and the lack of any one-bond scalar couplings with neighbouring spin-1/2 nuclei. Indeed, most current NMR methods for studying proteins rely upon $^{15}$N- or $^{13}$C-based heteronuclear approaches, and thus signals from slowly exchanging serine $^{1}$Hγ and threonine $^{1}$Hγ1 hydroxyl protons will not be readily detected. Furthermore, with average chemical shifts of 5.47 ± 1.1 ppm and 5.19 ± 1.2 ppm, respectively (Ulrich et al., 2008), their signals will usually be buried under the envelope of water and protein peaks in crowded $^{1}$H-NMR spectra. Accordingly, the limited number of hydroxyl assignments reported to date have been derived mainly from potentially ambiguous through-space $^{1}$H-$^{1}$H NOE interactions, and to a lesser extent, via $^{3}J_{H\gamma H\beta}$ couplings detected in scalar correlation spectra (Liepinsh et al., 1992; Knauf et al., 1993; Peelen and Vervoort, 1994; Takeda et al., 2011). It is noteworthy that hydroxyl signals have also been observed due to scalar couplings across OH•••OP hydrogen bonds in $^{31}$P-HMQC spectra of a flavoprotein (Lohr et al., 2000), as well as in dipolar-based solid state NMR spectra of a deuterated microcrystalline SH3 domain (Agarwal et al., 2010).
The goal of this study was to further develop NMR spectroscopic methods for detecting and unambiguously assigning the signals from serine and threonine hydroxyls in proteins. As a model system, I focused on the well characterized 20 kDa xylanase from *Bacillus circulans* (BcX). When the 1D $^1$H-NMR spectrum of the uniformly $^{13}$C/$^{15}$N-labeled protein is recorded with filtering against protons directly bonded to $^{13}$C or $^{15}$N nuclei, numerous signals are still detected (Fig. 2.1). These most likely arise from slowly exchanging hydroxyl groups as BcX lacks any cysteines and, with the exception of the catalytic general acid Glu172, all carboxyl moieties are deprotonated under the experimental conditions of pH 6 (McIntosh et al., 2011a; McIntosh et al., 2011b). The most downfield of the OH signals were assigned previously to 4 of the 15 tyrosines in BcX via 3-bond $^1$H-$^{13}$C$_\varepsilon$ couplings detected in a long-range $^{13}$C-HSQC spectrum (Baturin et al., 2011). Thus, I speculated that many of the remaining signals arose from some of the 18 serines and 25 threonines present in this protein. Using a combination of amino acid selective isotopic labeling and long-range $^{13}$C-HSQC experiments, I confirmed this hypothesis and assigned the signals from two serine and three threonine hydroxyl protons. These assignments enabled structural and dynamic studies of BcX through $^3$J$_{HC}$ scalar coupling and hydrogen exchange (HX) measurements, respectively.
Figure 2.1 (a) The $^1$H-NMR spectrum of uniformly $^{13}$C/$^{15}$N-labeled BcX recorded without $^{13}$C/$^{15}$N-decoupling. Downfield of amide envelope are apparent singlets arising from two tyrosine O$^\beta$H protons and a doublet at 12.2 ppm from the $^{15}$N-bonded $^1$H$^{\varepsilon 2}$ of His149 (Baturin et al., 2011; Plesniak, Connelly, et al., 1996). (b) The $^{13}$C/$^{15}$N-filtered spectrum of $^{13}$C/$^{15}$N-BcX reveals many additional $^1$H signals originating from oxygen-bonded protons, five of which were assigned in this study. Weak noisy peaks, such as those in the methyl region, arise from incomplete filtering. The spectra were recorded at pH 6 and 25 °C (850 MHz spectrometer) using the pulse sequence in Appendix Fig. A1. This sequence exploits two different spin echo delays to filter signals from $^1$H-$^{13}$C pairs with a wide range of $^1$J$_{CH}$ values (Ogura et al., 1996; Zwahlen et al., 1997; Breeze, 2000; Iwahara, 2001 #380), and a WATERGATE approach to dephase the signal from water (Piotto et al., 1992).
2.2 Materials and Methods

2.2.1 Protein expression and purification

The expression and purification of BcX followed previously described protocols (Sung et al., 1993; Joshi et al., 2001). Briefly, uniformly $^{13}\text{C}/^{15}\text{N}$-labeled BcX was produced in prototrophic *E. coli* BL21 (λDE3) cells transformed with a pET22b(+) vector encoding the enzyme. The cells were grown in LB media at 37 °C to an OD$_{600}$ of ~ 0.6, pelleted via gentle centrifugation, and resuspended in M9 media containing 1 g/L 98% $^{15}\text{NH}_4\text{Cl}$ and 3 g/L 99% $^{13}\text{C}_6$-glucose. After 45 minutes at 30 °C, BcX expression was induced with 0.1 mM IPTG.

BcX samples selectively labeled with $^{13}\text{C}/^{15}\text{N}$-serine or -threonine were produced using *E. coli* JC158 or JW0003-2 cells, respectively. The auxotrophic cells were obtained from the Yale Coli Genetic Stock Centre, transformed with a pCW-BcX vector, and grown in an artificial rich medium containing 0.5 g/L of $[^{13}\text{C}_3/^{15}\text{N}]$-L-serine (Sigma-Aldrich) or 0.12 g/L of $[^{13}\text{C}_4/^{15}\text{N}]$-L-threonine (Sigma-Aldrich), along with all other unlabeled amino acids (Muchmore et al., 1989). In each case, the auxotrophic *E. coli* cells were grown at 37 °C to OD$_{600}$ ~ 0.6 in LB media, pelleted by gentle centrifugation, and resuspended in the medium lacking either serine or threonine. After growth for 45 minutes at 30 °C, the labeled amino acid was added and protein expression induced with 1 mM IPTG.

Four to six hours post-induction, the cells were harvested by centrifugation and lysed by sonication in the presence of a protease inhibitor cocktail (Roche Diagnostics). The resulting BcX samples were purified using an SP-sepharose ion exchange column (GE Healthcare) with a 0 - 1 M NaCl gradient in 10 mM sodium phosphate buffer, pH 6, followed by gel filtration on a HiPrep 16/60 Sphacryl S-100 size exclusion column (GE Healthcare) equilibrated with 10 mM sodium phosphate buffer, pH 6. The final samples were concentrated to ~ 0.6 mM in the latter
buffer via ultrafiltration with a 10 KD MWCO Amicon Ultra spin column, and 5% D$_2$O was added as a lock solvent.

### 2.2.2 Protein characterization

*E. coli* JW0003-2 (F-, ΔthrC724::kan, Δ(araD-araB)567, ΔlacZ4787::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514) from the Keio collection (Baba et al., 2006) is a well-behaved threonine auxotroph. The resulting BcX was selectively $^{13}$C/$^{15}$N-labeled at only these residues, as verified by $^{13}$C- and $^{15}$N-HSQC spectra (Fig. 2.2). In contrast, *E. coli* JC158 (lacI22, λ-, ei4-, relA1, serA6, spoT1, thiE1) (Clark, 1963) is a leaky serine auxotroph, presumably using glycine as a metabolic precursor (Waugh, 1996). Although not quantitated, the resulting BcX appeared only partially enriched in $^{13}$C/$^{15}$N-serine based on relative $^{13}$C- and $^{15}$N-HSQC spectral intensities. Also, due to metabolic interconversion, some $^{15}$N-labeling of the alanine, tryptophan and glycine amides (~ 2%, 9% and 12% relative $^{15}$N-HSQC peak intensities versus those of the serine amides, respectively) and $^{13}$C-labeling of the alanine methyls occurred (Fig. 2.3). It is noteworthy that in preliminary experiments carried out with prototrophic *E. coli* BL21, significant metabolic interconversion of both $^{15}$N-threonine and $^{15}$N-serine occurred as evidenced by the $^{15}$N-HSQC spectra of the resulting proteins (Fig. 2.4). Thus, the advantages of using auxotrophic *E. coli* strains include increased labeling selectivity and minimization of the amount of labeled amino acid necessary to ensure adequate bacterial growth and protein expression (Muchmore et al., 1989).
Figure 2.2 In the (a) $^{15}$N-HSQC and (b) $^{13}$C-HSQC spectra of $^{13}$C/$^{15}$N-Thr labeled BcX, all of the expected amide $^1$H$_{\text{N}}$$^{15}$N and aliphatic $^1$H$_{\alpha}$-$^{13}$C$_{\alpha}$, $^1$H$_{\beta}$-$^{13}$C$_{\beta}$ and $^1$H$_{\gamma 1}$-$^{13}$C$_{\gamma 1}$ peaks from the 25 threonine residues can be observed. Also noteworthy, no additional residues were isotopically enriched by metabolic interconversion when using the threonine auxotroph *E. coli* JW0003-2.
Figure 2.3 (a) The $^{15}$N-HSQC spectrum of $^{13}$C/$^{15}$N-Ser labeled BcX. The expected amide $^1$H$^N$-$^{15}$N signals from all 18 serine residues can be seen as major peaks. Although not quantitated, the level of enrichment appeared low, presumably due to the production of unlabeled serine from glycine by the leaky serine auxotroph E. coli JC158. Due to metabolic interconversion, weak amide signals from glycine, tryptophan, and to a lesser extent, alanine, residues were also observed. (b) The $^{13}$C-HSQC spectrum of $^{13}$C/$^{15}$N-Ser labeled BcX. In addition to the expected major $^1$H$^α$, $^{13}$C$^α$ and $^1$H$^β$,$^{13}$C$^β$ signals from the serine residues, weaker signals from glycine, tryptophan and alanine (i.e., methyls) are observed. A minor buffer contamination yielded the signal with a $^{13}$C shift of ~33 ppm.
Figure 2.4 Auxotrophic E. coli strains facilitate selective labeling of BcX with (a, b) \(^{15}\text{N}/^{13}\text{C}\)-threonine and (c, d) \(^{15}\text{N}/^{13}\text{C}\)-serine. Shown are the \(^{15}\text{N}\)-HSQC spectra of the protein produced using (a) the threonine auxotrophic strain JW0003-2 and (b) the prototrophic strain BL21 grown in media supplemented with \(^{15}\text{N}/^{13}\text{C}\)-threonine and all other unlabeled amino acids. The protein produced with the auxotroph was highly enriched for only threonine. In contrast, many additional mainchain and sidechain amides were labeled by the prototrophic strain due to metabolic interconversion of the \(^{15}\text{N}/^{13}\text{C}\)-threonine. Also shown are the \(^{15}\text{N}\)-HSQC spectra of the protein produced using (a) the leaky serine auxotrophic strain JC158 and (b) the prototrophic strain BL21 grown in media supplemented with \(^{15}\text{N}/^{13}\text{C}\)-serine and all other unlabeled amino acids. Although additional non-serine amides were labeled in both cases, metabolic interconversion was worse with the prototrophic strain.
2.2.3  NMR spectroscopy

NMR spectra were recorded at 25 °C with Bruker Avance III 500, 600, and 850 MHz spectrometers equipped with xyz-gradient TCI cryoprobes. Data were processed with NMRpipe (Delaglio et al., 1995a), and analyzed using Sparky (Goddard and Kneeler, 1999) and Topspin. Signals from the serine and threonine non-hydroxyl $^1$H, $^{13}$C, and $^{15}$N nuclei were assigned based on the previously reported spectra of BcX (Connelly et al., 2000; Plesniak, Wakarchuk, et al., 1996) and confirmed using intra-residue H(C)TOCSY-NH and C(C)TOCSY-NH spectra recorded with the $^{13}$C/$^{15}$N-serine or -threonine labeled proteins (Fig. 2.5). These spectra differed from the more commonly used inter-residue versions as the amides following the serines and threonines were unlabeled. Signals from the serine and threonine hydroxyls were characterized with a variety of single- and multiple-bond (long range) correlation experiments, presented in the Results section. Details of most pulse sequences used in this study are documented as Appendix Fig. A1-A8.
Figure 2.5 An example assignment of the NMR signals from Thr143 (boxed) using $^{13}$C/$^{15}$N-Thr labeled BcX. (a) Peaks from all 25 amides in the $^{15}$N-HSQC spectrum were readily assigned from previously reported data. (b, c) Intra-residue C(C)TOCSY-NH and H(C)TOCSY-NH spectra were used to correlate aliphatic $^{13}$C and $^1$H signals with the amide $^{15}$N and $^1$H$^N$ of each threonine. Shown are planes taken at the $^{15}$N shift of Thr143. (d) Based on these shifts, the methyl $^1$H$^\alpha$-$^{13}$C$^\gamma$ signals of each threonine were then identified in a constant time $^{13}$C-HSQC spectrum. The 3D pulse sequences are presented in Appendix Figures A7 and A8.
2.2.4 $^3J_{CH}$ quantitation

Three-bond scalar couplings for the detected hydroxyls were measured using a quantitative spin-echo difference experiment with a $^{13}$C/$^{15}$N-filtered readout (Appendix Fig. A4). The data were recorded for uniformly $^{13}$C/$^{15}$N-labeled BcX at pH 6.0 and 25 °C using an 850 MHz spectrometer. Difference ($I_{\text{diff}}$) and reference ($I_{\text{ref}}$) peak heights were measured with spin echo Δ delay periods of 21 msec, and used to solve for $^3J_{CH}$ according to the equation 2.1 (Blake et al., 1992)

$$I_{\text{diff}}/I_{\text{ref}} = 1 - \cos(2\pi * ^3J_{CH} \Delta) \quad (2.1)$$

The reported $^3J_{CH}$ are the average values from data recorded with either a broadband $^{13}$C inversion pulse or on-resonance hard pulses during the final spin echo of the pulse sequence.

2.2.5 Hydrogen exchange kinetics

Hydroxyl HX kinetics were measured using a 1D CLEANEX-PM experiment (Hwang et al., 1997) with a $^{13}$C/$^{15}$N-filtered readout (appendix Fig. A5). The data were recorded for uniformly $^{13}$C/$^{15}$N-labeled BcX at pH 6.0 and 25 °C using a 600 MHz spectrometer. Peak heights ($I_i$) were measured as a function of transfer time ($t_i$), and fit with Mathworks Matlab to equation 2.2,

$$\frac{I_i}{I_0} = \left( \frac{k_{\text{ex}}}{k_{\text{ex}} + R_{\text{app}}} \right) \left( 1 - e^{-(k_{\text{ex}} + R_{\text{app}})t_i} \right) \quad (2.2)$$

where $I_0$ is the corresponding signal height in a control spectrum recorded without exchange, $k_{\text{ex}}$ is the exchange rate constant, and $R_{\text{app}}$ is an apparent rate constant due to both transverse and
longitudinal relaxation. Recycle delays were 2 sec, except in the control spectra for which a 12 sec delay was used to ensure full water relaxation. To account for the reduction of steady-state water magnetization by a factor of ~0.7, the reported $k_{ex}$ values were scaled by 1.4 fold (Hwang et al., 1997).

### 2.3 Results and Discussion

#### 2.3.1 Detection of serine and threonine hydroxyl signals

The use of selective isotopic labeling and long-range $^{13}$C-HSQC experiments provides a simple route for detecting and unambiguously assigning slowly exchanging hydroxyl proton signals. In a conventional gradient $^{13}$C-HSQC spectrum recorded with a standard INEPT "1/4J" delay of 1.7 msec, the expected signals from directly-bonded $^1$H-$^{13}$C groups were readily observed in samples of $^{13}$C/$^{15}$N-serine or -threonine labeled BcX. In contrast, spectra recorded with ~ 11 msec delays yielded additional weak crosspeaks due to long-range $^3$J$_{CH}$ couplings between the hydroxyl $^1$H$^\gamma_1$ and methyl $^{13}$C$^\gamma_2$ pairs of three threonines (Fig. 2.6) and the $^1$H$^\gamma$ and $^{13}$C$^\alpha$ pairs of two serines (Fig. 2.7). Importantly, these signals were apparent singlets with or without $^{13}$C- or $^{15}$N-decoupling during acquisition, thus confirming that they indeed corresponded to oxygen-bonded protons. By alignment with the reference one-bond spectra, the hydroxyls were readily assigned to Thr72, Thr145, Thr171, Ser130, and Ser176.
Figure 2.6 (a, b) Signals from three hydroxyl $^1\text{H}^\gamma_1$ nuclei in selectively $^{13}\text{C}/^{15}\text{N}$-threonine labeled BcX were detected in a long-range $^{13}\text{C}$-$^\beta$-decoupled $^{13}\text{C}$-HSQC spectrum that exploits a weak $^3\text{J}_{\text{CH}}$ coupling to the methyl $^{13}\text{C}^\gamma_2$. (c) The signals were then assigned by alignment with a one-bond constant time $^{13}\text{C}$-HSQC spectrum (contoured ~ 2x higher than a and b). To detect the $^3\text{J}_{\text{CH}}$ couplings, while also minimizing signals from directly bonded $^1\text{H}$-$^{13}\text{C}$ groups, an INEPT "1/4J" delay of 11.4 msec was used for the long-range experiment (Appendix Figs. A3 and A4)
Figure 2.7 Signals from two hydroxyl $^1H\gamma$ nuclei in selectively $^{13}C/^{15}N$-serine labeled BcX were detected in a (a, b) long-range $^{13}C$-HSQC spectrum that exploits a weak $J_{CH}$ coupling to the $^{13}C\alpha$, and then assigned by aligning to a (c) one-bond $^{13}C$-HSQC spectrum (contoured ~ 2x higher than a and b). An INEPT "1/4J" delay of 10.5 msec was used for the long-range experiment in order to also minimize signals from directly bonded $^{13}C$-$^1H$ groups (appendix Fig. A2). Note that Ser176 $^1H\gamma$ has a chemical shift of 4.16 ppm, which falls within the range for most serine $^1H\alpha$ and $^1H\beta$ nuclei. Thus, good suppression of the one-bond couplings, as seen in panel b, was important for detecting this signal.
Several details of this strategy, which has been demonstrated for tyrosines (Baturin et al., 2011) and should also work with cysteines, are worth comment. In addition to providing immediate amino acid-specific assignments, selective $^{13}$C- or $^{13}$C/$^{15}$N-labeling is advantageous for several reasons. The multitude of potential one-, two- and three-bond $^1$H-$^{13}$C correlations present in the long-range $^{13}$C-HSQC spectrum of a uniformly enriched protein could easily obscure the desired hydroxyl signal. Unlike the generally downfield shifted phenolic protons of tyrosines, these threonine and serine OH signals fall across the spectral envelope typical of many amide, aromatic, and aliphatic protons. Also, even if clearly identified as an OH signal (i.e., by comparing spectra recorded ± $^{15}$N/$^{13}$C decoupling during acquisition or in H$_2$O versus D$_2$O buffer), chemical shift degeneracy along the $^{13}$C dimension of a long-range $^{13}$C-HSQC spectrum recorded with a uniformly enriched protein might preclude an unambiguous assignment. In contrast, these complications are significantly reduced by using selective $^{13}$C-labeling, thus enabling the confident detection and assignment of signals from only serine or threonine hydroxyls. Of course, labeling with $^{13}$C$^{\alpha}$-Ser or $^{13}$C$^{\gamma2}$-Thr would provide even further spectral simplification, albeit at the cost of obtaining these compounds via synthetic or biosynthetic routes (Velyvis et al., 2012). Also, various deuteration approaches can be envisioned. However, the uniformly $^{13}$C/$^{15}$N-enriched amino acids purchased for this study were obtained from an algal lysate and thus relatively inexpensive. This labeling scheme also facilitated residue-specific assignments using amide $^{15}$N-resolved $^{13}$C-TOCSY approaches. On the other hand, a disadvantage of this scheme is peak broadening in $^{13}$C-HSQC spectra due to $^{13}$C-$^{13}$C couplings. Although the latter can be avoided with a constant-time pulse sequence (Appendix Figs. A2 and A3), the required ~ 26 msec delay resulted in substantial signal loss with the BcX samples (not shown). Fortunately, in the case of threonine, the $^{13}$C$^{\gamma2}$ resonate over 40 ppm upfield from the
$^{13}\text{C}^{\alpha/\beta}$, thus allowing decoupling in $t_1$ with a selective $^{13}\text{C}$ inversion pulse applied at 70 ppm to sharpen the detected $^{13}\text{C}^{\gamma_2}$-$^1\text{H}^{t_1}$ peaks (Appendix Fig. A4).

Even with selective labeling, the hydroxyl protons in BcX have chemical shifts ranging from 2.9 to 7.8 ppm and thus potentially overlap resonances from other serine and threonine protons (e.g., Ser176 in Fig. 2.8). Therefore, it was also critical to use empirically optimized INEPT "1/4J" delays of 10.5 msec for serine and 11.4 msec for threonine in order to minimize signals from directly bonded $^{13}\text{C}$-$^1\text{H}$ pairs (i.e., nulls occurring at $n/(2^1\text{J}_{\text{CH}})$) while favouring long-range couplings. Although the choice of these timings was best for BcX due to a balance of one-bond signal suppression and relaxation versus transfer efficiency for detecting long-range hydroxyl couplings, OH signals were still observed with delays spanning 4 to 32 msec. Parenthetically, another advantage of selective isotopic labeling for the simple long-range $^{13}\text{C}$-HSQC experiment lies with the difficulty in efficiently suppressing the signals from all $^{13}\text{C}$-$^1\text{H}$ pairs in a uniformly labeled protein sample due to the range of possible $^1\text{J}_{\text{CH}}$ couplings.
Figure 2.8 The $^3J_{CH}$ couplings for the hydroxyl protons in uniformly $^{13}C/^{15}N$-labeled BcX were measured using a quantitative spin-echo difference experiment with a $^{13}C/^{15}N$-filtered readout. Shown are the reference (top) and difference (lower) spectra recorded with $\Delta = 21$ msec in the pulse sequence of Appendix Fig A5. The signals from Tyr26 and Thr145 were deconvoluted using Topspin. In addition to the results given in the text for the serines and threonines, apparent $|^{1}J_{C\epsilon H\eta}|$ values were also fit for Tyr105 (7.3 Hz), Tyr53 (6.6 Hz), Tyr79 (7.0 Hz), and Tyr26 (5.6 Hz). As reported previously (Baturin et al., 2011), $|^{2}J_{C\zeta H\eta}|$ are small for tyrosines. Also, due to slow ring flipping and slow rotation about the $C^\zeta-O^\eta$ bond, the hydroxyl $^1H^\eta$ of Tyr79 is constrained by hydrogen bonding to a trans conformation relative to the coupled $^{13}C^\epsilon$. The same holds for Tyr105. In contrast, the aromatic rings of Tyr26 and Tyr53 flip rapidly, giving averaged chemical shifts and, hence, $^3J_{C\epsilon H\eta}$ for their two $^{13}C^\epsilon$ nuclei.
All five hydroxyls detected in the long-range $^{13}$C-HSQC spectra are hydrogen bonded and either partially (Thr145) or completely solvent inaccessible in the static X-ray crystal structure of BcX. This, of course, is expected as these hydroxyls must be well protected from HX in order to be observed via this NMR approach. Of the of the 18 serines and 25 threonines in BcX, several additional residues, including Thr33, Thr67, Thr110, Ser100, and Ser180, are also fully buried and hydrogen bonded, and thus likely yield the remaining unassigned OH peaks in the $^{13}$C/$^{15}$N-filtered spectrum of Fig. 1B. Accordingly, the absence of any corresponding signals in a long-range $^{13}$C-HSQC spectrum is somewhat surprising. Note that care was taken to limit the adverse effects of hydrogen exchange. In particular, the water magnetization was minimally perturbed through use of gradients, rather than selective saturation, and returned to the +z-axis at the end of the $^{13}$C-HSQC pulse sequences (Grzesiek and Bax, 1993). One possible explanation is that some of these serine and threonine residues may have conformationally-restricted hydroxyls with a gauche $\text{C}^{\alpha}/\gamma^2-\text{O}^{-1}/\gamma^1-\text{H}^\gamma/\gamma^1$ dihedral angle and thus relatively small $^{3}J_{\text{C}^{\alpha}-\text{H}^\gamma}$ and $^{3}J_{\text{C}^\gamma^2-\text{H}^\gamma^1}$ couplings, respectively (see below). Alternatively, the hydroxyls may be exchanging slowly enough to be detected in the 1D spectrum (i.e., $k_{\text{ex}} < ^{1}J \sim 100$ s$^{-1}$ for filtering $^{1}$H-$^{13}$C and $^{1}$H-$^{15}$N signals), yet fast enough ($k_{\text{ex}} > ^{3}J_{\text{CH}} \sim 10$ s$^{-1}$) to impair polarization transfer during the INEPT periods of the long-range $^{13}$C-HSQC pulse sequences (Henry and Sykes 1990; Segawa et al. 2008).

A complementary approach to detecting slowly exchanging serine and threonine hydroxyls is to use a $^{13}$C/$^{15}$N-filtered/edited HSQC-NOESY experiment (Ikura and Bax, 1992; Zwahlen et al., 1997). Although typically recorded to detect intermolecular NOEs within a complex of isotopically labeled and unlabeled species, intramolecular NOE signals will also be observed between oxygen-bonded and $^{13}$C/$^{15}$N-bonded protons within a single uniformly labeled
protein. Indeed many such NOEs can be detected for the slowly exchanging serine, threonine, and tyrosine hydroxyls in BcX (Baturin et al., 2011) and not shown). However, their unambiguous assignment proved very difficult due to potential chemical shift degeneracies. This precluded the confident identification of the remaining unassigned OH peaks in the $^{13}$C/$^{15}$N-filtered spectrum of Fig. 2.1b.

### 2.3.2 Coupling constant measurements

It is notable that potential signals from $^{2}J_{CH}$ hydroxyl proton couplings with the $^{13}$C$^{\beta}$ of either amino acid were not detected, nor were those from $^{3}J_{CH}$ couplings between the $^1H^{\gamma1}$ and $^{13}$C$^{\alpha}$ pairs of the threonines. For simple alcohols, $^{2}J_{C-OH}$ couplings of -2 to -3 Hz and dihedral angle-dependent $^{3}J_{C-OH}$ couplings of 2 to 9 Hz (assumed positive) have been reported (Hansen, 1981; Borisov et al., 1998). Thus, the absence of these additional signals likely reflects weak, averaged scalar interactions, for which detection is limited by the relaxation behaviour of this 20 kDa protein.

The expected dihedral angle-dependence of the $^{3}J_{C-OH}$ couplings can also provide structural insights into the hydroxyl conformations. Therefore, I used a quantitative spin-echo difference experiment (Blake et al., 1992) combined with $^{13}$C/$^{15}$N-filtration to measure the coupling constants for the five detected hydroxyls (Fig. 2.8). This yielded $|^{3}J_{C^{\alpha}-H^{\gamma}}|$ values of 6.9 Hz for Ser130 and 7.3 Hz for Ser176. For the threonines, $|^{3}J_{C^{\gamma2}-H^{\gamma1}}|$ values of 6.8 Hz for Thr72, 5.9 Hz for Thr145, and 5.4 Hz for Thr171 were measured. The fitting errors were estimated to be ± 0.5 Hz, and additional possible couplings were assumed to be negligible due to the absence of corresponding signals in long-range $^{13}$C-HSQC spectra. Although the Karplus equation for such $^{3}J_{C-OH}$ couplings has not been parameterized, the measured values suggest that all five hydroxyl
protons are at least moderately restrained within BcX and predominantly in trans conformations relative to the coupled $^{13}$C nucleus. Parenthetically, for each threonine with a trans $C^\gamma_2$-$C^\beta$-$O^\gamma_1$-$H^\gamma_1$ dihedral angle, the corresponding $C^\alpha$-$C^\beta$-$O^\gamma_1$-$H^\gamma_1$ would be in the gauche conformation. The latter would likely have a relatively small $^3J_{C^\alpha-H^\gamma_1}$, thus explaining the lack of $^{13}C^\alpha$-$H^\gamma_1$ crosspeaks for the three threonines detected in long-range $^{13}$C-HSQC spectra. By way of comparison, in a long-range $^{13}$C-HMQC spectrum of a CAP-Gly domain, a slowly exchanging threonine $H^\gamma_1$ with an unusual shift of -0.97 ppm showed a strong correlation to its $^{13}C^\alpha$ and a weak correlation to its $^{13}C^\gamma_2$ (Plevin et al., 2008). This threonine participates in an OH/$\pi$ interaction with an adjacent phenylalanine ring, and indeed has a trans $C^\alpha$-$C^\beta$-$O^\gamma_1$-$H^\gamma_1$ dihedral angle.

2.3.3 Hydrogen exchange kinetics

Serine and threonine hydroxyls undergo both specific and general acid- and base-catalyzed HX. Accordingly, long-range $^{13}$C-HSQC spectra were recorded for the $^{13}$C/$^{15}$N-threonine labeled BcX sample at pH values ranging from 4 to 8 and temperatures from 5 °C to 25 °C. However, no additional hydroxyl signals were detected beyond those found initially at pH 6. This is consistent with the observation that serines and threonines in model peptides exchange most slowly at a pH$_{\text{min}}$ of ~ 6.5 (Liepinsh et al., 1992; Liepinsh and Otting, 1996).

To measure the HX kinetics of the BcX serine and threonine residues, a CLEANEX-PM experiment (Hwang et al., 1997) with $^{13}$C/$^{15}$N-filtered detection was implemented. As shown in Fig. 2.9, a reliable transfer curve was measurable for Thr145, yielding a fit exchange rate constant $k_{\text{ex}}$ of $3.5 \pm 0.6$ s$^{-1}$. Based on the data of Liepinsh and Otting (Liepinsh et al., 1992), a reference $k_{\text{ex}} \sim 500$ s$^{-1}$ can be estimated for a serine or threonine in a random coil peptide under
similar experimental conditions. Thus, the HX rate of Thr145 is retarded by ~ 150 fold due to its structural and electrostatic environment in BcX. Although Ser130 appeared to be exchanging, overlap with neighbouring peaks in combination with low signal intensity made quantitation unreliable. In contrast, CLEANEX transfer was not detected for Thr72, Thr171 and Ser176. Thus, these residues likely have $k_{ex} < 1 \text{ s}^{-1}$ and protection factors > 500. Consistent with these relative HX rate constants, the latter three residues yielded stronger hydroxyl crosspeaks in long-range $^{13}\text{C}$-HSQC spectra than did Thr145 and Ser130 (Figs. 2.6 and 2.7). Conversely, several additional unassigned hydroxyls showed more rapid HX (60 to 110 s$^{-1}$; Fig. 2.8), which likely precluded their detection and assignment in long-range $^{13}\text{C}$-HSQC spectra. Hydroxyl HX rate constants of a similar magnitude have also been measured recently using an elegant approach that exploits highly specific $^{13}\text{C}/^2\text{H}$ serine and threonine labeling along with deuterium isotope shifts (Takeda et al., 2011).
Figure 2.9 The hydroxyl HX rates were measured at pH 6 and 25 °C using a CLEANEX-PM experiment with a $^{13}$C/$^{15}$N-filtered readout. (a) Overlaid spectra with mixing times of 0.42, 0.84, 2, 4, 10, 20, 40, 80, 150 ms (light grey to dark grey) are shown along with a normalized reference spectrum (black). (b) The build-up exchange curve of Thr145 was fit to yield $k_{ex} = 3.5 \pm 0.6$ s$^{-1}$ and $R_{app} = 43 \pm 10$ s$^{-1}$. Errors were estimated via a Monte Carlo approach. Note that Tyr26 does not show detectable exchange, and the reference signal intensity for the adjacent Thr145 was deconvoluted using Topspin. Several additional unassigned hydroxyls showed more rapid HX (A: 60 ± 20 s$^{-1}$; B: 70 ± 20 s$^{-1}$; and C: 110 ± 30 s$^{-1}$), which likely precluded their detection and assignment in long-range $^{13}$C-HSQC spectra. The pulse sequence is summarized in Appendix Fig. A6.
2.3.4 Structural insights

By combining NMR spectroscopic and X-ray crystallographic data, insights into the structural features of several hydroxyls in BcX can be obtained (Fig. 2.10). Based upon their relative positions in the X-ray crystallographic structure of BcX, it is very reasonable to propose that the sidechains of Thr72 and Ser176 both accept a hydrogen bond from a neighboring amide or indole, respectively, and donate a hydrogen bond to a carbonyl oxygen. This proposal is supported by the observed protection of these hydroxyls from rapid HX, and by coupling constant measurements, indicating that each is restrained in a trans $C^{\alpha\gamma_2}-C^{\beta}-O^{\gamma_1}-H^{\gamma_1}$ conformation.

In a cluster of buried polar groups, Thr171 bridges Ser180 and Tyr26. The $O^{\gamma_1}$ threonine appears positioned to both accept a hydrogen bond from the $O^{\gamma}$H of Ser180, and donate one to the phenol of Tyr26, which in turn donates a hydrogen bond to carbonyl of Met169. Although this is consistent with $^3J_{CH}$ measurements, indicating a trans $C^{\gamma_2}-C^{\beta}-O^{\gamma_1}-H^{\gamma_1}$ dihedral angle for Thr171, the hydrogen bond to Tyr26 seems rather long in the crystal structure of BcX. Thus, it is somewhat surprising that the hydroxyl proton of Thr171 is well protected from HX, especially since that of Ser180 was not detected in long-range $^{13}C$-HSQC despite an apparently strong interaction with this threonine. This highlights the difficulty in inferring such energetic and dynamic details from static crystal structures.

Thr145 and Ser130 are involved in a fascinating network of buried polar groups (Joshi et al., 1997; Plesniak, Connelly, et al., 1996). Thr145 must donate a hydrogen bond to the negatively-charged carboxylate of Asp101, which is also linked via a bound water to the N$^{\epsilon_2}$H of neutral His149. On the opposite side of the imidazole ring, Ser130 appears to donate its $O^{\gamma}$H to the N$^{\delta_1}$ of this histidine, while accepting a hydrogen bond from the indole of Trp153. In this
proposed arrangement, both Thr145 and Ser130 have trans $C^{\alpha\gamma_2}-C^{\beta}-O^{\gamma_1}-H^{\gamma_1\gamma_1}$ conformations that are consistent with their measured $^3J_{CH}$ couplings. Although the hydroxyl $H^{\gamma_1}$ of Thr145 is partially solvent exposed in the BcX crystal structure, it is still well protected from HX due to these presumably strong interactions.

It is notable that the five hydroxyl protons in detected BcX have chemical shifts ranging from 2.9 to 7.8 ppm (Fig. 2.1), and thus differ substantially from the average values of 5.47 ± 1.1 ppm (serine $^1H^\gamma$) and 5.47 ± 1.1 ppm (threonine $^1H^{\gamma_1}$) reported in the BioMagResBank (Ulrich et al., 2008). Inspection of the static crystal structure of BcX does not provide any simple consistent explanation for this chemical shift variation in terms of hydrogen bonding or proximity to aromatic rings (Fig. 2.10). One exception to this statement is that, of the five serines and threonines characterized in this study, only Thr145 is involved in an ionic hydrogen bond to negatively-charged acceptor (Asp101), and its $H^{\gamma_1}$ has the most downfield shifted signal at 7.8 ppm.

2.4 Summary

The direct characterization of protein hydroxyl protons is usually a difficult challenge for structural biologists. Using a combination of selective isotope labeling along with various isotope filtering and editing strategies, I have detected and assigned the $^1H$ signals from five slowly exchanging serine and threonine hydroxyls in BcX. These approaches were also exploited to measure $^3J_{CH}$ couplings and HX kinetics and thereby gain structural and dynamic insights into the local environments of these residues within this model protein. These methods should prove useful for further studies of functionally important hydroxyls in a variety of proteins and protein complexes.
Figure 2.10 NMR spectroscopy provides confirmatory insights into the structural features of several BcX serine and threonine hydroxyl protons determined previously by X-ray crystallography. Both (a) Thr72 and (b) Ser176 donate a hydrogen bond to a carbonyl oxygen and accept one from an amide or indole, respectively. (c) Thr171 is a weak hydrogen bond donor to Tyr26 and a strong acceptor for Ser180. (d) Thr145 and Ser130 participate in an extensive hydrogen-bonding network also involving a charged aspartate, a buried neutral histidine, and an internal water. The program Reduce (Word et al., 1999) was used to add hydrogens to the protein coordinate file 1XNB.pdb (oxygen, red; nitrogen, blue; carbon, green; hydrogen, grey). Hydrogen bonds are indicated with yellow dashes.
Chapter 3: The pK\textsubscript{a} values of the catalytic residues in the retaining glycoside
hydrolase T26H mutant of T4 lysozyme

3.1 Overview

T4 phage lysozyme (T4L) is an enzyme that cleaves bacterial cell wall peptidoglycan. Remarkably, the single substitution of the active site Thr26 to a His (T26H) converts T4L from an inverting to a retaining glycoside hydrolase with transglycosylase activity. It has been proposed that T26H-T4L follows a double displacement mechanism with His26 serving as a nucleophile to form a covalent glycosyl-enzyme intermediate (Kuroki et al., 1999). To gain further insights into this or alternative mechanisms, I used NMR spectroscopy to measure the acid dissociation constants (pK\textsubscript{a} values) and/or ionization states of all Asp, Glu, His, and Arg residues in the T4L mutant. Most notably, the pK\textsubscript{a} value of the putative nucleophile His26 is 6.8 ± 0.1, whereas that of the general acid Glu11 is 4.7 ± 0.1. If the proposed mechanism holds true, then T26H-T4L follows a reverse protonation pathway in which only a minor population of free enzyme is in its catalytically competent ionization state with His26 deprotonated and Glu11 protonated. These studies also confirm that all arginines in T26H-T4L, including the active site Arg145, are positively charged under neutral pH conditions.

3.2 Introduction

Lysozymes are glycoside hydrolases that cleave the β-1,4-glycosidic linkage between N-acetylMuramic acid (NAM) and N-acetylglucosamine (NAG) of peptidoglycan, a major component of the bacterial cell wall. These enzymes are classified into at least 5 families by the
Carbohydrate Active Enzyme (CAZy) database based on their amino acid sequences (Lombard et al., 2014; Wohlkonig et al., 2010). Members of a family typically utilize a similar hydrolytic mechanism that leads to either retention or inversion of anomeric carbon stereochemistry at the site of hydrolysis (Davies and Henrissat, 1995; Rye and Withers, 2000).

The lysozyme from Enterobacteria phage T4 (T4L) is a member of the GH24 family. Although T4L is an extremely well-characterized model system for investigating protein folding and dynamics (Baase et al., 2010), detailed enzymology studies have proven difficult as it acts on a complex glycopeptide substrate (Jensen et al., 1976) for which analogs amenable to rigorous kinetic measurements are not readily available. In 1995, Matthews and co-workers reported that T4L is an inverting glycoside hydrolase (Kuroki et al., 1995). This conclusion was based in part on the observation that hydrolysis of a peptidic tetrasaccharide by T4L yielded a disaccharide product with the opposite anomeric stereochemistry to that generated by hen egg white lysozyme (HEWL), a classic GH22 retaining hydrolase (Vocadlo et al., 2001). Accordingly, it was proposed that T4L catalyzes a single-displacement inverting reaction in which a nucleophilic water, activated by the general base Asp20 (with a pKₐ value of 3.6 (Anderson, 1992; Anderson et al., 1993)), attacks the β-1,4-linked anomeric C1 carbon of the peptidic NAM, while the general acid Glu11 (pKₐ 5.4) concomitantly donates a proton to the leaving aglycone, NAG.

In the X-ray crystallographic structures of wild-type T4L, a candidate nucleophilic water is hydrogen bonded between the sidechains of Asp20 and Thr26. As part of an earlier attempt to engineer metal binding sites into T4L, the Matthews group introduced a Glu at position 26. Serendipitously, the T26E mutant was isolated from E. coli with Glu26 covalently bound to a peptidic NAG-NAM disaccharide via an α-linkage to the anomeric carbon of a distorted NAM moiety (Kuroki et al., 1993). This stable species appeared to be analogous to the glycosyl-
enzyme intermediate formed in the first step of a double-displacement retaining mechanism. Subsequently, it was found that a T26H mutant is catalytically active, yet produces a hydrolysis product with the opposite anomeric stereochemistry to that yielded by wild-type T4L (Kuroki et al., 1995). Remarkably, a single amino acid substitution at position 26 changes T4L from an inverting to a retaining glycoside hydrolase. Furthermore, unlike the wild-type enzyme, T26H-T4L also functions as a transglycosylase (Kuroki et al., 1999).

Based on mutational and X-ray crystallographic studies, a two-step double-displacement mechanism was proposed for T26H-T4L (Kuroki et al., 1995, 1999) (Fig. 3.1a). In the glycosylation step, His26 is postulated to be a nucleophile, forming an α-linked glycosyl-enzyme intermediate. This is facilitated by Glu11 serving as a general acid to protonate the leaving aglycone. In the subsequent deglycosylation step, Glu11 functions as a general base to activate either a water (hydrolysis) or carbohydrate (transglycosylation) for nucleophilic displacement of His26 and release of the glycone with a net retention of β-anomeric stereochemistry.
Figure 3.1 (a) Proposed double-displacement retaining mechanism for the hydrolysis reaction catalyzed by T26H-T4L (Kuroki et al., 1995, 1999). (R1 and R2 correspond to NAG; R3 to peptide-linked lactyl substituent of NAM). A minor population of the enzyme is shown with the nucleophile His26 and the general acid/base Glu11 in their catalytically competent protonation states. (b) An alternative mechanism involving substrate-assisted catalysis. Asp20 is shown as a general base, but could also electrostatically stabilize a charged oxazolinium intermediate.
An important aspect of understanding the proposed mechanism, or any alternative, for T26H-T4L lies with defining the protonation states of catalytic residues along the reaction pathway. To this end, Kuroki and co-workers (Hiromoto et al., 2017) used room temperature neutron diffraction to characterize a sample of perdeuterated T26H-T4L*, crystallized in D$_2$O buffer (pD 7.0). In the reported 2.09 Å resolution structure (5XPF.pdb), His26 is predominantly in its neutral state and the carboxyl of Glu11 partially deuterated. This is consistent with the scheme of Fig. 3.1a. However, Asp20 and nearby Arg145 were also modeled to be neutral. This is perplexing since there are no obvious physicochemical reasons why these active site residues should have such highly perturbed pK$_a$ values and unusual sidechain deuteration states at pD ~ 7. Indeed, in the 2.2 Å resolution neutron crystallographic structure of the perdeuterated wild-type enzyme at cryogenic temperatures (pD 6 - 7, 5VNQ.pdb), all ionizable residues are in their charged forms. (pD 6 - 7, 5VNQ.pdb), all ionizable residues are in their charged states (Li et al., 2017). Inspired by these reports, I used NMR spectroscopy to define the protonation states and, when possible, pK$_a$ values of the His, Asp, Glu, and Arg residues in T26H-T4L. These measurements provide key insights into the catalytic mechanism and electrostatic properties of this mutant retaining glycoside hydrolase.
3.3 Materials and Methods

3.3.1 Protein expression and purification

The clones encoding T4L* (plasmid 18111) and T26H-T4L* (plasmid 18251) were obtained from AddGene. T4L* is a cysteine-free pseudo-wild-type lysozyme with the C54T and C97A substitutions. Protein expression and purification followed previously described protocols (Vallurupalli et al., 2009). Briefly, $^{13}\text{C}/^{15}\text{N}$- or $^{15}\text{N}$-labeled lysozyme was expressed in transformed prototrophic E. coli HMS174 cells. An LB seed culture was used to inoculate M9 media containing 3 g/L 99 % $^{13}\text{C}_6$-glucose and/or 1 g/L 98 % $^{15}\text{N}$H$_4$Cl to an initial OD$_{600}$ of ~ 0.1. Expression was induced with 1 mM IPTG at OD$_{600}$ ~ 0.6, and the cells were harvested after ~ 4 hours of growth at 37 °C by centrifugation.

Samples of T26H-T4L* selectively labeled with $^{13}\text{C}^\gamma$-aspartic acid or $^{13}\text{C}^\delta$-glutamic acid were expressed in auxotrophic E. coli strains EA1 (aspC, tyrB, asnAB::Tn5) and DL39 (aspC, ilvE, tyrB, avtA::Tn5), respectively (Waugh, 1996). Overnight seed cultures in LB were used to inoculate minimal media containing 500 mg/L $^{13}\text{C}^\gamma$-D/L-aspartic acid (Cambridge Isotopes) or 800 mg/L $^{13}\text{C}^\delta$-D/L-glutamic acid (Cambridge Isotopes), along with all other unlabeled amino acids (Muchmore et al., 1989; McIntosh et al., 1990; Anderson et al., 1993). Cells were grown to an OD$_{600}$ of ~ 0.6, induced with 1 mM IPTG, and harvested by centrifugation after growth for 4 hours at 37 °C.

Cells were lysed by homogenization, loaded onto a SP-Sepharose cation exchange column (GE Healthcare) in 50 mM sodium phosphate, 2 mM EDTA, pH 6.5, and eluted using a linear gradient of NaCl from 0 to 1 M. Samples were further purified by passage through a Superdex 75 size exclusion column in NMR buffer (100 mM KCl, 30 mM potassium phosphate,
pH 5.5), and concentrated to ~ 1.5 mM using a 3 kDa MWCO Amicon ultracentrifugal filter. Finally, ~ 5 % D₂O was added as a lock solvent. Protein concentrations were determined using ultraviolet absorbance (predicted ε₂₈₀ = 25440 M⁻¹ cm⁻¹).

3.3.2 NMR Spectroscopy

NMR spectra were recorded at 25 °C with Bruker Avance III 500, 600, and 850 MHz spectrometers equipped with xyz-gradient TCI cryoprobes. Data were processed with NMRpipe (Delaglio et al., 1995a, 1995b) and analyzed in NMRFAM-SPARKY (Lee et al., 2015) and Topspin. Standard heteronuclear scalar correlation experiments, including 3D HNCACB, CBCACONH, H( CC O)-TOCSY-NH and C( CC O)-TOCSY-NH (Sattler et al., 1999), were automatically interpreted using PINE (Bahrami et al., 2009) and verified manually to assign relevant backbone and sidechain ¹H, ¹³C, and ¹⁵N signals. The sidechain carboxyl signals of Asp/Glu were assigned from a 3D Hβ²/γ²/ζ-Cβ/γ/δ CO/ζ experiment, which was based on a modified HCACO experiment (Kay et al., 1989). Carboxyl deuterium isotope shifts were measured from 2D Hβ²/γ²( Cβ/γ) CO/ζ/δ spectra of uniformly ¹⁵N/¹³C-labeled T26H-T4L* in H₂O NMR sample buffer and then after lyophilization and resuspension in an equal volume of D₂O. The initial sample pH was 5.8, and the final sample pH* (uncorrected for isotope effects) was 5.7. Histidine imidazole signals were readily identified in conventional and constant time ¹³C-HSQC spectra (Plesniak, Connelly, et al., 1996; Vuister and Bax, 1992). Arginine ¹⁵Nε signals were assigned with HνNεCζ spectra recorded using a modified HNCA experiment (Kay et al., 1990), ¹³Cζ signals with HeNeCζ spectra, recorded using a modified HNCO experiment (Kay et al., 1990), and ¹⁵Νη signals with a cross polarization HνNνCζ experiment (Yoshimura et al., 2017).
3.3.3 pKₐ Determination of His, Asp, and Glu

NMR-monitored pH titrations were recorded on a Bruker Avance III 850 MHz spectrometer. The titrations were carried out by transferring the sample between the NMR tube and a 1.5 mL Eppendorf tube and adding small aliquots of NMR buffer containing 0.1 M HCl or NaOH. The average pH meter reading, measured before and after spectra acquisition with a ThermoFisher Orion 9110DJWP electrode at room temperature (~ 20 °C), was taken as the sample pH value. Histidines in ¹³C/¹⁵N-labeled T26H-T4L* were monitored between pH 5.0 and 10.7 via ¹³C-HSQC spectra. The sidechain carboxyl signals of Asp and Glu in selectively labeled protein samples were monitored between pH 1.8 to 7.8 and 1.8 to 9.8, respectively, via direct detection 1D ¹³C-NMR with ¹H decoupling during acquisition. The pH-dependent chemical shifts of individual ¹³C and ¹H nuclei were fit with GraphPad Prism to the Henderson-Hasselbalch equation for one or macroscopic ionization equilibria to obtain corresponding pKₐ values (McIntosh et al., 2011a). In the case of each His, results from four reporter nuclei were averaged to obtain the reported pKₐ values with standard deviations. In the case of Asp and Glu, errors were estimated ± 0.1 as limited by the accuracy in measuring the sample pH value.
3.4 Results

T26H-T4L* yielded high quality NMR spectra, and thus assignment of the $^1$H, $^{13}$C, and $^{15}$N signals from mainchain and selected sidechain nuclei was straightforward using standard heteronuclear correlation experiments (Fig. 3.2). A comparison of the $^{15}$N-HSQC spectrum of this mutant with that of pseudo-wild-type T4L* revealed moderate $^1$H$^N$ and $^{15}$N chemical shift differences for many corresponding amides (Fig. 3.3). When mapped onto the structure of lysozyme, these amides cluster in the N-terminal lobe around position 26. The crystallographic coordinates of T4L* (1CX6.pdb) and T26H-T4L* (1QT8.pdb) superimpose closely (0.28 Å rmsd for all atoms), and thus these chemical shift differences likely arise from small conformational changes due to the amino acid substitution, combined with electric field and ring current effects from the charged, aromatic imidazolium sidechain of histidine. These two coordinate files were used to provide the inter-residue distances cited below.
Figure 3.2 The $^{15}$N-HSQC spectrum of T26H-T4L* in 100 mM KCl, 30 mM potassium phosphate, and 5% D$_2$O at pH 5.5 and 25 °C. Assigned mainchain amide $^1$H-$^{15}$N signals are labeled.
Figure 3.3(a) Overlaid $^{15}$N-HSQC spectra of T26H-T4L* (red) and wild-type T4L* (blue) with both samples at pH 5.5 and 25 °C. Signals from amides, as well as arginine sidechains (~ 85 ppm in $^{15}$N), are shown. (b) Histogram of amide chemical shift differences between corresponding amides in the two proteins, calculated as $\Delta \delta = (\Delta \delta_H)^2 + (0.16 \Delta \delta_N)^2$ $^{1/2}$. Blank values are for prolines and residues for which assignments were not obtained. (c) When mapped on the structure of T26H-T4L* (1QT8.pdb), amides with different chemical shifts cluster around position 26 (*). The largest shift perturbation (4.1 ppm) occurs for the amide of Ile29.
3.4.1 Histidine $pK_a$ and tautomer state determination

The signals from the two histidines in uniformly $^{13}$C/$^{15}$N-labeled T26H-T4L* were monitored via $^{13}$C-HSQC spectra as the protein was titrated between pH 5 and 11 (Fig. 3.4a and Fig. 3.5). Fitting the pH-dependent chemical shifts of the ring $^{13}$C$^{\varepsilon_1}$, $^1$H$^{\varepsilon_1}$, $^{13}$C$^{\delta_2}$ and $^1$H$^{\delta_2}$ reporter nuclei yielded average $pK_a$ values of 6.8 ± 0.1 for His26 and 8.9 ± 0.1 for His31 (Fig. 3.4b and Table I). Furthermore, with highly diagnostic $^{13}$C$^{\delta_2}$ chemical shifts ~ 127 ppm, the neutral forms of both histidines can be confidently assigned to the less common N$^{\delta_1}$H tautomer (Sudmeier et al., 2003; Platzer et al., 2014).

These results are consistent with the previously reported $pK_a$ value of 9.1 for His31 in wild-type T4L (Anderson et al., 1990), and the X-ray crystallographic structure of the protein with this histidine donating a hydrogen bond from its ring N$^{\delta_1}$ to the sidechain carboxylate of Asp70 ($pK_a < 0.8$, see below). Similarly, the structure of T26H-T4L* shows that His26 donates a hydrogen bond from its N$^{\delta_1}$ to the buried mainchain carbonyl oxygen of Tyr24, leaving its N$^{\varepsilon_2}$ solvent exposed to aid catalysis. Presumably both hydrogen bonds are maintained when His26 and His31 are N$^{\varepsilon_2}$-deprotonated under alkaline conditions.

The $pK_a$ value of an ionizable functional group in a folded protein reflects the pH-dependent free energy difference between states with that group in its neutral versus charged forms, and hence a complex interplay of entropic and enthalpic terms, including desolvation, hydrogen bonding and electrostatic interactions with surrounding charged and polar moieties. Dissecting the relative contributions of these terms is challenging experimentally and computationally.(Bosshard et al., 2004) Nevertheless, it is well accepted that the highly elevated $pK_a$ value of His31, compared to that of ~ 6.5 for a histidine in a random coil polypeptide
(Platzer et al., 2014), arises in large part from its formation of a very favorable salt-bridge (i.e. a hydrogen-bonded ion pair) with ionized Asp70. (Anderson et al., 1990) This conclusion is supported by a comparison of the pH-dependent stability against thermal denaturation of the wild-type protein versus variants with H31N and/or D70N substitutions. (Anderson et al., 1990)

In contrast, the relatively unperturbed pKₐ value of His26 may reflect a balance of unfavorable desolvation of the partially buried imidazolium sidechain and favorable electrostatic interactions with nearby Glu11 (~ 6 Å) and Asp20 (~ 3.5 Å). Based on their pKₐ values presented below, these latter residues are both predominantly negatively charged around the midpoint pH 6.8 for the deprotonation of positively charged His26.
Figure 3.4(a) Superimposed $^{15}$N-decoupled $^{13}$C-HSQC spectra of uniformly $^{13}$C/$^{15}$N-labeled T26H-T4L* showing the $^{1}$H$^{e1}$-$^{13}$C$^{e1}$ signals of His26 and His31 as the protein was titrated between pH 5 and 11 at 25 °C (100 mM KCl, 30 mM potassium phosphate, 5 % D$_2$O). Spectra for the $^{1}$H$^{δ2}$-$^{13}$C$^{δ2}$ signals are provided in Fig. 3.5 (b) Fit titration curves for the $^{13}$C$^{e1}$, $^{1}$H$^{e1}$, $^{13}$C$^{δ2}$ and $^{1}$H$^{δ2}$ nuclei yielded average pK$_{a}$ values of 6.8 ± 0.1 for His26 and 8.9 ± 0.1 for His31 (Table I). The biphasic titration curve for the $^{13}$C$^{e1}$ of His31 has a second fit pK$_{a}$ ~ 6.3 for a small chemical shift change of -0.2 ppm with increasing sample pH that can be attributed to conformational or electrostatic perturbations (McIntosh et al., 2011a) upon deprotonation of nearby (~ 8 Å) His26. The structures of the imidazolium and neutral imidazole N$^{δ1}$H tautomer are shown.
Figure 3.5 Superimposed $^{15}$N-decoupled $^{13}$C-HSQC spectra of uniformly $^{13}$C/$^{15}$N-labeled T26H-T4L* showing the $^1$H$^{\delta_2}$-$^1^3$C$^{\delta_2}$ signals of (a) His26 and (b) His31 as the protein was titrated between pH 5 and 11 at 25 °C. The $^{13}$C$^{\delta_2}$ signals are doublets due to $^1$J$_{CC}$ coupling with the adjacent $^{13}$C'. For clarity, only representative titration points are shown, and all data are presented in Fig. 3.4b. Both neutral histidines adopt the N$^{51}$H tautomer (structure shown) based on fit plateau $^{13}$C$^{\delta_2}$ chemical shifts of 126.7 ppm (His26) and 127.2 ppm (His31).
3.4.2 Aspartic acid pK\(_a\) determination

A sample of T26H-T4L* selectively labeled with \(^{13}\)C\(\gamma\)-aspartic acid was titrated between pH 1.8 and 7.8. The pH-dependent chemical shifts of the \(^{13}\)C\(\gamma\) nuclei were monitored via 1D \(^{13}\)C-NMR spectra (Fig. 3.6a). The Asp \(^{13}\)C\(\gamma\) and Glu \(^{13}\)C\(\delta\) are part of the carboxylic acid functional group and thus generally reliable reporter nuclei for pK\(_a\) measurements by NMR spectroscopy. (Platzer et al., 2014) Of the ten Asp residues, seven yielded well-defined titration curves with downfield chemical shift changes of ~ 3.4 ppm upon increasing sample pH (Fig. 3.6c). Such spectral changes are diagnostic of sidechain deprotonation (Platzer et al., 2014). Fitting the titration curves yielded the pK\(_a\) values summarized in Table I. These values match well those reported previously and discussed in detail for the wild-type protein (Anderson et al., 1993).

**Figure 3.6** Stacked \(^{13}\)C-NMR spectra of T26H-T4L* selectively labeled with (a) \(^{13}\)C\(\gamma\)-Asp and (b) \(^{13}\)C\(\delta\)-Glu, recorded as a function of sample pH at 25 °C (100 mM KCl, 30 mM potassium phosphate, 5 % D\(_2\)O). Due to metabolic interconversion, glutamine (Q) sidechains were also labeled in the latter sample. The titration curves were fit to obtain the (c) Asp and (d) Glu pK\(_a\) values reported in Table I. The biphasic titration curve for Asp20 follows one pK\(_a\) ~ 1.5 with a restrained \(^{13}\)C\(\gamma\) chemical shift change of 3.4 ppm upon increasing sample pH that is attributed to its own ionization and a second small upfield chemical shift change of -0.5 ppm with pK\(_a\) ~ 6.6 that likely reflects structural or electrostatic perturbations (McIntosh et al., 2011a) due to the deprotonation of nearby (~ 3.5 Å) His26. Similarly, the titration curve for Glu11 follows a predominant \(^{13}\)C\(\delta\) chemical shift change of 2.7 ppm with pK\(_a\) 4.7 assigned to its own ionization and a smaller downfield change of 0.5 ppm with pK\(_a\) ~ 7 that also likely reflects the titration of His26 (~ 6 Å).
In contrast, the $^{13}$C\textsubscript{\gamma} chemical shifts of Asp10 and Asp70 did not change markedly over the pH range studied, and that of Asp20 appeared to follow a titration with $pK_a < 2$. However, in the absence of a complete titration curve, it is difficult to infer the ionization state of an Asp residue based on chemical shift alone. For example, Asp10 has a $^{13}$C\textsubscript{\gamma} chemical shift indicative of a protonated carboxylic acid, whereas that of Asp70 corresponds to a deprotonated carboxylate. An alternative approach for distinguishing a carboxylic acid from a carboxylate lies with detecting the presence or absence, respectively, of a 2-bond isotope shift of ~ 0.23 ppm for the $^{13}$C\textsubscript{\gamma} due to a proton versus deuteron on the adjacent oxygen (Ladner et al., 1975; Led and Petersen, 1979; Wang et al., 1996). As shown in Figure 3.7a, none of the ten Asp residues exhibited a $^{13}$C\textsubscript{\gamma} chemical shift change of this magnitude when T26H-T4L* was transferred from H\textsubscript{2}O at pH 5.8 to D\textsubscript{2}O at pH* 5.7. This confirms that all are ionized under these conditions. Furthermore, since no NMR spectroscopic evidence for their protonation was seen even at pH 1.8 in $^{13}$C-NMR monitored titrations, Asp10 and Asp70 are strong acids with $pK_a$ values < 0.8 (i.e., from the Henderson-Hasselbalch equation for a simple acid dissociation equilibrium, at pH $= pK_a + 1$, the population ratio of conjugate base to acid is 0.1, which should be detectable). The same conclusions were previously reported for wild-type T4L (Table 1), as evidenced by the lack of any NMR-detectable titrations for Asp10 and Asp70 down to pH 1.5, combined with a careful study of the pH-dependent stability of the protein under highly acidic conditions (Anderson et al., 1990; Anderson et al., 1993) These highly reduced $pK_a$ values, compared to that of ~ 4 for an aspartic acid in a random coil polypeptide (Platzer et al., 2014) can be rationalized at least in part by the salt-bridges between Asp10-Arg148 and Asp70-His31 observed by X-ray crystallography.
Most important for this study, I conclude that Asp20 has a $pK_a \sim 1.5$ in T26H-T4L* (obtained by fitting the partial titration curve with a restrained chemical shift change of 3.4 ppm due to deprotonation). This is substantially lower than the $pK_a$ of 3.6 reported for this residue in the wild-type protein (Anderson et al., 1993) and may at least in part be due to the presence of a positively charged histidine versus neutral threonine at the nearby ($\sim 3.5 \, \AA$) position 26. Parenthetically, in wild-type T4L, Asp20 shows an anomalously small $^{13}C_\gamma$ chemical shift change over the pH range examined (Anderson et al., 1993). Although this limits the confidence that can be placed on assigning the fit $pK_a$ value of 3.6 to the deprotonation of this catalytic residue, no other titrations were detected $^{13}C_\gamma$ of Asp20 between pH 1.5 and 5.5.

**Figure 3.7** Selected regions of the 2D $H^{\beta\gamma} (C^{\beta\gamma})^{\gamma\delta}$ spectra of uniformly $^{15}N/^{13}C$-labeled T26H-T4L* in H$_2$O (blue; pH 5.8) and D$_2$O (red; pH* 5.7) NMR sample buffer. The spectra show correlations between (a) the $H^{\beta\beta}$ and $^{13}C_\gamma$ of Asp and (b) the $H^{\gamma\gamma}$ and $^{13}C_\delta$ of Glu. The vertical insets are the 1D $^{13}C$-NMR spectra of protein selectively labeled with (a) $^{13}C_\gamma$-Asp at pH 5.9 or (b) $^{13}C_\delta$-Glu at pH 5.9 (taken from Fig. 3.6). The inset scale bar shows the expected $\sim 0.23$ ppm isotope shift for a neutral carboxylic acid ($^{15}COOH$ versus $^{13}COOD$). The lack of any significant isotope shifts demonstrates that the Asp (including Asp10, not shown) and Glu residues in T26H-T4L* are predominantly ionized under these conditions. The small shift of 0.07 ppm exhibited by Glu11 ($pK_a$ 4.7) is attributed to a change in its fractional ionization due to the slightly different pH/pH* conditions. The slight "reversed" isotope shifts of Asp70 and Asp92 may result from protonation versus deuteration of their hydrogen bonded partners (Wang et al., 1996; Guo et al., 2012). Also seen in (a) are $^{15}N$-coupled signals from several Gln sidechains, which show expected (Ladner et al., 1975; Liu et al., 2008) deuterium isotope shifts ($^{13}CONH_2$ versus $^{13}COND_2$) of $\sim 0.12$ ppm. Signals from Asp10 and the Asn sidechains are outside of the presented $^{13}C$ chemical shift windows.
3.4.3 Glutamic acid pKₐ determination

A sample of T26H-T4L* selectively labeled with δ-¹³C glutamic acid was also titrated between pH 1.8 and 9.3 (Fig. 3.6b,d). In ¹³C-NMR spectra, all eight Glu residues showed clear titration curves with downfield ¹³Cᵣ chemical shift changes of ~ 3.8 ppm diagnostic of deprotonation with increasing sample pH (Fig. 3.6b,d). Sidechain ionization was confirmed by the absence of a significant deuterium isotope shift for each residue in H₂O at pH 5.8 versus D₂O at pH* 5.7 (Fig. 3.7b). Fitting the titration curves yielded the pKₐ values summarized in Table I. These results also generally match those reported previously for wild type T4L (Anderson et al., 1993), with Glu62 having the lowest pKₐ value of 2.9, likely due in part to its close proximity to Arg52 (3.2 Å).

Glu11 has a pKₐ of 4.7 in T26H-T4L*. This is only marginally higher than the pKₐ of ~ 4.4 for an unperturbed glutamic acid in a random coil polypeptide (Platzer et al., 2014), yet lower than that of 5.4 measured for the wild-type protein (Anderson et al., 1993). Favorable electrostatic interactions with nearby (~ 6 Å) His26 in the T26H mutant may cause this pKₐ reduction. However, as noted by Andersen et al. (Anderson et al., 1993), the elevated pKₐ value of Glu11 in wild-type T4L is somewhat unexpected given its hydrogen bonding with Arg145, albeit offset by charge repulsion from Asp10 (~ 7 Å) and Asp20 (~ 7 Å). These authors suggested that conformational dynamics of T4L in solution, including a hinge bending motion between its N- and C-lobes, may alter the solvent accessibility and electrostatic environment of active site residues relative to those seen in static crystal structures, thus precluding a simple rationalization of pKₐ values.
Table 3.1. *pKa* values of T26H-T4L*

<table>
<thead>
<tr>
<th>Residue</th>
<th>T26H-T4L*</th>
<th>T4L a</th>
</tr>
</thead>
<tbody>
<tr>
<td>His26</td>
<td>6.8 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>His31</td>
<td>8.9 ± 0.1</td>
<td>9.1</td>
</tr>
<tr>
<td>Asp10</td>
<td>&lt; 0.8 b</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Asp20</td>
<td>~ 1.5 c d</td>
<td>3.6</td>
</tr>
<tr>
<td>Asp47</td>
<td>3.1 ± 0.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Asp61</td>
<td>3.6 ± 0.1</td>
<td>3.6</td>
</tr>
<tr>
<td>Asp70</td>
<td>&lt; 0.8 b</td>
<td>&lt; 0.5</td>
</tr>
<tr>
<td>Asp72</td>
<td>3.3 ± 0.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Asp89</td>
<td>4.0 ± 0.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Asp92</td>
<td>2.7 ± 0.1</td>
<td>2.5</td>
</tr>
<tr>
<td>Asp127</td>
<td>3.5 ± 0.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Asp159</td>
<td>3.4 ± 0.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Glu5</td>
<td>3.7 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Glu11</td>
<td>4.7 ± 0.1</td>
<td>5.4</td>
</tr>
<tr>
<td>Glu22</td>
<td>4.3 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Glu45</td>
<td>3.8 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Glu62</td>
<td>2.9 ± 0.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Glu64</td>
<td>4.4 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Glu108</td>
<td>4.2 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Glu128</td>
<td>4.4 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>Arg95</td>
<td>&gt; 11 d</td>
<td></td>
</tr>
<tr>
<td>Arg145</td>
<td>&gt; 11 d</td>
<td></td>
</tr>
<tr>
<td>Arg148</td>
<td>&gt; 11 d</td>
<td></td>
</tr>
</tbody>
</table>

a Values measured at 10 °C as reported by Anderson et al. (Anderson et al., 1990; Anderson, 1992; Anderson et al., 1993)
b Estimated upper limit as no 13C chemical shift changes due to protonation were seen at the titration endpoint of pH 1.8 (Fig. 3.6). c Fit with a restrained chemical shift change of 3.4 ppm upon deprotonation.
d Estimated lower limit as no 15N chemical shift changes due to deprotonation were seen at the titration endpoint of pH 9.9 (Fig. 3.8).
3.4.4 All arginines are $N^\varepsilon$ protonated at neutral pH

I also sought to determine the protonation states of the arginines in T26H-T4L*. The sidechain $^{15}N^\varepsilon$-$^1H^\varepsilon$ signals from all thirteen arginines in both the wild-type and mutant protein were readily detected and assigned in $^{15}N$-HSQC spectra recorded at pH 5.9 (Fig. 3.8). This unequivocally demonstrated that under these conditions, all, including Arg145, are protonated at the $N^\varepsilon$ position. However, it is possible that an arginine could be neutral due to loss of a terminal $H^\eta$ proton. Therefore, I assigned the arginine $^{13}C^\zeta$ and $^{15}N^\eta$ signals in T26H-T4L* (Fig. 3.8). Although caution must be exercised in inferring charge state from chemical shifts, all thirteen have $^{15}N^\varepsilon$, $^{13}C^\zeta$, and $^{15}N^\eta$ signals are highly indicative of a fully protonated guanidinium sidechain (e.g. 84.8 ppm, 159.5 ppm, and 75.0 and ~ 71 ppm, respectively for Arg145 (Platzer et al., 2014; Fitch et al., 2015)). Furthermore, the $^{13}C$- and $^{15}N$-NMR signals of corresponding arginines in the T26H-T4L* mutant and in the wild-type protein are very similar. (Werbeck et al., 2013; Gerecht et al., 2017; Mackenzie and Hansen, 2017) Of these, several arginines, including Arg145, in T4L at pH 5.5 have been unambiguously demonstrated to be fully protonated and hence positively charged on the basis of $^1H^\eta$-$^{15}N^\eta$ scalar coupling patterns (Yoshimura et al., 2017).
Figure 3.8 Assignment of the signals from the guanidinium sidechain of Arg145 in uniformly $^{13}$C/$^{15}$N-labeled T26H-T4L* at pH 6.5 and 25 °C. (a) The $^1$H$^{13}$C plane from a 3D C(CCO)-TOCSY-NH spectrum taken at the $^{15}$N shift of Ala146 (119.2 ppm). This provided assignments for the $^{13}$Cα (58.8 ppm), $^{13}$Cβ (31.2 ppm), $^{13}$Cγ (25.5 ppm) and $^{13}$Cδ (44.8 ppm) of Arg145. (b) The $^1$H$^{13}$C plane at the indicated $^{15}$N shift (84.8 ppm) of a 3D Hε Nε Cd spectrum yielded the assignment of the $^1$Hε (6.86 ppm) and $^{15}$Nε (84.8 ppm) signals of Arg145 by virtue of a scalar correlation to the $^{13}$Cδ (see also Fig. 3.9). (c) The $^{13}$Cζ (159.5 ppm) was then assigned from the corresponding $^1$H$^{13}$C plane of a 3D Hε Nε Cζ spectrum. (d, e) Finally, $^{15}$Nη signals at 75 ppm and ~71 ppm were identified from a 2D Nε/η-Cζ correlation spectrum. Panels d and e are from the same $^{13}$C-detected spectrum, but e is displayed at a higher contour level for clarity. These assignments are similar to those reported for wild-type T4L*. (f) The cartoon summarizes the expected $^{15}$Nη, $^{13}$Cζ, and $^{15}$N chemical shifts of an arginine sidechain upon titration from its positively charged to neutral form. Under these conditions (pH 6.5), all thirteen arginines in T26H-T4L* have chemical shifts diagnostic of a fully protonated guanidinium group.
I also attempted to measure the pK_a values of the arginines in T26H-T4L* using 2D H^δ_2(C^ő)N^ε-type experiments (Andre et al., 2007). However, this was not successful as spectral crowding precluded the confident measurement of pH-dependent \(^{15}\text{N}^\varepsilon\) chemical shifts. In principle, these reporter shifts, as well as those of the \(^{13}\text{C}^\zeta\) nuclei, could be obtained from a N^δ/n–C^ζ correlation experiment, (Yoshimura et al., 2017) although this was not pursued due to the low sensitivity of such a \(^{13}\text{C}\)-detected approach and signal overlap in the resulting 2D spectra (Fig. 3.8). Moreover, the unperturbed pK_a value of an arginine in a random coil polypeptide is ~13.8 (Platzer et al., 2014; Fitch et al., 2015), and it is unlikely that lysozyme would remain folded, or even chemically intact, under the harsh alkaline conditions required to measure complete titration curves. Therefore, I simply recorded \(^{15}\text{N}\)-HSQC spectra of T26H-T4L* as a function of sample pH value (Fig. 3.9c-e). At pH 9.9, the \(^{15}\text{N}^\varepsilon\text{-}^{1}\text{H}^\varepsilon\) signals of all but three highly protected arginines (Arg95, Arg145, and Arg148) disappeared due to rapid base-catalyzed hydrogen exchange with water (Henry and Sykes, 1995). At pH 10.75, only Arg95 was observed. This protection, which matches the trends in arginine exchange rate constants recently measured for wild-type T4L* (Mackenzie and Flemming Hansen, 2018), can be attributed at least in part to hydrogen bonds between Arg95 with the mainchain oxygen of Phe153, Arg145 with the sidechain oxygen of Asn101, and Arg148 with the carboxylate of Asp10. Importantly, no significant changes in the \(^{15}\text{N}^\varepsilon\) chemical shifts of these arginines were observed over the pH range examined. Since the \(^{15}\text{N}^\varepsilon\) signal of an arginine moves downfield by ~ 6 ppm upon deprotonation (Platzer et al., 2014; Fitch et al., 2015), this implies that the pK_a value of Arg145, as well as Arg95 and Arg148, must be > 11 and likely higher.
Figure 3.8 $^{15}$N-HSQC spectra showing the arginine guanidinium $^1H^\varepsilon - ^{15}N^\varepsilon$ signals from uniformly $^{15}$N/$^{13}$C-labeled (a) wild-type T4L* and (b) T26H- T4L* at pH 5.9. The data unambiguously demonstrate that all thirteen arginines, including Arg145, are $N^\varepsilon$-protonated in both proteins. Superimposition of these spectra (Fig. 3.3a) reveals small $^1H^\varepsilon - ^{15}N^\varepsilon$ chemical shift differences several corresponding arginines of the wild-type and mutant protein attributed to the T26H substitution and differences in exact sample conditions. (C-E) With increasing pH, the $^1H^\varepsilon - ^{15}N^\varepsilon$ signals from T26H-T4L* disappear due to base-catalyzed hydrogen exchange, with only well-protected arginines detected under alkaline conditions. The lack of any significant pH-dependent $^{15}N^\varepsilon$ chemical shift changes shows that the detected arginines remain positively charged. The small downfield shift in the $^1H^\varepsilon$ signal of Arg148 follows an apparent $pK_a$ of 7.2. There are no ionizable groups proximal to Arg148 with a $pK_a$ value in this range and thus I speculate that the this chemical shift change reflects altered pH-dependent interactions with buffer phosphate ($pK_{a2}$ ~ 7).
3.5 Discussion

Using NMR spectroscopy, I measured the pKₐ values of the Asp, Glu, and His residues in T26H-T4L* and, further, demonstrated that all arginines are positively charged at neutral pH. Along with previously reported data for T4L, these results provide important constraints for understanding the catalytic mechanisms of the wild-type and mutant lysozymes.

3.5.1 Inverting mechanism of wild-type T4L

Wild-type T4L is an inverting glycoside hydrolase. A reasonable single-displacement mechanism (Kuroki et al., 1993) involves Asp20 acting as a general base to activate a nucleophilic water that is also hydrogen bonded to Thr26. This water carries out a nucleophilic substitution, with inversion of stereochemistry, at the anomeric carbon of a peptidic NAM. Concomitantly, Glu11 serves as a proton donor to facilitate departure of the NAG aglycone.

This mechanism leads to the prediction that T4L should display a bell-shaped activity versus pH profile, with a pH optimum of 4.5 where Asp20 (pKₐ 3.6) is predominantly deprotonated to act as a general base and Glu11 (pKₐ 5.4) predominantly protonated to function as a general acid. In this context (and below), "predicted activity" refers to kₕ/Kₘ, the second order rate constant for reaction of free enzyme and substrate, and neglects likely contributions of pH-dependent interactions between the charged substrate peptide moiety, LAla-DGlu-DAP-LAla, with an extended binding site on the protein surface (Kuroki et al., 1993). (The first order rate constant for turnover of the Michaelis complex, kₕ, will depend upon the pKₐ values of catalytic residues with bound substrate, which have yet to be measured.) Unfortunately, these fundamental catalytic parameters have not been determined for T4L due to the lack of a well-defined substrate amenable to detailed kinetic analyses. Although it has been reported that the lytic activity of T4L
on chloroform-treated *E. coli* cells is maximum at pH ~ 7.3, with sensitivity to ionic strength, amines and divalent cations (Jensen and Kleppe, 1972; Hardy and Poteete, 1991), numerous factors complicate the mechanistic interpretation of these results.

### 3.5.2 Retaining mechanism of T26H-T4L with His26 as a nucleophile

Remarkably, the single substitution of Thr26 with a histidine converts T4L from an inverting to a retaining glycoside hydrolase with transglycosylase activity. Based on the classical Koshland double-displacement mechanism, a plausible pathway (Fig. 3.1a) was proposed for T26H-T4L that involves nucleophilic attack of His26 on the peptidic NAM moiety to form an α-linked glycosyl-enzyme intermediate (Kuroki et al., 1993, 1995, 1999). This glycosylation step is facilitated though protonation of the leaving aglycone by the general acid Glu11. In the subsequent deglycosylation step, this same residue serves as a general base to activate either a nucleophilic water (hydrolysis) or carbohydrate (transglycosylation) for His26 displacement and glycone release with overall retention of β-anomeric stereochemistry. The alternating role of Glu11 as a general acid and base likely results from changes in its environment, and hence pKₐ values, along the reaction pathway (McIntosh et al., 1996b). In contrast to the wild-type enzyme, Asp20 in T26H-T4L would not play any obvious direct catalytic role except perhaps to facilitate substrate binding by hydrogen bonding to the nitrogen of the N-acetyl group of NAM. Evidence supporting this mechanism includes the observation of a stable peptidic NAG-NAM moiety α-linked to Glu26 in T26E-T4L (Kuroki et al., 1993), as well as the structural and enzymatic characterization of an extensive series of lysozymes with mutations at residues 11, 20, and 26 (Kuroki et al., 1995, 1999). In particular, the Nε² of His26 is approximately positioned for nucleophilic attack on a docked substrate, and the T26Q mutant is inactive.
Accepting this proposed mechanism leads to the prediction that the pH-dependent activity ($k_{\text{cat}}/K_m$) of T26H-T4L will follow a bell-shaped profile with a maximum at pH 5.75. However, the $pK_a$ value of the general acid Glu11 (4.7) is lower than that of the postulated nucleophile His26 (6.8). Thus, at this predicted pH optimum, only a small fraction of the protein ($\sim 1\%$) will be in the catalytically competent ionization state with Glu11 protonated and His26 deprotonated. Such a "reverse protonation mechanism" is well established (albeit underappreciated) for many enzymes, and requires that the low population of suitably ionized residues be offset by their high catalytic efficiency (Mock, 1992). It may not seem intuitive that the pH-dependence of $k_{\text{cat}}/K_m$ for a reverse protonation mechanism would reflect the population-averaged macroscopic $pK_a$ values of the catalytic residues measured by NMR spectroscopy. (McIntosh et al., 2011a). That is, due to electrostatic interactions, the microscopic $pK_a$ values of proximal Glu11 and His26 will likely differ depending on whether the other is neutral or charged. However, at the predicted pH optimum of 5.75, the major population of T4L with His31 protonated and Glu11 deprotonated will be in a $pH$-independent equilibrium with the minor population having the catalytically competent deprotonated His31 and protonated Glu11.

3.5.3 Alternative mechanism for T26H-T4L

Although in principle histidine can serve as a nucleophile for retaining glycoside hydrolases, it is striking that none to date has ever been demonstrated or proposed to play this role in any such enzyme other than T26H-T4L. This prompts the question whether T26H-T4L could follow an alternative catalytic mechanism. One possibility, exemplified by GH20 $\beta$-hexosaminidases (Mark et al., 2001), involves the nucleophilic participation of the N-acetyl group of NAM (Fig. 3.1b). Such participation could become feasible if the T26H mutation
results in interactions that are favourable to the requisite alignment of the side chain amide moiety. A particularly attractive feature of this hypothesized mechanism is that we know it is inherently possible for the N-acetyl moiety to adopt the correct geometry for nucleophilic attack since it does so in other enzymes. In contrast, we do not know with certainty that this is the case for the nitrogen of His26.

In this hypothesized substrate-assisted catalysis mechanism, Asp20 might act as a general base to abstract a proton from the N-acetyl nitrogen and assist formation of a bicyclic oxazoline intermediate. Glu11 would serve as a general acid for aglycone departure and then as a general base for hydrolysis or transglycosylation, with net retention of β-anomeric stereochemistry. The predicted pH optimum in $k_{cat}/K_m$ for such a mechanism would be ~ 3.1 with Asp20 ($pK_a \sim 1.5$) predominantly deprotonated and Glu11 ($pK_a 4.7$) predominantly protonated in the free enzyme. However, with such a low $pK_a$ value, Asp20 would be a poor general base and might rather remain negatively charged to stabilize a positively charged oxazolinium intermediate without proton transfer. (Coines et al., 2018) It is also worth considering that, in the Michaelis complex with bound substrate, the $pK_a$ values of these catalytic residues could be better tuned for efficient catalysis.

In contrast to the mechanism proposed previously, Asp20, if involved, serves a direct catalytic role while that of His26 is not obviously defined. This is conformationally plausible as Asp20 is hydrogen bonded to the N-acetyl group of NAM in the X-ray crystallographic structure of glycosylated T26E-T4L (Kuroki et al., 1993), yet less consistent with the observations that the T4L-T26Q mutant is inactive, whereas the D20C/T26H double mutant is still active as a retaining glycoside hydrolase (Kuroki et al., 1999). However, wild-type T4L is tolerant to substitutions to substitutions at position 20, including D20C for which cysteine could serve as a
general base (Hardy and Poteete, 1991). Furthermore, the very fact that substituting Thr26 with His changes the stereochemical outcome of peptidoglycan hydrolysis suggests an active site plasticity that tips the balance between different catalytic mechanisms. Resolving these mechanisms will require more detailed enzymatic studies with well-defined substrates.

3.5.4 Arg145 is protonated with a high pKa value

Using NMR spectroscopy, I showed conclusively that the Nε of Arg145 in both wild-type T4L* and T26H-T4L* is protonated under typical experimental conditions. Also, from chemical shift arguments, the sidechain guanidinium group of this residue in the mutant protein is most certainly ionized with a pKₐ > 11. Although largely buried in the X-ray crystal structures of both proteins, stabilization of the delocalized positive charge of Arg145 may result from extensive hydrogen bonding between the terminal Nᵪ¹ and Nᵪ² with the carboxylate of Glu11 and the Nᵪ² and Nε with the sidechain oxygen of Asn101. Considerable debate has centred on the possibility of a deprotonated arginine sidechain in a protein under physiological conditions (Harms et al., 2011; Fitch et al., 2015). With an intrinsic pKₐ value of nearly 14, a substantial energetic penalty would be paid to accommodate a neutral guanidino group. This penalty is on the order of global protein folding energetics, and thus any such species is unlikely to exist except perhaps in a minor population along an enzymatic pathway in which arginine functions as a general base (Schlippe and Hedstrom, 2005).

3.5.5 Ionization states determined from neutron crystallography

Neutron crystallography is a well-established approach for identifying hydrogens in molecules, including proteins. Indeed, the ionization states of the Asp, Glu, His and Arg residues
found herein by NMR spectroscopy are consistent with those determined in a recently published joint neutron and high-resolution X-ray diffraction study of perdeuterated T4L* at pD 6 - 7 (5VNQ.pdb and 5VNR.pdb) (Li et al., 2017). Thus it somewhat perplexing why in a similar joint diffraction study of perdeuterated T26H-T4L* at pD 7, Arg145 was reported to be Nε deprotonated, Asp20 carboxyl protonated, Glu11 partially (~ 60 %) carboxyl protonated, and His26 deprotonated at Nε2 (5XPE.pdb and 5XPF.pdb) (Hiromoto et al., 2017). As I have demonstrated for this protein in solution, Asp20 has a pKά ~ 1.5 and Arg145 is Nε protonated with an inferred pKά > 11. Thus both are effectively completely charged under physiologically relevant pH conditions. Also, at pH 7, Glu11 (pKά 4.7) and His 31 (pKά 6.8) should be ~ 1 % and 40 % protonated, respectively. There is no obvious physicochemical reason as to why substitution of Thr26 to His would dramatically alter the pKά values of these active site residues in T26H-T4L versus wild-type T4L, or why T26H-T4L would differ between solution and crystalline states. Although crystal packing or protein and solvent deuteration could explain the smaller discrepancies for Glu11 and His31, it is difficult to imagine how this could lead to such energetically unfavorable changes for Asp20 and Arg145. Unexplained disagreements between results obtained by NMR spectroscopy and neutron diffraction can also be found in the case of photoactive yellow protein (Yoshimura et al., 2017; Yonezawa et al., 2017). This speaks to the challenges in defining protein protonation states and pKά values, and the need to use complementary methods to obtain a comprehensive understanding of the pH-dependent ionization states of biological macromolecules.
Chapter 4: Conclusions and future studies

The general aim of my thesis was to investigate labile protons with biologically interesting roles in proteins. To this end, I developed new NMR methods for characterizing serine and threonine hydroxyl protons, and utilized established NMR approaches to determine the protonation states of a T4 lysozyme variant with an altered catalytic mechanism. This continues the overall goal of biophysical chemists to study proteins with ever increasing detail in order to better parameterize, model and ultimately predict their properties. There are still several outstanding questions that, with considerable effort, could expand our understanding of these projects.

4.1 Characterizing hydroxyl protons by NMR spectroscopy: Summary, significance, limitations, and potential applications

In Chapter 2, I utilized selective labeling strategies and minor modifications to standard pulse sequences to characterize hydroxyl protons found on serines and threonines by NMR spectroscopy. By optimizing the “1/4J” for 3-bond couplings in standard INEPT pulse sequences for 3-bond couplings, I was able to detect and characterize five protected hydroxyl protons (from two serines and three threonines) in the xylanase from Bacillus circulans. This strategy has several advantages over previously utilized NMR techniques for hydroxyl proton assignment. The most significant is that a through-bond approach allows for unambiguous assignments. In contrast, previous techniques relied on potentially ambiguous through-space NOE interactions. The second benefit is that the use of selective isotopic labeling greatly simplifies acquired spectra by reducing the number of signals. This allowed me to further parametrize hydroxyl proton exchange rates and coupling constants.
The developed techniques have several limitations that, if overcome, would greatly expand their applications. A primary limitation of this strategy is that it is restricted to well protected hydroxyl groups. Hydroxyls in intermediate to fast exchange with water have broadened or undetectable chemical shifts, thus precluding their observation by NMR spectroscopy. However, this general approach could be applied to solid state NMR, where it may be possible to detect much more solvent exposed residues in a frozen sample.

Although advantageous for spectral simplification, a second limitation of this technique is the requirement of selective amino acid isotopic labeling. Optimization of the “1/4J” delay for 3-bond couplings was chosen as a combination of three factors: suppressing the strong 1-bond couplings, allowing adequate time for the buildup of the much weaker 3-bond couplings, and minimizing the total experiment time to limit signal loss from relaxation. In principle, these experiments could be carried out using uniformly $^{13}$C/$^{15}$N labeled samples. However, it would be unlikely one could optimize this delay to suppress a large background of 1-bond couplings (with variable magnitudes) in order to unambiguously assign the weak signals from a small number of hydroxyls. Moreover, a plethora of possible 2- and 3-bond $^1$H-$^{13}$C from the uniformly labeled protein would surely obscure those from the hydroxyls.

Additionally these results in combination with structural data on BcX could be used to begin to parametrize the Karplus equation for serine and threonine dihedral $^3J_{C-OH}$ couplings. Although given the small $^3J_{C-OH}$ coupling constants determined (5 - 8 Hz) and in the case of threonine and the lack of complementary $^3J_{Ca-OH}$ for the detected $^3J_{Cy-OH}$, it is very likely that the detected hydroxyls are in a trans conformation to the coupled $^{13}$C nucleus in order to have a large enough $^3J_{C-OH}$ to be detected by current techniques. This parameterization may not be necessary as detection alone would dictate dihedral geometry.
This technique can be extended to other residues with labile protons, as previously done with the tyrosine hydroxyl. Another logical choice would be to apply these strategies to the thiol hydrogens of cysteines. These experiments would be straightforward as there are well-behaved cysteine auxotrophs available to aid isotopic labeling, and only minor modifications to the pulse sequences would be required.

4.2 Characterizing T4L* T26H: Summary, significance, limitations, and potential applications

In Chapter 3, I examined a most peculiar mutant of T4 lysozyme which had previously been shown to change its catalytic mechanism from inverting to retaining. The authors of a neutron diffraction study (Hiromoto et al., 2017) suggested this arose from highly unusual protonation states of ionizable residues in the catalytic site of T26H-T4L. Through NMR-monitored pH titrations, I was able to determine protonation states and pKₐ values of most ionizable residues in this lysozyme mutant.

This work further resolves the limits to which local protein environment perturbs the protonation state of its constituent amino acids. In the case of arginine (and despite reports to the contrary), it is difficult to envision its folded environment in a globular domain being capable of overcoming the energetic penalty required to reduce the intrinsic pKₐ (~ 13.8) to < 7 in order to be neutral at physiological pH conditions. Indeed when the reportedly highly perturbed Arg145 of T26H-T4L was monitored in the solution state, the residue proved quite normal. This highlights the challenges of characterizing labile hydrogens and the limitations of current structural techniques. This also provides strong evidence that all arginines are positively charged at physiologically relevant pH values in folded proteins.
The protonation states and $pK_a$ values of the Asp, Glu and His residues in T26H-T4L were determined. These residues had varying levels of perturbations from their intrinsic $pK_a$ values. These experimental data may help in the computational prediction of protein $pK_a$ values for which accurate algorithms are currently lacking. Comparing these results to previously reported results for WT-T4L also provides excellent examples of how minor changes to local environment can have large effects on protein electrostatics. The most striking example of this is Asp20. In the WT enzyme, Asp20 is hydrogen bonded to the hydroxyl of Thr26 and has a $pK_a$ value of 3.6. In the T26H mutant, Asp20 is rotated by 90° about its $\chi_1$ dihedral angle, This allows its carboxyl group to hydrogen bond with the backbone and perturbs its $pK_a$ value down two units to 1.5.

Perhaps of highest interest, the proposed nucleophile His26 of T26H-T4L has a higher $pK_a$ (6.8) than the general acid Glu11 ($pK_a$ 4.7). This suggests that T26H-T4L utilizes reverse protonation. If the proposed double-displacement mechanism is correct, it adds to the body of evidence that reverse protonation is utilized much more frequently than previously assumed as an effective catalytic scheme.

However, it is possible that His26 is not the nucleophile. If so, the most likely alternative mechanism would utilize neighboring group participation of the 2-acetamido group of NAG to act as an internal nucleophile. Differentiating between these two mechanisms for T4L is exceedingly difficult. In principle, this could be accomplished by producing inhibitors tailored to either mechanism. If His26 is indeed the nucleophile, replacement of the hydrogen at the two position of NAG in the peptidoglycan substrate with an electron withdrawing fluorine could trap the glycosyl-enzyme intermediate. Alternatively, if the 2-acetamido group is acting as an internal
nucleophile, a thiazoline peptidoglycan analogue could inhibit T4L\* T26H by mimicking the oxazoline intermediate. Preparing both proposed inhibitors would be a daunting synthetic task.

Obtaining a pH-dependent activity profile of T26H-T4L would also help differentiate potential mechanisms. This is also challenging albeit more feasible. Purified peptidoglycan has previously been prepared for activity assays and could be used to obtain a pH-activity profile. There are two major drawbacks to the assay. The first is that the purification of the substrate peptidoglycan is tedious. The second is that the reaction products are monitored by reverse phase HPLC, which would likely be insufficient for the rigors of enzyme kinetics and why previous studies have only characterized reaction products. It is also likely that the non-catalytic ionizations would be influencing the activity profile and thus complicate its interpretation.

The work presented has increased the number of methods available for labile hydrogen detection, added to the library of explicitly determined protonation states and pK\(_a\) values linked to structure, characterized perturbed ionizable groups to give insight into an enzymatically interesting system and posed questions on the utility of ionizable residues. In summary, I have continued a long line of work to nudge the field forward and raised new questions for a seemingly banal and extensively studied topic.
Bibliography


Kuroki, R., L. H. Weaver, and B. W. Matthews. 1999. 'Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site', Proceedings of the National Academy of Sciences of the United States of America, 96: 8949-54.


Appendices

Appendix A  Pulse sequences for the detection and characterization of hydroxyls

Figure A.1 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered experiment to detect signals from $^1$H nuclei bonded directly to oxygen (or sulfur) atoms while suppressing those bonded directly to $^{15}$N or $^{13}$C nuclei (see Figure 1). Carrier frequencies: $^1$H, 4.8 ppm (water); $^{13}$C, 80 ppm (between the aliphatic and aromatic regions); and $^{15}$N, 119.5 ppm (amide region). Delays (not including pulse widths): $\tau_a = 2.72$ ms; $\tau_b = 1.78$ ms; and $\tau_c = 1.32$ ms. Narrow (wide) rectangular pulses have flip angles of $90^\circ$ ($180^\circ$). The last three $^1$H pulses represent WATERGATE with the hard $180^\circ$ pulse surrounded by rectangular 1.8 ms $90^\circ$ pulses. The trapezoids are “smoothed Chirp” (20% smoothed) $180^\circ$ $^{13}$C pulses with a 60 kHz sweep width (from low field to high field) and a duration of 1730 $\mu$s (for a 600 MHz NMR spectrometer). The triangles represent $z$- (xyz- for g5) gradients with smoothed square shapes: g1 = 9.8 G/cm (1 ms); g2 = 14.7 G/cm (300 $\mu$s); g3 = 19.6 G/cm (1 ms); g4 = 14.7 G/cm (100 $\mu$s); and g5 = 34.3 G/cm (1 ms). Although not required, GARP-4 (WALTZ-16) decoupling with 4.5 (1.5) kHz of $^{13}$C ($^{15}$N) field strength (for a 600 MHz spectrometer) improves filtering.

All pulses are along the x-axis, except for the following phase cycling:

$\phi_1 = x, y, -x, -y$
$\phi_2 = 4x, 4y, 4(-x), 4(-y)$
$\phi_3 = 4(-x), 4(-y), 4x, 4y$
$\phi_4 = x, y, -x, -y$
Receiver = x, -y, -x, y
Figure A.2 Pulse sequence for a 2D $^{13}$C-HSQC experiment. Carrier frequencies: $^1$H, 4.78 ppm (water); $^{13}$C, 46.5 ppm (aliphatic region); and $^{15}$N, 119.5 ppm (amide region). Delays (not including pulse widths): $\tau_a = 1.7$ ms (conventional one-bond "$1\text{/}4J""), 11.4 ms (threonine long range), or 10.5 ms (serine long range); and $\tau_b$ = duration of gradient $g1$ (1 ms). Narrow (wide) rectangular pulses have flip angles of 90° (180°). The trapezoids are “smoothed Chirp” (20% smoothed) 180° $^{13}$C pulses a 60 kHz sweep width (from low field to high field) and a duration of 500 μs, whereas the triangles represent “Gauss cascade Q3” 180° $^{13}$CO pulses (176 ppm) with a duration of 384 μs (for a 600 MHz NMR spectrometer). GARP-4 decoupling was applied with 4.5 kHz of $^{13}$C field strength (for a 600 MHz NMR spectrometer). The triangles also represent $xyz$-gradients (smoothed square shapes): $g1 = 34.3$ G/cm (1 ms); and $g2 = 8.62$ G/cm (1 ms).

All pulses are along the x-axis, except for the following phase cycling:

- $\phi_1 = x$, -x
- $\phi_2 = 2(x), 2(-x)$
- $\phi_3 = 4(x), 4(-x)$
- Receiver = x, -x, x, 2(-x), x, -x, x

Quadrature in $t_1$ is obtained by “echo-antiecho” method, changing the $g1$ sign for each successive $t_1$ data point.
Figure A.3 Pulse sequence for a 2D long-range constant-time methyl $^{13}$C-HSQC experiment. Carrier frequencies: $^1$H, 4.8 ppm (water); and $^{13}$C, 20 ppm (methyl region). Delays (not including pulse widths): $\tau_a = 1.7$ ms (conventional one-bond) or 11.4 ms (threonine long range); and $T_c = 13.0$ ms. Narrow (wide) rectangular pulses have flip angles of 90° (180°). The small triangle represents a methyl 180° “IBURP2” pulses with a duration of 1200 μs, whereas the tall triangle represents a “Gauss cascade Q3” 180° $^{13}$C pulse at 40 ppm with a duration of 256 μs. WALTZ-16 decoupling was applied with 3.1 kHz of $^{13}$C field strength (for a 600 MHz NMR spectrometer). The triangles also represent z- (g1) and xyz- (g2, g3) gradients (smoothed square shapes of 1 ms duration, followed by 200 μs delays): g1 = 15 G/cm; g2 = 34.3 G/cm; and g3 = 8.62 G/cm.

All pulses are along the x-axis, except for the following phase cycling:
- $\phi_1 = 8(y), 8(-y)$
- $\phi_2 = 4(x), 4(-x)$
- $\phi_3 = x, y, -x, -y$
- Receiver = x, -x, x, 2(-x), x, -x, x, -x, x, -x, 2x, -x, x, -x

Quadrature in $t_1$ is obtained by “echo-antiecho” method, changing the g2 sign for each successive $t_1$ data point.
**Figure A.4** Pulse sequence for a 2D methyl-selective $^{13}$C$^\beta$-decoupled $^{13}$C-HSQC experiment. Carrier frequencies: $^1$H, 4.78 ppm (water); $^{13}$C, 46.5 ppm (aliphatic region); and $^{15}$N, 119.5 ppm (amide region). Delays (not including pulse widths): $\tau_a = 1.7$ ms (direct coupling) or 11.4 ms (long range); and $\tau_b = \text{duration of gradient } g_1$ (1 ms). Narrow (wide) rectangular pulses have flip angles of 90° (180°). The triangles represent “IBURP2” $^{13}$C pulses with a duration of 847 μs at 71 ppm. WALTZ-16 decoupling was applied with 3.8 kHz of $^{13}$C field strength (for an 850 MHz NMR spectrometer). The triangles also represent xyz-gradients (smoothed square shapes): $g_1 = 34.3$ G/cm (1 ms); and $g_2 = 8.62$ G/cm (1 ms).

All pulses are along the x-axis, except for the following phase cycling:

- $\phi_1 = x, -x$
- $\phi_2 = 2(x), 2(-x)$
- $\phi_3 = 4(x), 4(-x)$
- Receiver = $x, -x, x, 2(-x), x, -x, x$

Quadrature in $t_1$ is obtained by “echo-antiecho” method, changing the $g_1$ sign for each successive $t_1$ data point.
Figure A.5 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered quantitative spin-echo difference experiment to measure $^{13}$C couplings to hydroxyl protons. Carrier frequencies: $^1$H, 4.78 ppm (water); $^{13}$C, 80 ppm (between the aliphatic and aromatic regions); and $^{15}$N, 118 ppm (amide region). Delays (not including pulse widths): $\tau_a = 2.78$ ms; $\tau_b = 1.92$ ms; and $\tau_c = 1.25$ ms. The spin-echo delay $\Delta = 21$ ms (including pulse widths) was chosen empirically to maintain good $^{13}$C/$^{15}$N-filtering. Narrow (wide) rectangular pulses have flip angles of $90^\circ$ ($180^\circ$). The last three $^1$H pulses represent WATERGATE with the hard $180^\circ$ pulse surrounded by rectangular 1.8 ms $90^\circ$ pulses. The small (tall) trapezoids are “20% smoothed Chirp” $180^\circ$ $^{13}$C pulses with a 60 kHz (80 kHz) sweep width, from low field to high field for the 850 MHz NMR spectrometer, and a duration of 1500 $\mu$s (400 $\mu$s). Alternatively, the final broadband $^{13}$C $180^\circ$ Chirp pulse (or its absence) can be replaced using hard $90^\circ$,$90^\circ$,$x$ (or $90^\circ$,$90^\circ$,$x$) pulses at 22 pm (threonine $^{13}$C$^\beta$), 57 ppm (serine $^{13}$C$\alpha$), or 118 ppm (tyrosine $^{13}$C$\epsilon$). The triangles represent z- (xyz- for g5) gradients (g1 and g3 are rectangular, others with smoothed square shapes): g1 = 1.5 G/cm (300 $\mu$s); g2 = 20 G/cm (1 ms); g3 = 1.0 G/cm (100 $\mu$s), g4 = 11.5 G/cm (1 ms), g5 = 30.0 G/cm (1 ms).

The reference spectrum (N scans) is obtained by omitting the final broadband $^{13}$C Chirp $180^\circ$ while maintaining the $\Delta$ delay (or using a hard $90^\circ$,$90^\circ$,$x$). All pulses are along the x-axis, except for the phase cycling:

$\phi_1$ = x, y, (-x), (-y)
Receiver = x, -y, -x, -y

The difference spectrum (2N total scans) is collected in interleaved mode, whereby the final $^{13}$C Chirp $180^\circ$ pulse is present ($^{13}$C Chirp $180^\circ$ or a hard $90^\circ$,$90^\circ$,$x$) for even scans and absent (no $^{13}$C Chirp $180^\circ$ or a hard $90^\circ$,$90^\circ$,$x$) for odd scans. All pulses are along the x-axis, except for the phase cycling:

$\phi_1$ = 2x, 2y, 2(-x), 2(-y)
Receiver = x, -x, -y, y, -x, x, y, -y
Figure A.6 Pulse sequence for a 1D $^{13}$C/$^{15}$N-filtered CLEANEX-PM to measure the hydrogen exchange kinetics of $^1$H nuclei bonded directly to oxygen (or sulfur) atoms (see Figure 1). Frequencies: $^1$H, 4.8 ppm (water); $^{13}$C, 70 ppm (between the aliphatic and aromatic regions); and $^{15}$N, 119 ppm (amide region). Delays (not including pulse widths): $\tau_a = 2.72$ ms; $\tau_b = 1.78$ ms; and $\tau_c = 1.32$ ms. Narrow (wide) rectangular pulses have flip angles of 90° (180°). The first selective water 180° $^1$H pulse has 7.5 ms Gaussian shape, and the last three $^1$H pulses represent WATERGATE with the hard 180° pulse surrounded by rectangular 1.8 ms 90° pulses. The trapezoids are “smoothed Chirp” (20% smoothed) 180° $^{13}$C pulses with a 60 kHz sweep width (from low field to high field) and a duration of 1730 μs (for a 600 MHz NMR spectrometer). The CLEANEX-PM spin-lock was applied with 4.8 kHz of $^1$H field strength, and although not required, GARP-4 (WALTZ-16) decoupling with 4.5 (1.5) kHz of $^{13}$C ($^{15}$N) field strength (for a 600 MHz spectrometer) improves filtering. The triangles represent $z$- (xyz- for g6) gradients (smoothed square shapes of 1 ms duration followed by 200 μs delays): $g_1 = 9.3$ G/cm; $g_2 = 9.8$ G/cm; $g_3 = 14.7$ G/cm; $g_4 = 19.6$ G/cm; $g_5 = 14.7$ G/cm; and $g_6 = 34.3$ G/cm. The rectangle represents a 0.02 G/cm $z$-gradient (g0) applied constantly during the CLEANEX-PM sequence.

All pulses are along the x-axis, except for the following phase cycling:

$\phi_1 = x, y$

Receiver = $x, -x$
Figure A.7 Pulse sequence for a 3D intra-residue C(C)TOCSY_NH used to correlate the sidechain $^{13}$C and amide $^{15}$N-$^1$H signals of selectively labeled BcX. Carrier frequencies: $^1$H, 4.8 ppm (water) and $^{15}$N, 115 ppm (amide region). $^{13}$C carrier frequency is staying at 43 ppm (aliphatic region) at the beginning of the pulse sequence and moving to 176 ppm (CO) right after gradient pulse $g_3$. Delays (not including pulse widths): $\tau_a = 1.7$ ms; $\tau_b = 1.1$ ms; $\tau_c = 12.4$ ms; $\tau_d = 5.5$ ms; $\tau_e = 2.3$ ms; and $\tau_f = 1.21$ ms. Narrow (wide) rectangular pulses have flip angles of 90° (180°). The triangles represent “Gauss cascade Q3” 90° (180°) $^{13}$C pulses with a duration of 384 μs (307 μs) (for a 600 MHz NMR spectrometer). WALTZ-16 decoupling was applied with 4.5 (1.0) kHz of $^1$H ($^{15}$N) field strength, and the DIPSI-2 spin-lock was 10.4 kHz of $^{13}$C field strength (for a 600 MHz NMR spectrometer). The triangles also represent z- (xyz- for $g_4$ and $g_7$) gradients (smoothed square shapes of 1 ms duration followed by 200 μs delay): $g_1 = 7.3$ G/cm; $g_2 = 24.5$ G/cm; $g_3 = -14.7$ G/cm; $g_4 = 34.3$ G/cm; $g_5 = 14.7$ G/cm; $g_6 = 29.4$ G/cm; and $g_7 = 3.47$ G/cm.

All pulses are along the x-axis, except for the following phase cycling:

$\phi_1 = 8(y), 8(-y)$
$\phi_2 = 2(x), 2(-x)$
$\phi_3 = 4(x), 4(-x)$
$\phi_4 = x, -x$
Receiver = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x

Quadrature in $t_1$ is obtained by States-TPPI method with phase $\phi_2$. Quadrature in $t_2$ is obtained by the “echo-antiecho” method, with FIDs recorded separately for $\phi_5 = -y$ ($g_4 = 34.3$ G/cm) and for $\phi_5 = y$ ($g_4 = -34.3$ G/cm).
Figure A.8 Pulse sequence for a 3D intra-residue H(CC)TOCSY-NH used to correlate the sidechain $^1$H and amide $^{15}$N-$^1$H$^N$ signals of selectively labeled BcX. Carrier frequencies: $^1$H, 4.8 ppm (water) and $^{15}$N, 115 ppm (amide region). $^{13}$C carrier frequency is staying at 43 ppm (aliphatic region) at the beginning of the pulse sequence and moving to 176 ppm (CO) right after gradient pulse g2. Delays (not including pulse widths): $\tau_a = 1.7$ ms; $\tau_b = 1.1$ ms; $\tau_c = 12.4$ ms; $\tau_d = 5.5$ ms; $\tau_e = 2.3$ ms; and $\tau_f = 1.21$ ms. Narrow (wide) rectangular pulses have flip angles of 90° (180°). The triangles represent “Gauss cascade Q3” 90° (180°) $^{13}$C pulses with a duration of 384 μs (307 μs) (for a 600 MHz NMR spectrometer). WALTZ-16 decoupling was applied with 4.5 (1.0) kHz of $^1$H ($^{15}$N) field strength, and the DIPSI-2 spin-lock was 10.4 kHz of $^{13}$C field strength (for a 600 MHz NMR spectrometer). The triangles also represent z- (xyz- for g3 and g6) gradients (smoothed square shapes of 1 ms duration followed by 200 μs delay): g1 = 24.5 G/cm; g2 = -14.7 G/cm; g3 = 34.3 G/cm; g4 = 14.7 G/cm; g5 = 29.4 G/cm; and g6 = 3.47 G/cm.

All pulses are along the x-axis, except for the following phase cycling:

$\phi_1 = 2(x), 2(-x)$
$\phi_2 = 8(x), 8(-x)$
$\phi_3 = 4(x), 4(-x)$
$\phi_4 = x, -x$
Receiver = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x

Quadrature in $t_1$ is obtained by States-TPPI method with phase $\phi_1$ incremented by 90° for each successive $t_1$ data point. Quadrature in $t_2$ is obtained by the “echo-antiecho” method by changing the $\phi_5$ phase by 180° and the g3 sign simultaneously for each successive $t_2$ data point.