Studies of Amino-Specific Chemical Tags. Prospects for $^{19}$F and $^{13}$C Protein NMR and Materials Chemistry

by

Sacha Thierry Larda

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Chemistry

University of Toronto

© Copyright by Sacha Thierry Larda 2015
Studies of Amino-Specific Chemical Tags. Prospects for $^{19}$F and $^{13}$C Protein NMR and Materials Chemistry

Sacha Thierry Larda

Doctor of Philosophy

Department of Chemistry
University of Toronto

2015

ABSTRACT

Site-specific chemical tagging of proteins is a precursor to many structurally-focused biological studies by NMR and mass spectrometry. In particular, reductive methylation is a well-known approach to incorporate methyl groups onto lysine and N-terminal amines in proteins. Interestingly, the NMR spectra of reductively dimethylated lysine residues are often complex, due to methyl inequivalence and chemical exchange. Whereas prior studies on small molecules have eluded to some of the underlying exchange processes, the fundamental origins of methyl inequivalence for dimethyllysines have been largely neglected. Furthermore, prior to this thesis, no comprehensive model had been proposed which encapsulates all elements of the exchange phenomena, showing how they are interconnected and give rise to experimentally observed NMR spectra of reductively methylated proteins. The first part of this thesis addresses the methyl exchange for both small molecule substituted dimethylamines and for dimethylamino groups in proteins and should therefore be useful for better understanding methyl spectra of reductively methylated proteins as well as methyl NMR spectra in general.

The second half of this thesis is primarily concerned with the application of $^{19}$F NMR toward the study of protein chemical exchange as well as the design of novel $^{19}$F tags for site-specific labeling of amino acid side chains. The former pertains to a study of exchange between monomeric and oligomeric states
of the toxic PrP$^\text{Sc}$ form of the prion protein, which has implications for understanding the mechanisms responsible for aggregation and pathogenesis. The subsequent two chapters describe a computational method for the design of highly-sensitive $^{19}\text{F}$ chemical tags and also introduce a new general bioconjugation strategy for chemically modifying amino groups with reactive indoles, with an emphasis on $^{19}\text{F}$ NMR applications. The final chapters introduce a cross-saturation, spin-diffusion-based technique that allows for the detection of subtle conformational changes in proteins.

The methodology described in this thesis extends the repertoire of techniques for designing new chemical tags for NMR applications, for site-specifically labeling proteins, and demonstrates NMR-based approaches to studying disease-relevant chemical exchange processes, characterizing conformational intermediates, and for understanding how biophysical characterization of side chain dynamics and kinetic processes can give insight into structure-function relationships in proteins.
Acknowledgements

I would first like to give a big thank you to my supervisor/advisor, Scott Prosser, for his tutelage, support, and guidance throughout the years. Undoubtedly during the duration of a PhD, the student encounters difficult or frustrating periods where the only remedy is the advice of a good mentor to get you over a wall, and for that I’m thankful to Scott. His advice was critical to my moving forward on countless occasions and I found his openness to new ideas always refreshing and stimulating. The freedom to explore, experiment, and investigate ideas that piqued my interest at any given moment was also certainly a privilege I might not have had elsewhere. It was very much an interdisciplinary experience and I wouldn’t have wanted it any other way. Scott has always supported my endeavors and provided the assistance and directive needed to reach my goals. I truly feel that this has been one of the most fulfilling and mentally stimulating times of my life. Thanks again Scott!

I’d also like to thank my lab colleagues and everyone whom I have collaborated with. In particular I’d like to thank Karen Simonetti, Carlene Starck, Libin Ye, Joshua-Hoang, William Thach, Michael Bokoch, Aashish Manglik, Ranjith Muhandiram, and Professors Simon Sharpe, Voula Kanelis, Peter Macdonald, Ulli Krull, Drew Woolley, Julie Forman-Kay, Lewis Kay (University of Toronto), and Professors Brad Hanna (University of Guelph), and Brian Kobilka (Stanford University) for various discussions, protein, plasmid or cell samples, advice, and/or equipment access. Thank you especially to Professor Kanelis and Professor Macdonald as well as the members of their respective groups. I’d also like to thank several current and former lab members of the Prosser group including: Rohan Alvares, Tae Hun Kim, Evelyn Cheung, Sameer Al-Abdul-Wahid, Juli Kitevski-LeBlanc, and of course, Ferenc Evanics. These last three were instrumental in getting a newbie grad set up and familiarized with the lab, protocols, and equipment. Ferenc deserves a special thank you for his constant assistance with NMR spectrometer issues throughout the years, over which we shared numerous coffees. Sameer also deserves special recognition for providing regular NMR training as well as helpful advice and assistance during my earlier years as a grad student.

Finally, I’d like to thank my close friends and family, most especially my parents who have always given me plenty of support and encouragement. Without you, I would certainly not be where I am today.

-Sacha
Table of Contents

ABSTRACT ............................................................................................................................................. ii

ACKNOWLEDGEMENTS .................................................................................................................... iv

TABLE OF CONTENTS .......................................................................................................................... v

LIST OF FIGURES ................................................................................................................................. ix

LIST OF TABLES .................................................................................................................................... xv

1 Chapter 1: Reductively methylated proteins and dimethyllysine dynamics and kinetics

1.1 The chemistry of lysine amino acid side chains and N-terminal amines in proteins ......................... 1
1.2 The chemical modification of lysine side chain amino groups ............................................................ 2
1.3 Other approaches to chemical modification of lysine sidechain amines ............................................. 7
1.4 Reductive methylation of lysine side chain and N-terminal amino groups for $^{13}$C and $^1H$,$^{13}$C NMR of proteins ................................................................................................................................. 10
1.5 Introduction to dimethylamine methyl proton and carbon chemical shift inequivalence .................. 12
1.6 Nitrogen inversions ............................................................................................................................ 16
1.7 Studies on $N^\varepsilon$,$N^\varepsilon$,$N^\alpha$,$N^\alpha$-tetramethyllysine: an analog for dimethylamines in reductively methylated proteins .............................................................................................................................. 21

1.7.1 Abstract ........................................................................................................................................ 21
1.7.2 Introduction ................................................................................................................................. 21
1.7.3 Methods .................................................................................................................................... 22
1.7.4 Results and Discussion ................................................................................................................... 22
1.7.5 Conclusions ............................................................................................................................... 22

1.8 Intramolecular asymmetry and methyl inequivalence: trimipramine maleate .................................. 37
1.9 The $\gamma$-gauche effect for leucine side chain $\delta$-methyls .................................................................... 38

1.10 DFT-based calculations of $^1H$ and $^{13}$C chemical shifts for the $\delta$CH$_3$ groups of leucine side chains ........................................................................................................................................ 43

1.10.1 Abstract..................................................................................................................................... 43
1.10.2 Methods .................................................................................................................................. 43
1.10.3 Results and Discussion .............................................................................................................. 44
1.11  DFT-based calculations of $^1$H and $^{13}$C chemical shifts for the N\alpha amino methyl resonances of protonated tetramethyllysine..................................................................................48

1.11.1 Abstract...........................................................................................................48

1.11.2 Introduction.....................................................................................................48

1.11.3 Methods.........................................................................................................48

1.11.4 Results and Discussion.................................................................................49

1.12  Methyl exchange and $\Delta \delta$ for protonated vs deprotonated states of dimethyl amines.................51

1.12.1 Abstract...........................................................................................................51

1.12.2 Methods.........................................................................................................51

1.12.3 Results and Discussion.................................................................................52

1.12.4 Conclusions...................................................................................................54

1.13  $^{13}$C resonance inequivalence for N\epsilon dimethylamino methyls of lysine side chains in reductively methylated proteins.............................................................................................................55

1.13.1 Abstract...........................................................................................................55

1.13.2 Introduction.....................................................................................................55

1.13.3 Methods.........................................................................................................56

1.13.4 Results and Discussion.................................................................................57

1.14  Average $^1$H resonance inequivalence for N\epsilon dimethylamino methyls of lysine side chains in reductively methylated proteins.................................................................61

1.15 Summary...........................................................................................................68

1.16 References...........................................................................................................69

2  Chapter 2: Lysine methylation strategies for characterizing protein conformations by NMR

2.1 Abstract.............................................................................................................75

2.2 Introduction.......................................................................................................76

2.3 Materials and Methods.....................................................................................77

2.4 Results and Discussion.....................................................................................80

2.5 Conclusions......................................................................................................89

2.6 Supplementary Info..........................................................................................90

2.7 References.......................................................................................................98
3 Chapter 3: Dynamic equilibria between monomeric and oligomeric misfolded states of the mammalian prion protein measured by $^{19}$F NMR

3.1 Abstract ..............................................................................................................................101
3.2 Introduction......................................................................................................................102
3.3 Methods..........................................................................................................................104
3.4 Results.............................................................................................................................107
3.5 Discussion.......................................................................................................................116
3.6 Conclusions.....................................................................................................................119
3.7 Associated Content.........................................................................................................120
3.8 References......................................................................................................................123

4 Chapter 4: Site-specific labeling of protein lysine residues and N-terminal amino groups with indoles and indole-derivatives

4.1 Abstract..........................................................................................................................129
4.2 Introduction.....................................................................................................................129
4.3 Methods..........................................................................................................................130
4.4 Results and Discussion...................................................................................................132
4.5 Conclusions....................................................................................................................141
4.6 Supporting Info.............................................................................................................141
4.7 References......................................................................................................................146

5 Chapter 5: Comparison of chemical shift sensitivity of trifluoromethyl tags – optimizing $^{19}$F NMR studies of proteins

5.1 Abstract..........................................................................................................................150
5.2 Introduction.....................................................................................................................151
5.3 Materials and Methods..................................................................................................153
5.4 Results and Discussion.................................................................................................154
5.5 Conclusions and Final Remarks....................................................................................159
5.6 Supplementary Information.........................................................................................160
5.7 References......................................................................................................................162

6 Chapter 6: Saturation transfer NMR: aliphatic cross saturation spin diffusion
6.1 Cross-saturation (CS) and spin Diffusion ................................................................. 165
6.2 $^{15}$N, $^1$H cross-saturation spin diffusion of proteins as a general technique for monitoring core packing and subtle conformational changes: application to Ca$^{2+}$-calmodulin .......................... 167

6.2.1 Abstract .................................................................................................................. 167
6.2.2 Introduction ............................................................................................................. 168
6.2.3 Materials and Methods ........................................................................................ 168
6.2.4 Results and Discussion ........................................................................................ 169
6.2.5 Conclusions .......................................................................................................... 174

6.3 Application of $^{15}$N, $^1$H cross-saturation spin-diffusion experiments toward monitoring subtle differences in protein stability associated with the presence of chemical stabilizers or unnatural amino acid incorporation ................................................................. 174

6.3.1 Introduction .......................................................................................................... 175
6.3.2 Methods ................................................................................................................ 175
6.3.3 Results and Discussion ........................................................................................ 176
6.3.4 Conclusions .......................................................................................................... 180
6.3.5 References ............................................................................................................ 181

7 Chapter 7: Conclusion and Future Directions

7.1 Conclusion ................................................................................................................. 185
7.2 References ............................................................................................................... 188
List of Figures

Chapter 1

Figure 1.1 Number of lysine and arginine residues across all protein structures in PDB.

Figure 1.2 Equilibrium between deprotonated and protonated states of the lysine sidechain amine.

Figure 1.3 Proton-exchange at the Nα amino group of a terminal glycine residue.

Figure 1.4 Scheme showing proposed mechanistic steps in the reductive methylation of amines.

Figure 1.5 Correlation between pKa and nucleophilicity of amino groups in model compounds.

Figure 1.6 Nucleophilicity and pKa trends for primary, secondary, and tertiary amines.

Figure 1.7 Methylation of lysine side chain amino groups.

Figure 1.8 Rational for why tertiary amines cannot undergo reductive methylation.

Figure 1.9 Structural comparison between unmethylated and dimethylated hen egg white lysozyme.

Figure 1.10 Chemical reagents used for the modification of lysine residues in proteins.

Figure 1.11 Acetylation of lysine residues with anhydrides.

Figure 1.12 Partial charges and delocalization of electron density across the amide bond for acetylated lysine side chain amino groups.

Figure 1.13 Two common fluorogenic reagents used to site-specifically label lysine residues in proteins.

Figure 1.14 Full dimethylation of hen egg white lysozyme (hewl) achieves two methyl groups (red stars) per amino group.

Figure 1.15 Splitting of resonance K13 from reductively methylated calmodulin.

Figure 1.16 Splitting of resonance 7 as pH decreases for reductively methylated Lysozyme.

Figure 1.17 Dimethyllysine spectra for reductively methylated hen egg white lysozyme and ubiquitin.

Figure 1.18 Broadening and splitting of $^{13}$C 1D NMR resonances for the sidechain methyls and N-terminal methyl groups of the small molecule N$^\epsilon$N$^\epsilon$.N$^\alpha$-tetramethyllysine as a function of pH.

Figure 1.19 Depiction of nitrogen inversion.

Figure 1.20 Inversions for the nitrogen center of dimethyllysine.

Figure 1.21 One proposed model for the origin of $^1$H and $^{13}$C line broadening for the methyl resonances of dimethyllysine.
Figure 1.22 Distinct methyl groups for substituted dimethylamines.

Figure 1.23 Leucine and tetramethyllysine atomic labeling.

Figure 1.24 Structure of $N^eN^e,N^aN^a$-tetramethyllysine in the deprotonated and protonated states.

Figure 1.25 $^{13}$C 1D NMR (at 150 MHz carbon frequency) of reductively methylated lysine.

Figure 1.26 Proton exchange rates for several amino acids under various solution conditions.

Figure 1.27 Lineshape simulation #1.

Figure 1.28 Lineshape simulation #2.

Figure 1.29 Simulation of two site exchange between inequivalent methyl groups of a substituted dimethylamine.

Figure 1.30 General free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines.

Figure 1.31 Dynamic equilibrium between protonated and deprotonated states of substituted dimethylamines.

Figure 1.32 Arrhenius plots for the protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates versus temperature given in Table 1.

Figure 1.33 Eyring plots for protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates versus temperature given in Table 1.

Figure 1.34 Completed free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines.

Figure 1.35 Structures of trimipramine (1), imipramine (2), and amitriptyline (3) and both $^1$H and $^{13}$C spectra for (1).

Figure 1.36 Leucine side chain $\chi$ rotamers and the $\gamma$-gauche effect.

Figure 1.37 Correlation between sterically induced force along the C-H bond of a methyl group to induced $^{13}$C chemical shift perturbation.

Figure 1.38 Leucine sidechain $\chi_2$ conformers, representative spectra and distributions.

Figure 1.39 $\Delta\delta$ as a function of temperature for the $^{13}$C resonances of the inequivalent $\delta$CH$_3$ groups of leucine side chains in calbindin D$_{9k}$.

Figure 1.40 Sampling distributions of leucine side chain conformations.

Figure 1.41 Conformers of 5-methyl-3-(methylamino)hexan-2-one assessed via DFT.
Figure 1.42 Three major rotamer conformations for $N_\alpha$ of tetramethyllysine.

Figure 1.43 Structures for the three major rotamer conformations for the $N_\alpha$ methyls of dimethylalanine.

Figure 1.44 Gauche plus, trans and gauche minus rotamers for $N_\alpha$ dimethylamino groups.

Figure 1.45 Gauche plus, trans and gauche minus rotamers for $N\text{Na}_\alpha$-dimethylalanine in the protonated (top) and deprotonated (bottom) states.

Figure 1.46 Line shape simulations (blue) fit to experimental $^{13}$C NMR spectra (red) of the K1ε methyl resonances (in exchange) for reductively dimethylated hen egg white lysozyme.

Figure 1.47 General free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines which are mediated by proton-exchange and inversion events. Identical to Figure 30.

Figure 1.48 Eyring plots for the protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates for the K1ε dimethyl amino methyl exchange in reductively methylated hen egg white lysozyme.

Figure 1.49 Completed free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines which are mediated by proton-exchange and inversion events.

Figure 1.50 Representative spectra of reductively methylated proteins in which one or more diethyllysine residues exhibit slow exchange.

Figure 1.51 Methyl resonances in the dimethylamino region of the $^1$H,$^{13}$C HSQC NMR spectra for $^{13}$C-reductively methylated hen egg white lysozyme at pH 7.5 (A) and 9.0 (B).

Figure 1.52 $^1$H-$^{13}$C HSQC spectrum of the region containing the $C\varepsilon$ resonances of lysine in [ε-$^{13}$C] lysine-labeled Pol $\lambda$ lyase domain.

Figure 1.53 Normalized side chain and total residue solvent exposure estimates for two lysine residues (K1 of Lysozyme (193L) and K27 of Ubiquitin (1UBQ)) that exhibit methyl $^1$H resonance inequivalence.

Figure 1.54 Total amino acid residue B-factors for lysines in hen egg white lysozyme (193L) and human ubiquitin (1UBQ).

Figure 1.55 Cation-pi interactions between lysines and aromatic side chains in the crystal structure of hen egg white lysozyme (193L).

Figure 1.56 Methylene side chain conformations for lysine residues K291 (g+) and K287 (trans) for the Lyase Domain of DNA Polymerase $\gamma$. 
Chapter 2

**Figure 2.1** ($^{13}$C, $^1$H) HSQC spectra at 37°C and pH 7.5 of partially methylated HEWL showing assignments for mono- (top) and dimethyl-lysine resonances (bottom)

**Figure 2.2** ($^{13}$C, $^1$H) HSQC spectra of reductively methylated HEWL, as a function of temperature at pH 7.4.

**Figure 2.3** ($^{13}$C, $^1$H) HSQCs of reductively methylated HEWL at 37 °C showing the mono- (left) and dimethyl regions (right) at pH 6.0, 7.5, and 9.0.

**Figure 2.4** Theoretical deprotonation rates for mono- and dimethyl lysines as a function of pH.

**Figure 2.5** Reaction scheme for $^{13}$C-reductive methylation of lysines.

**Figure 2.6** Monomethyl to dimethyl ratios for HEWL labeled at pH 5.7, 7.0, and 8.0.

**Figure 2.7** Reduction of imine with deuteride generates mono- and di-deuteromethyl labels

**Figure 2.8** $^{13}$CH$_2$D and $^{13}$CD$_3$ deuteromethyl labeling is preceded by imine formation with $^{13}$C,d$_2$ formaldehyde.

**Figure 2.9** Comparison of a standard $^{13}$C, $^1$H HSQC spectrum of $^{13}$CH$_2$D-deuteromethylated apo-calmodulin in 300 mg/mL polyvinylpyrrolidone (cyan) versus a $^2$H-edited $^{13}$C, $^1$H HSQC spectrum (black) acquired on the same sample.

**Figure S2.2** Change in averaged chemical shift for each dimethyl lysine of reductively methylated HEWL.

**Figure S2.3** Change in $^1$H chemical shift for each dimethyl lysine of reductively methylated HEWL.

**Figure S2.4** Change in $^{13}$C chemical shift for each dimethyl lysine of reductively methylated HEWL.

**Figure S2.5** Average weighted change in $^1$H and $^{13}$C chemical shift for each monomethyl lysine of reductively methylated HEWL.

**Figure S2.6** Change in $^1$H chemical shift for each monomethyl lysine of reductively methylated HEWL.

**Figure S2.7** Change in $^{13}$C chemical shift for each monomethyl lysine of reductively methylated HEWL.

**Figure S2.8** Comparison of $^{13}$C, $^1$H HSQC spectra of reductively methylated HEW lysozyme, as a function of increasing $^{13}$C-formaldehyde concentrations.

**Figure S2.9** Comparison of $^{13}$C, $^1$H HSQC spectra of monomethyl lysine resonances for reductively methylated BSA, as a function of increasing $^{13}$C-formaldehyde concentrations.

**Figure S2.10** Comparison of $^{13}$C, $^1$H HSQC spectra of dimethyl lysine resonances for reductively methylated BSA, as a function of increasing $^{13}$C-formaldehyde concentrations.
Chapter 3

Figure 3.1 $^{19}$F NMR spectra of 3-fluorophenylalanine-labeled ShaPrP(90–231)$^{\beta}$.

Figure 3.2 Equilibrium populations of $\beta$-state monomer, octamer, and large oligomers formed by ShaPrP(90–231) as a function of temperature.

Figure 3.3 Free energy ($\Delta G$) of association between $\beta$-state components as a function of temperature for WT (●) and F198S ShaPrP(90–231) (△).

Figure 3.4 Normalized populations of monomeric and oligomeric ShaPrP(90–231)$^{\beta}$ as a function of pressure.

Figure 3.5 Equilibrium free energies associated with pressure-induced transitions in the oligomeric state of WT ShaPrP(90–231)$^{\beta}$.

Figure 3.6 $^{19}$F NMR saturation transfer measurements on ShaPrP(90–231)$^{\beta}$. Magnetization decay profiles of total oligomer (purple), octamer (red), and large oligomer (blue) components for WT ShaPrP(90–231)$^{\beta}$.

Figure 3.7 Model for equilibrium exchange between assembled states within ShaPrP(90–231)$^{\beta}$.

Figure S3.1 Dynamic light scattering analysis of ShaPrP(90-231)$^{\beta}$.

Figure S3.2 Diffusion filtered $^{19}$F NMR spectra and physical diffusion profiles of $\beta$-state F198S ShaPrP(90-231).

Figure S3.3 $^{19}$F NMR saturation transfer measurements on Sha(PrP90-231)$^{\beta}$ $\beta$-monomers.

Chapter 4

Figure 4.1 Reaction schemes describing 5-fluoroindole (A) and 5-fluoroindole-2-carboxylate (B) labeling of protein lysine residues.

Figure 4.2 Sites of lysine residues in the crystal structure of hen egg white lysozyme.

Figure 4.3 ESI+ Mass spectra for 5-Fluoroindole- and 5-Fluoroindolecarboxylate-labeled HEWL.

Figure 4.4 5-fluoroindole labeling of HEWL versus time.

Figure 4.5 5-fluoroindole-2-carboxylate labeling of HEWL versus time.

Figure 4.6 Structures of six fluorine tags used for $^{19}$F-labeling of cysteine (BTFA, TFET, BTFMA) and lysine (SETFA, FINDOLE, FINDCBXY) residues.

Figure 4.7 DFT-based normalized $^{19}$F chemical shifts for each fluorine tag across varying solvent polarities.
Figure S4.1 Diffusion-edited $^{19}$F NMR for 5-fluoroindole and 5-fluoroindole-2-carboxylate labeled HEWL at 35 °C.

Figure S4.2 $^{19}$F NMR spectra for reactions between 5-fluoroindole, formaldehyde and unmethylated HEWL (cyan) and 5-fluoroindole, formaldehyde, and fully $^{13}$C-Dimethylated HEWL (red).

Figure S4.3 $^{19}$F NMR comparisons between the following reactions: (Green) 3 mg Tris(hydroxymethyl)aminomethane in ~90% H$_2$O:D$_2$O with 3 mg 5-fluoroindole and 0.5 uL formaldehyde. (Blue) 3mg BisTris in ~90% H$_2$O:D$_2$O with 3 mg 5-fluoroindole and 0.5 uL formaldehyde. (Red) 3 mg 5-fluoroindole with 0.5 uL formaldehyde in ~90% H$_2$O:D$_2$O.

Figure S4.4 Lysine hydrochloride + 5-fluoroindole + $^{12}$C formaldehyde ESI$^+$ mass spectrum.

Figure S4.5 $^{19}$F NMR spectra of fluoroindole and fluoroindole-carboxylate labeled HEWL versus Temperature. Figure S4.6 $^{19}$F NMR spectra of fluoroindole and fluoroindole-carboxylate labeled HEWL versus pH. 5-fluoroindole (left column) and 5-fluoroindole-2-carboxylate labeled (right column) HEWL. $^{19}$F NMR spectra acquired at pH 6.0 (top), 7.2 (middle), and 8.0 (bottom) at 35 °C. Samples were prepared as described in Figure 1 but each sample was dialyzed into 20 mM K+/Na+ phosphate at the respective pH indicated. Note the improved resolution of resonances for fluoroindole-2-carboxylate labeled HEWL at pH 6.0 (bottom right). Note also that at all pHs investigated, the samples showed no signs of degradation over extended periods of storage at room temperature.

Figure S4.7 Unlabeled HEWL ESI$^+$ mass spectrum.

Chapter 5

Figure 5.1 Structures of cysteine-specific trifluoromethyl probes used in this study.

Figure 5.2 Changes in $^{19}$F chemical shift ($\Delta$δ) of various CF$_3$ tags as a function of solvent polarity.

Figure 5.3 (A) DFT-based normalized chemical shifts for three CF$_3$ tags conjugated to methanethiol (shown in C) and (B) DFT-based normalized average partial charges for $^{19}$F atoms for each of the above three CF$_3$ tags, as a function of solvent polarity.

Figure S5.1 $^{19}$F NMR chemical shifts (normalized to $^{19}$F shifts in H$_2$O) as a function of solvent polarity.

Figure S5.2 Correlation between normalized $^{19}$F chemical shifts and normalized $^{19}$F electrostatic partial charges.

Chapter 6

Figure 6.1 Saturation Spin Diffusion Experiment.
Figure 6.2 Difference spectra (on minus off resonance) for $^{15}$N,$^1$H cross-relaxation/spin-diffusion NMR experiments for Ca$^{2+}$-Calmodulin.

Figure 6.3 Spin-diffusion measurements reflect the rigidity of the protein interior along the temperature denaturation pathway.

Figure 6.4 $^{15}$N,$^1$H HSQC of fluoro phenylalanine-labeled Ca$^{2+}$-Calmodulin at 70 °C.

Figure 6.5 Changes in the degree of the tertiary structure of 3F-Phe CaM as a function of temperature, as monitored by the near-UV ellipticity signature, ⟨θ⟩$_{253−275}$.

Figure 6.6 (A) Temperature dependence of the $^{19}$F solvent isotope shifts, Δδ*(H2O), and (B) Temperature-dependent changes in oxygen accessibility to the protein interior as measured by the temperature dependence of $^{19}$F paramagnetic shifts, Δδ*(O2).

Figure 6.7 The aliphatic aromatic contact region of $^1$H,$^1$H NOESY spectra of 3-F-Phe enriched Ca$^{2+}$-CaM at 30°C (black), 50°C (red), and 70°C (blue).

Figure 6.8 Aliphatic saturation spin diffusion Plots comparing standard $^{15}$N,$^1$H-labeled Ca$^{2+}$-CaM in the presence and absence of 5% TFEtOH.

Figure 6.9 X-ray crystal structure and $^{19}$F NMR spectrum of CaM.

Figure 6.10 Aliphatic Saturation cross-relaxation plots comparing spin diffusion efficiencies at various temperature for several proteins.

Chapter 7

Figure 7.1 Spin diffusion efficiency versus temperature for residues in either the C-terminal domain or N-terminal domain of Ca$^{2+}$-calmodulin.

List of Tables

Chapter 1

Table 1.1 Kex for methyl exchange for the Nα dimethylamino of tetramethyllysine.

Table 1.2 Deprotonation (k$_{PD}$) and protonation (k$_{DP}$) rates associated with the methyl exchange for the Nα of tetramethyllysine.

Table 1.3 Summary of the thermodynamic parameters associated with the Nα methyl exchange for tetramethyllysine.
Table 1.4 DFT-based average $^{13}$C chemical shifts for the $\delta_1$ and $\delta_2$ methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) for each of the three rotameric states (gauche +, trans, and gauche -). Each chemical shift is an average across the three C$\beta$-C$\gamma$ ($\chi^1$) side chain reorientations.

Table 1.5 DFT-based average $^{13}$C chemical shifts for the $\delta_1$ and $\delta_2$ methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) across all nine major side chain conformers (three $\chi^1$ and three $\chi^2$ rotamers).

Table 1.6 DFT-based average $^1$H chemical shifts for the $\delta_1$ and $\delta_2$ methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) for each of the three rotameric states (gauche +, trans, and gauche -). Each chemical shift is an average across the three C$\beta$-C$\gamma$ ($\chi^1$) side chain reorientations.

Table 1.7 DFT-based average $^1$H chemical shifts for the $\delta_1$ and $\delta_2$ methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) across all nine major side chain conformers (three $\chi^1$ and three $\chi^2$ rotamers).

Table 1.8 DFT-based average $^{13}$C and $^1$H chemical shifts and shift differences between inequivalent methyls for the N$\alpha$ of tetramethyllysine across all three $\chi^1$ rotamers.

Table 1.9 Average $^{13}$C and $^1$H chemical shift differences between inequivalent methyls for the N$\alpha$ of dimethylalanine across three major $\chi^1$ rotamers.

Table 1.10 DFT predicted dimethylalanine methyl carbon and methyl proton $\Delta\delta$ values for the protonated and deprotonated states.

Table 1.11 Comparison between the 6-31+G* basis set which includes diffuse functions to 6-31G** for prediction of dimethylalanine methyl carbon $\Delta\delta$ values in the protonated and deprotonated states.

Table 1.12 Methyl exchange ($K_{ex}$) and both protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates versus temperature for the dimethylamino of K1$\varepsilon$ in reductively methylated hen egg white lysozyme.

Table 1.13 Summary of the thermodynamic parameters associated with the methyl exchange for K1$\varepsilon$ dimethyllysine.

Table 1.14 Comparison between the thermodynamic parameters associated with the N$\alpha$ methyl exchange in tetramethyllysine to those associated with the methyl exchange for the salt-linked K1$\varepsilon$ dimethyllysine in HEWL.

Chapter 2

Table S2.1 $^1$H and $^{13}$C chemical shifts and line widths for mono- and dimethyl lysines. Chemical shift and linewidht data is reported for partially labeled HEWL in phosphate buffer, pH 7.4. Where peaks are split, data are reported for both resonances. pKa measurements were determined from chemical shift vs pH titration curves (supplementary) for spectra acquired at 15 °C.
Chapter 3

Table 3.1 Thermodynamic Parameters Associated with the Temperature-Induced Transition from Monomer to Oligomer and Octamer to Large Oligomer for WT and F198S PrP(90-231)\textsuperscript{\textalpha}.

Table 3.2 Exchange Rate Constants Observed for WT and F198S ShaPrP(90-231)\textsuperscript{\textbeta}.

Table S3.1 Estimated line widths (Δv) at 28°C for the most downfield monomeric species (M1), the octamer, and the prefibrillar aggregate, based upon spectral deconvolution and T2 relaxation measurements of F198S ShaPrP(90-231)\textsuperscript{\textbeta}.

Table S3.2 Estimates of the thermodynamic parameters associated with the pressure-induced transition from monomer to oligomer and octamer to large oligomer for WT PrP(90-231)\textsuperscript{\textbeta}.

Chapter 4

Table 4.1 Percent Modification of Proteins as quantified by ESI-MS.

Table S4.1 Relative % solvent exposure for lysine residues in hen egg white lysozyme. Calculated from structure 193L.

Chapter 5

Table 5.1 Fitting parameters (A, x\textsubscript{o} and w) used to interpret the dependence of chemical shift change Δδ of CF3 tags to solvent polarity, using the formula Δδ = A/(1+exp(-(Py-x\textsubscript{o})/w))

List of Abbreviations:

- BSA: Bovine Serum Albumin
- CaM: Calmodulin
- ESI: Electrospray Ionization
- HEWL: Hen Egg White Lysozyme
- NMR: Nuclear Magnetic Resonance
- PDB: Protein Data Bank
- PrP: Prion Protein
- STD: Saturation Transfer Difference
- TFE\textsubscript{EtOH}: Trifluoroethanol
- TML: Tetramethyllysine
- WT: Wild Type
CHAPTER 1

Reductively methylated proteins and dimethyllysine dynamics and kinetics

1.1 The chemistry of lysine amino acid side chains and N-terminal amines in proteins

Lysine is one of three basic amino acids found in eukaryotes. It comprises ~6-7% of the amino acid identity of most proteins (King and Jukes 1969) and often plays a role in protein solubility, stability, catalysis, and/or ligand binding. A survey of all amino acid sequences deposited in the Protein Data Bank (PDB) shows that more than half of all proteins have 4-20 lysines (Figure 1.1).

![Figure 1.1](image-url)

Figure 1.1 Number of lysine and arginine residues across all protein structures in PDB. All ~100,000 FASTA sequences from PDB were screened against the number of lysine residues (left) or number of arginine residues (right) in their sequence using AWK scripts. Sequences with 0 lysines or arginines were excluded from the sample as were those with >100 lysine or arginines. Histogram plotted using gnuplot. Figures produced using data from the protein data bank (PDB) www.rcsb.org.

Lysine side chains exist in two predominant forms at physiological pH - a protonated, cationic state and a deprotonated, neutral state (or free base). The pKa of the side chain amino group has been shown to vary between 9.3 and 13.2, depending on its local environment in the protein and on the pI of the protein (Kesvatera et al., 1996, Gao et al., 2006). In general, however, the pKa is expected to be between 9.5 and 10.6 for most solvent-exposed lysine residues (Means and Feeney 1971, Borisenko et al. 2000, Zhang and Vogel 1993, Gao et al. 2006). Those lysines involved in catalytic mechanisms in which Schiff-base formation is a pre-requisite, tend to exhibit lower pKa's in order to facilitate deprotonation (and thereby activation) of the nucleophilic amine (Gao et al. 2006). The population of the protonated, charged form relative to the deprotonated, neutral form of the amino group in solution can be approximated by the Henderson-Hasselbalch relationship:

\[ \text{pH} = \text{pKa} + \log([\text{NH}_2]/[\text{NH}_3^+]) \]  \hspace{1cm} \text{Eq. (1)}

Assuming a default pKa of 10.5 and at a physiological pH of 7.5, the side chain protonated to deprotonated ratio is roughly 1000:1. As per Eq (1), when the pH is equivalent to the pKa, the sidechain...
is approximately 50 percent deprotonated. This is important because it is the deprotonated state of the amino group that is reactive as a base and nucleophile, whereas the protonated state is generally unreactive toward electrophiles (Figure 1.2).

![Equilibrium between deprotonated and protonated states of lysine sidechain amine.](image)

The same equilibrium exists for Nα-terminal amines in proteins, such as for a terminal glycine amino group, shown below (Figure 1.3). One difference between the lysine sidechain amine and N-terminal amine is that, generally N-terminal amines exhibit a significantly lower pKa (pKa~6.8-9.6) (Grimsley et al. 2009, Means and Feeney 1971, Lundblad 2004). This suggests that at neutral pH, for a pKa of 8, the ratio of protonated to deprotonated states is roughly 10:1 and explains why this amino group is often the most reactive, even though its nucleophilicity is perhaps lower than for Nε side chain amino groups (Gerken et al. 1982, Goux et al. 1984, Larda et al. 2012). One consideration, however, is that the pKa of any given amino group is highly dependent on local environment and significantly lower or higher pKas than the average have been reported (Kesvatera et al. 1996, Lundblad 2004, Schmidt and Westheimer 1971) Hydrogen bonding, salt-linkages, and cation-pi interactions can all have an effect on the pKa of a given amino group (Gao et al. 2006).

![Proton-exchange at the Nα amino group of a terminal glycine residue.](image)

### 1.2 The chemical modification of lysine side chain amino groups.

Lysine side chain and N-terminal amine groups have a free lone pair (non-bonded electrons, NBEs) that serve as the reactive center for acid-base and nucleophilic S_N2 reactions. These amine groups are hard nucleophiles and they preferentially react with hard electrophiles exhibiting either a formal positive charge or a partial but highly polarized bond (Gancarz 1995). The initial S_N2 attack in a reductive methylation reaction is a good example of such a hard nucleophile – hard electrophile interaction. Aldehydes, ketones, and the α-position of α,β-unsaturated carbonyl compounds, for example,
preferentially react with amines over soft nucleophilic thiolate anions (Clayden et al. 2012). Therefore, protein lysine and terminal amine groups will preferentially react with formaldehyde in reductive methylation reactions. Initial $S_N^2$ attack by the nitrogen lone pair, followed by proton transfer and subsequent dehydration result in an iminium ion intermediate (Borch et al. 1971). This intermediate can subsequently be reduced to generate the methyl-amine product (Figure 1.4).

Figure 1.4 Scheme showing proposed mechanistic steps in the reductive methylation of amines.

In the case of side chain amines, the resulting secondary amine (methyllysine) is more basic than it is prior to the addition of the methyl group. The average pKa of a monomethyllysine is roughly 11.2-11.9 indicating a substantial increase in the nucleophilicity following monoalkylation (Bradbury and Brown 1975, Larda et al. 2012, Gerken et al. 1982). In general, the pKa of a side chain functional group can be used as an approximate measure of nucleophilicity (Lundblad 2004, Freedman and Radda 1968) (Figure 1.5).

Figure 1.5 Correlation between pKa and nucleophilicity of amino groups in model compounds. Reprinted with permission from Biochem J. (1968) 108(3): 383–391. Copyright © the Biochemical Society.

Furthermore, for small atoms, especially those in the first and second rows of the periodic table, nucleophilicity mirrors basicity; and so it would be expected that following monoalkylation of a lysine residue, the newly formed monomethyllysine (secondary amine) would be more nucleophilic than the corresponding primary amine (Figure 1.6).
It is therefore somewhat expected that a second round of Schiff base (imine) formation and reduction would occur immediately following monomethylation of a lysine sidechain. This holds true to experiment as dimethylation of lysines is highly favoured over monomethylation, at neutral pH (Larda et al. 2012, Taylor and Webb 2001). It is important to note that given the significant increase in the pKa of the methylamine group, the equilibrium shifts toward a greater number of protonated amine (Larda et al. 2012 Gerken et al. 1982, Bradbury and Brown 1975) (Figure 1.7). As mentioned previously, the protonated amine (with a formal positive charge) is unreactive as a nucleophile. Though more of the amine is protonated following monoalkylation; at physiological pHs and above, it is quite apparent that the reaction still proceeds rapidly to dimethylated state. At lower pHs, however, it is possible to improve mono:dimethyl ratios (Larda et al. 2012). Furthermore, given the pKa differences between Nα terminal amines and lysine side chain amino groups, it is possible to control reaction selectivity for N-terminal labeling by performing reactions at lower pH.

Figure 1.6 Nucleophilicity and pKa trends for primary, secondary, and tertiary amines. R groups represent hydrocarbon chain.
Interestingly, upon dimethylation, the pKa drops to between 10.2-10.5, a value very near to that of unmodified lysines (Gerken et al. 1982, Bradbury and Brown 1975, Larda et al. 2012, Jentoft et al. 1981). One reason for this reduction in pKa following dialkylation may be that the nitrogen center is now sterically hindered. For reductive alkylation reactions performed with small aldehydes and ketones, dialkylation is possible whereas for larger moieties, steric factors prevent anything further than monoalkylation (Means and Feeney 1968).

The mechanism underlying reductive alkylation does not permit trialkylation at amine groups given the inability of a tertiary amine (dimethyl-, dialkyl-lysine) to form an imine intermediate. Following nucleophilic attack, the nitrogen of a tertiary amine has no transferrable proton and therefore no longer has a free lone pair with which to form the Schiff base (Figure 1.8).

For Dimethyllysine of pKa = 10.5,
At pH 6.0 Protonated:Deprotonated = 31622:1
At pH 7.5: Protonated:Deprotonated = 1000:1
At pH 8.0: Protonated:Deprotonated = 316:1
At pH 10.5: Prot:Deprot = 1:1

Figure 1.7 Methylation of lysine side chain amino groups. Mono and dimethyllysine pKas and the associated populations of protonated and deprotonated states at four different pHs according to the Henderson-Hasselbalch equation.
It may be expected that protein function would be significantly perturbed by the addition of methyl groups onto lysine residues. However, crystal structure comparisons between native and fully dimethylated egg white lysozyme indicate little to no structural perturbations (Figure 1.9) (Rypniewski et al. 1993, Vaney et al. 1996).

In fact, methylation is regularly employed as a means by which to enhance the crystallization properties of proteins (Kim et al. 2008). For the addition of larger moieties including ethyl, propyl, isopropyl, benzyl, and cyclohexyl groups, whether perturbations occur, and if they do occur whether they affect function, seems to be highly system dependent. Some proteins are quite amenable the alkylation with large bulky moieties whereas others are quite intolerant (Means and Feeney 1968).

Reductive methylation is highly site-specific for lysine and N-terminal amino groups (Means and Feeney 1968, Cabacungan et al. 1982, Means 1984). Arginine, while having free amine in its guanidino
group, does not appear to be methylated between the pH range of 5-9.5 likely because the formal positive charge is delocalized due to resonance in the guanidino group, making either of the nitrogens less nucleophilic than in the absence of such resonance. Furthermore, the arginine amine is a strong base, having a pKa above 12, meaning that it is almost entirely protonated under the mild conditions employed during reductive methylations (Cai et al. 2008). However, studies employing \(^{14}\)C formaldehyde have demonstrated minor incorporation of radioactivity at arginine and histidine residues indicating that some modification is possible at these sites (Lundblad 1995).

1.3 Other approaches to chemical modification of lysine sidechain amines.

Many reagents exist for the chemical modification of lysine and N-terminal amine groups. In fact, lysines are some of the most common sites on proteins used for site-specific chemical tagging, crosslinking, and fluorophore labeling. Whereas chemical alkylation of lysine amino groups via reductive methylation was considered above, there are still many other approaches to in vitro labeling of lysines including: acylation, carbamylation, guanidation, amidination, reactions with sulfonyl chlorides, isocyanates and isothiocyanates, as well as enzymatic-mediated acylations and methylations (Means and Feeney 1971, Lundblad 2004, Hermanson 2008, Walker 2009, Syed et al. 1992). Some, but not all, of these common amino group bioconjugation reagents are shown below (Figure 1.10)
Acetylations using acetic anhydride are one of the oldest known and most used modifications of amino groups (Fraenkel-Conrat 1957). Unfortunately, all acylations of lysine side chain and N-terminal amines eliminate the positive charge through the formation of a neutral, planar amide bond (Figure 1.11).
Figure 1.11 Acetylation of lysine residues with anhydrides. Note the formation of an amide bond at the side chain amino group which destroys the native charge of the lysine.

The amide bond is often depicted statically (1) but is more appropriately represented by a resonance structure where the lone pair electrons of the nitrogen are delocalized into the π* orbital of the carbonyl carbon and where the nitrogen has a partial positive charge whereas the carbonyl oxygen possesses a partial negative charge (2) (Figure 1.12) (Kemnitz and Loewen 2007).

Figure 1.12 Partial charges and delocalization of electron density across the amide bond for acetylated lysine side chain amino groups.

Whereas the acylation of lysines is a common modification of amino groups, one might argue that the complete conversion of all charged lysine side chain amino groups in a protein to neutral species could be structurally and or functionally perturbing to a protein. Indeed, as with most chemical and enzymatic labeling approaches, the tolerance of a protein toward a given modification must be taken into consideration. A modification that is perturbing to one protein may be perfectly benign to another. Though even if a modification is deemed perturbing, it may still find utility as a method by which to probe reactivity at a given site in a protein; which may vary as a function of conformation and/or topology. This was demonstrated in the use of isotope-coded affinity tags (ICAT) targeting cysteine and lysine side chains for the elucidation of ligand-specific conformations of the complex β2-Adenergic Receptor protein (Kahsai et al. 2011).
Aside from applications involving the incorporation of isotopically labeled chemical tags onto protein amine groups for mass spectrometry and NMR based studies, other common tagging methodologies employ side chain labeling with fluorophores for single molecule fluorescence and/or in vivo localization studies. Many of these fluorogenic tags are sulfonyl chlorides or reactive esters that not only convert the positively charged amino group to a neutral amide but also introduce a large chemical moiety onto the surface of proteins (Figure 1.13).

![Figure 1.13 Two common fluorogenic reagents used to site-specifically label lysine residues in proteins.]

Dansyl Chloride

Alexa Fluor 532 NHS ester-5'

There are numerous examples of studies where lysine amino groups or N-terminal amines are modified by large, hydrophobic chemical moieties. Again, the extent to which function is affected by labeling is highly system dependent and must be tested on a case by case basis. In the design of new chemical tags for site-specific labeling of proteins, it is often most desirable that the chemical tag be small, non-perturbing, highly sensitive (whether via fluorescence yield or large gyromagnetic ratio in the case of isotopic labeling for NMR), permit retention of the native charge at the site of labeling, and be amenable to the mild reactions conditions necessary for preservation of native protein conformations.

Reductive methylation of proteins with $^{13}$C formaldehyde, satisfies all of the above conditions for most proteins and is increasingly used to monitor protein-ligand interactions as well as both hydrogen bond and salt-linked mediated conformational changes (Jentoft et al. 1981, Chavan et al. 2013, Hattori et al. 2013, Bokoch et al. 2010).

1.4 Reductive Methylation of lysine side chain and N-terminal amino groups for $^{13}$C and $^1$H, $^{13}$C NMR of proteins

Several advantages of reductive methylation over other protein labeling methods include: i) size of the probe, ii) retention of native protein charge, iii) minimal perturbation of structure and function, and iv) mild reaction conditions. Reductive methylation has been used since the late 1960s for the modification of proteins (Means and Feeney 1968) and somewhat later was employed as a method for probing intramolecular structure of proteins (Brown et al. 1978, Jentoft et al. 1981). More recently it has proven generally useful for many structural biological investigations, where it is used as a “go-to”
method to enhance crystallization properties of proteins for x-ray crystallography (Tan et al. 2014, Kim et al. 2008). One proposed explanation for enhanced crystallization following methylation is a favourable entropy of solvent reorganization around methyllysines which thereby promote crystal packing (Joachimiak 2010). In addition to crystallographic applications, reductive methylation can also provide insight into protein structure and function via the introduction of small, isotopically labeled methyl tags onto lysine side chain and N-terminal amino groups for NMR. These methyl tags have been used to report on intramolecular salt links, hydrogen bonds, protein conformational changes, ligand binding, protein-protein interactions, and as a way to conformationally “fingerprint” active and inactive states of enzymes/receptors (Brown et al. 1978, Gerken et al. 1982, Jentoft et al. 1981, Zhang and Vogel 1993, Huque and Vogel 1993, Hattori et al. 2013, Bokoch et al. 2010). Many early NMR studies of proteins reductively methylated with $^{13}$C formaldehyde, directly detected the newly introduced $^{13}$C nuclei (Brown et al. 1978, Jentoft et al. 1981, Gerken et al. 1982, Goux et al. 1984). Later, 2-dimensional $^1$H,$^{13}$C NMR experiments were employed, largely because of improved resolution of methyl resonances in a 2D experiment, and because of the improved sensitivity afforded by detection of methyl protons over methyl carbons ($\gamma_H/\gamma_C = \sim 4:1$). Under most circumstances, full $^{13}$C methylation of proteins affords two methyl groups per lysine side chain amino group as well as two methyls per N-terminal amino (Figure 1.14).

![Figure 1.14](image)

**Figure 1.14** Full dimethylation of hen egg white lysozyme (hewl) achieves two methyl groups (red stars) per amino group (A). In the depiction, only lysine residues and the N-terminus are presented along with the “cartoon” of the protein secondary structure. There are 6 lysine side chains and one N-terminal amine in hewl for a total of 7 reactive amino groups. Reductive methylation with $^{13}$C formaldehyde results in the incorporation of 14 methyl groups into the protein. Generally, at high pH (>9) each of the two methyls on a given amino group (or sidechain) exhibit a single resonance due to chemical shift equivalence (shown in the $^1$H,$^{13}$C NMR spectrum (B). Note that the orange asterisk indicates signal originating from noise. Reprinted with permission from J Biomol NMR. (2012) 54:199–209. Copyright © Springer Science+Business Media B.V.

Interestingly, it has been frequently observed that each of the two terminal methyl groups in dimethyllysine side chains exhibit chemical shift inequivalence, resulting in line broadened and/or “split” resonances in $^1$H,$^{13}$C NMR spectra (Figure 1.15, 1.16). This chemical shift inequivalence, as
will be shown, relates to dynamic and kinetic processes of the lysine side chain. An understanding of these processes can give mechanistic insight into structurally or functionally relevant interactions in proteins.

Note that for the remainder of this thesis, the terms “dimethylamine” and “dimethylamino” will refer to molecules with the following substructures: \( \text{RN}^+\text{H(CH}_3\text{)}_2 \) or \( \text{RN(CH}_3\text{)}_2 \), where the nitrogen is either protonated or deprotonated, respectively.

1.5 Introduction to dimethylamine methyl proton and carbon chemical shift inequivalence.

A number of studies on reductively methylated proteins and small molecule dimethylamines demonstrate that both line broadening and/or resonance splitting for methyl resonances typically become more pronounced as pH decreases (Abraham et al. 2009, Gerken et al. 1982, Goux et al. 1984, Larda et al. 2012) (Figure 1.15, 1.16).

However, for reductively methylated proteins, the extent of line broadening is variable among different dimethyllysines and only a subset of the methyllysines exhibit the distinct \( ^{13}\text{C} \) resonance splitting mentioned earlier (Hattori et al. 2013, Goux et al. 1984, Larda et al. 2012, Gerken et al. 1982, Jentoft et al. 1981) (Figure 1.17).
The aforementioned variation in pH-dependence of resonance line broadening correlates well to differences in pKa among the different lysine residues (Abraham et al. 2009, Gerken et al. 1982, Larda et al. 2012, Goux et al. 1984). At lower pHs, dimethyllysine resonances tend to be broad but coalesce as pH is increased. This suggests some correlation to the protonation state of the dimethylamine nitrogen. A more complete description of the mechanisms underlying these line broadening effects will be the subject of subsequent sections (Section 1.7 and Chapter 2).

\(^{13}\text{C}\) resonance “splitting” is an extreme case of slow exchange and is thought to be a consequence of hydrogen bonding and/or salt-bridge interactions between dimethyllysine amino groups and adjacent functional groups in proteins (Zhang and Vogel 1993, Goux et al. 1984, Jentoft et al. 1981, Gerken et al. 1982). Yet, for small molecule dimethyamines in the absence of strong hydrogen bonding interactions and at low pH, \(^{13}\text{C}\) line broadening and/or resonance “splitting” can still occur suggesting that there is more to the description than the slowed methyl exchange due to stable hydrogen bonding interactions (Figure 1.18).

**Figure 1.17** Dimethyllysine spectra for reductively methylated hen egg white lysozyme and ubiquitin. (A) \(^1\text{H},^{13}\text{C}\) spectrum for reductively methylated hen egg white lysozyme at pH 7.5. The most upfield resonance (red arrow) exhibits substantial line broadening in the \(^1\text{H}\) dimension whereas several resonances downfield of 2.6 ppm exhibit resonance “splitting” in the indirect \(^{13}\text{C}\) dimension (black arrows). Reprinted with permission from J Biomol NMR (2012) 54:199–209. Springer. (B) \(^1\text{H},^{13}\text{C}\) spectrum for reductively methylated ubiquitin. Resonances K27, K11, and K29 exhibit line broadening and/or resonance “splitting”. Reprinted with permission from J Biomol NMR. (2012) 54:199–209. Copyright © Springer Science+Business Media B.V.
An overview of chapter 1 content:

To date, there has been no attempt to provide a complete description of the methyl exchange and methyl inequivalence observed in NMR spectra of small molecule dimethylamines and in dimethylamino groups of reductively methylated proteins. These spectroscopic signatures reflect dynamic and kinetic events as well as bonding interactions and are therefore important for understanding how methyl NMR spectra relate to structure and function. The reminder of Chapter 1 focuses on understanding all of these underlying phenomena and developing a model that describes experimentally observed NMR spectra of these species. This has general implications for understanding methyl NMR spectroscopy of small molecules and proteins.

To aid in understanding the order of the forthcoming sections, the following overview will summarize a number of key points of Chapter 1. Section 1.6 introduces the concept of nitrogen inversions, whereby an amino nitrogen center can undergo an umbrella-like interconversion, thereby inverting the stereocenter. This has been known to be a critical process for methyl exchange in dimethylamines and potential reasons for why this is the case will be given (Figure 1.22). Section 1.7 includes a $^{13}$C NMR study of the methyl exchange associated with the Nα-dimethylamino group of $^{13}$C-tetramethyllysine. Tetramethyllysine serves as a suitable model for understanding methyl exchange in small molecules and in dimethyllysines in the absence of protein-specific interactions such as hydrogen bonding. This section attempts to explain the pH and temperature-dependence of the methyl exchange in dimethylamines. Earlier studies have shown that while methyl exchange is due to inversions at the nitrogen center, inversions are rate limited by proton-transfer kinetics. Yet, none of these earlier studies have undertaken a complete thermodynamic description of the methyl exchange process. This section therefore begins by first exploring the microscopic reactions and rates governing proton-exchange at an amino group and then leads into modeling the methyl exchange process via $^{13}$C NMR line shape analyses. The line shape analysis permits estimation of the global rate for the methyl exchange, which can then be dissociated into its component protonation and deprotonation rates. The experimentally determined temperature-dependent methyl exchange rate profile affords a description of the free energy changes associated with the methyl exchange process. The above picture is entirely consistent with the hypothesis of deprotonation-mediated nitrogen inversions governing methyl exchange. Up to this point however, no attempts have been made to understand the origin of the methyl inequivalence in

Figure 1.18 Broadening and splitting of $^{13}$C 1D NMR resonances for the sidechain methyls and N-terminal methyl groups of the small molecule Nε,Nε,Nα,Nα-tetramethyllysine as a function of pH. Figure modified from original (addition of sidechain, and N-terminal labels) published in Biophysical Chemistry 19 (1984) 363-373. Copyright © Elsevier B.V.
dimethylamines. Whereas modeling correctly describes the exchange process, the origin of the non-degeneracy has not been fully addressed. Section 1.8 introduces the concept of asymmetry and proximity of asymmetric centers to the dimethylamino group whereas section 1.9 describes the \( \gamma \)-gauche effect in leucines which results in methyl inequivalence not unlike that seen in protonated dimethylamines. The latter section also addresses methyl exchange in leucine side chains resulting from fast rotameric averaging. Whereas some evidence also exists for slow rotameric exchange, this section, along with a computational study (Section 1.10-1.11) attempt to show why a fast-rotameric exchange hypothesis is more relevant for dimethylamines including dimethyllysine side chains. Section 1.12 attempts to demonstrate computationally that methyl exchange at dimethylamino groups does not simply arise solely from protonation-deprotonation, but that it indeed requires nitrogen inversion. Section 1.13 applies the same methods demonstrated in section 1.7 (\(^{13}\)C line shape simulations) to understanding the methyl exchange for a salt-linked dimethyllysine side chain in \(^{13}\)C-reductively methylated lysozyme. This demonstrates how the prior analyses are relevant to understanding structurally and functionally important interactions in proteins via methyl NMR spectroscopy. Finally section 1.14 addresses \(^1\)H resonance inequivalence in methyl groups for reductively methylated lysine side chains. Together, these results provide a comprehensive description for methyl \(^{13}\)C and \(^1\)H resonance inequivalence and exchange in both small molecule dimethylamines and for reductively methylated proteins. The underlying processes do, however, share similarities to exchange processes in other amino acid side chains and so the first chapter should be useful for further understanding methyl NMR spectroscopy in general.
1.6 Nitrogen inversions

Amines and phosphines are known to undergo what are known as inversions at their nitrogen or phosphorus centers. The inversion involves an umbrella-like motion which interconverts the molecule between two states related through an imaginary plane (Figure 1.19).

**Figure 1.19** Depiction of nitrogen inversion. (A) Inversions of N-ethyl-N-methylpropan-1-amine are preceded by deprotonation of the cation. (B) 3D models depicting inversion of the nitrogen center for the deprotonated state of methylamine. Note that in depiction (B); immediately following inversion, a high-energy state is populated. This is usually remedied by C-N bond rotations which eliminate the steric frustration on a picosecond timescale. The conformers related through inversion are therefore enantiomeric pairs if the nitrogen is a chiral center as in the protonated state shown in (A).

A computational, *ab initio* study on inversions of simple amines and phosphines found that the energy barrier for inversion of ammonia is between 18 and 30 kJ/mol (Kölmel *et al.* 1992). For simple amines, such as trimethylamine, the barrier lies between 25 and 39 kJ/mol though other studies have suggested a higher barrier, around 75 kJ/mol (18 kcal/mol) (Morgan and Leyden 1970). For bulky, hindered tertiary amines, steric constraint enforces a roughly planar geometry and thereby reduces the energy barrier to inversion. Tri-isopropyl amine, for example, exhibits a low barrier to inversion between 0.7 and 3.4 kJ/mol (Kölmel *et al.* 1992).

In addition to small molecule amines, nitrogen inversions also occur for lysine side chain amino groups following base-mediated deprotonation events (Figure 1.20). These inversion processes are thought to be important in characterizing chemical exchange phenomena for dimethyl amines and methylated lysine side chain amino groups (Goux *et al.* 1984, Wilson *et al.* 1995, Saunders and Yamada 1963).
local intramolecular asymmetry in a manner similar to that depicted in the figure below (Figure 1.21).

Indeed, many NMR studies on small molecule dimethyl amines attribute the pH-dependent line spectra of dimethyllysines to the frequency of nitrogen inversion events that occur following deprotonation following a planar intermediate state on the way to the inverted conformer. Protonation after inversion re-establishes the tetrahedral geometry with bond angles of ~109.5°.

It is perhaps attractive to relate the extent of line broadening or peak splitting in 	extsuperscript{1}H or 	extsuperscript{13}C NMR spectra of dimethyllysines to the frequency of nitrogen inversion events that occur following deprotonation at the nitrogen. It is assumed in these cases that nitrogen inversion is responsible for exchange between the non-degenerate (inequivalent) methyl groups on a given dimethylamine: RN\\textsuperscript{+}H(CH\\textsubscript{3})\textsubscript{2}. Indeed, many NMR studies on small molecule dimethylamines attribute the pH-dependent line broadening to altered inversion rates (Saunders and Yamada 1963, Goux et al. 1984, Wilson et al. 1995). Furthermore, several studies on reductively methylated proteins attribute methyl ineqivalence to local intramolecular asymmetry in a manner similar to that depicted in the figure below (Figure 1.21) (Goux et al. 1984, Larda et al. 2012).
they are never truly equivalent in the presence of asymmetry at, and around, the nitrogen center.

The diastereotopic methyl groups do truly exchange chemical environments (Figure 1.22). Whereas in the absence of inversion, each methyl always represents a distinct average between all rotameric states, they are never truly equivalent in the presence of asymmetry at, and around, the nitrogen center.

Figure 1.21 One proposed model for the origin of $^1$H and $^{13}$C line broadening for the methyl resonances of dimethyllysine. Lysine side chains in proteins are usually in the vicinity of other functional groups which create different local chemical environments. It is possible that inversion of the nitrogen center of dimethyllysines would put degenerate methyls in distinct chemical environments, therefore causing chemical exchange. While this is a valid hypothesis, other models (described later) seem to be a better fit to experimental observations.

The nitrogen inversion hypothesis seems more reasonable however, given that upon inversion, the diastereotopic methyl groups do truly exchange chemical environments (Figure 1.22). Whereas in the absence of inversion, each methyl always represents a distinct average between all rotameric states, they are never truly equivalent in the presence of asymmetry at, and around, the nitrogen center.
Figure 1.22 Distinct methyl groups for substituted dimethylamines. Each methyl group for a dimethyllysine side chain is chemically distinct from the other until nitrogen inversion occurs. It may seem that the sampling of rotameric states around the Cε-Nε bond would make the methyls equivalent. While this does serve to average the chemical shifts for each methyl among all rotameric states, the rotamer positions do not truly exchange the non-degenerate methyl groups. In the Newman projections shown, one methyl is labeled blue while the other red to distinguish them from each other. It is evident that no matter which rotameric state is sampled, the methyls do not truly exchange environments. Comparing projections 2 and 2*, this is readily apparent. Arguably, one might assume that 3 and 1* are almost identical. However, it becomes apparent that for side chains with intervening methylene groups between the tetrahedral center (N in this case) and a second asymmetric center (often Ca), the C-N bond angle is angled such that one methyl group is pointed “outward” while another is closer “inward” toward the asymmetric center. Therefore, in the absence of inversion, the methyl groups can be considered chemically inequivalent (non-degenerate).
Given that there is a true “swap” of the methyl positions following inversion, this seems a suitable hypothesis for the description of the chemical exchange process. Similarities in side chain dynamics can be drawn between dimethyllysine methyl groups and the isopropyl moiety of leucine. In the case of leucine side chains, the tetrahedral $C_\gamma$ carbon serves as an asymmetric (chiral) center as does the protonated nitrogen in the case of dimethylamines (Figure 1.23).

![Figure 1.23](image)

**Figure 1.23** Leucine and tetramethyllysine atomic labeling. Leucine (A) and Tetramethyllysine (B) with relevant backbone and side chain atoms labeled in red. Note that the $N_\epsilon$ of lysine side chains is sometimes assigned as $N_\zeta$ in literature, given that the nitrogen is indeed in the zeta position. However, throughout this thesis, the side chain amino nitrogen will be referred to as the $N_\epsilon$ to avoid confusion as prior literature to which many references will be made, uses the $N_\epsilon$ convention. When referencing sources that use the $N_\zeta$ convention, a note will be made to indicate a change in the assignment. The $N_\alpha$ is always designated as the nitrogen immediately adjacent to the $C_\alpha$ chiral center in the backbone.

Unlike a dimethylamine nitrogen, the $C_\gamma$ of leucine cannot undergo base-mediated deprotonation and therefore never undergoes inversions. Interestingly, chemical exchange is still observed for the terminal $\delta CH_3$ groups in leucine and the NMR spectra are remarkably similar to those of certain dimethylamines. Clearly some exchange can occur in the absence of inversion processes and it is likely that the same exchange process also applies to substituted dimethylamino groups (this will be addressed in section 1.9 and pertains to rotameric exchange).

In the following section, however, we focus on protonation-deprotonation kinetics for substituted dimethylamines and how this manifests in pH and temperature-dependent methyl exchange. It will be demonstrated that deprotonations followed by inversions at the nitrogen center are most likely responsible for the pH-dependent coalescence behaviour observed in the NMR spectra of such dimethylamines. An alternative hypothesis in which deprotonations alone are sufficient to eliminate asymmetry between inequivalent methyls is also plausible but will be tested in the later computational section (Section 1.12).
1.7 Studies on $\text{N}^\varepsilon\text{N}^\varepsilon\text{N}^\alpha\text{N}^\alpha$-tetramethyllysine: an analog for dimethylamines in reductively methylated proteins.

1.7.1 Abstract

Lineshape simulations were performed on $^{13}\text{C}$ 1D NMR spectra of $^{13}\text{C}$-labeled Tetramethyllysine (TML) as a function of temperature in order to provide both kinetic and thermodynamic descriptions of the methyl exchange processes at each dimethylamino group. Prior modeling, based on earlier work by Goux et al. 1984, demonstrated that methyl exchange in dimethylamino groups is due to relatively slow base-mediated deprotonations at the nitrogen center, followed by fast nitrogen inversions. The conclusion of this earlier study is that methyl exchange by nitrogen inversion is thus rate-limited by deprotonation events. Both proton-exchange and nitrogen inversion are therefore intrinsically coupled to observed methyl line broadening and pH-dependent exchange effects in NMR spectra of substituted dimethylamines. We demonstrated how inversion at the nitrogen center results in “true” methyl population exchange (Figure 1.22), a concept which will be important when considering methyl exchange in leucines and the origin of methyl inequivalence (in later sections of Chapter 1). Earlier studies, however, have not addressed the origins of the methyl inequivalence in substituted dimethylamines, nor have they provided a comprehensive model linking the multiple underlying kinetic processes to a thermodynamic energy landscape which fully describes the methyl exchange. The subject of this section is to put forth such a model and to demonstrate how it explains experimentally observed spectra of dimethylamino groups in both small molecules and in reductively methylated proteins. This model can be applied to numerous systems and has general implications for understanding methyl exchange NMR spectroscopy.

1.7.2 Introduction

Small molecule lysine salts and their derivatives serve as a good model system for studies of lysine residues in proteins and are important for understanding of the side chain dynamics in the absence of protein-specific interactions, kinetics and dynamics. Goux et al. (1984) performed lineshape analysis of $^{13}\text{C}$ NMR spectra of $\text{N}^\varepsilon\text{N}^\varepsilon\text{N}^\alpha\text{N}^\alpha$-tetramethyllysine (TML) to obtain a kinetic description of the exchange process underlying the $\text{N}^\alpha$ dimethyl resonance broadening (high pH) and splitting (low pH). They found that a model involving base-catalyzed deprotonations of the nitrogen, followed by inversions at the nitrogen center was sufficient to describe the chemical exchange process between the two diastereotopic methyl groups. The assumption that inversion is necessary for methyl exchange is reasonable given that inversions are much faster than the rates associated with deprotonation of a lysine or N-terminal amino group. Therefore, it might seem intuitive that the slow exchange observed for the N-terminal (N$\alpha$) methyls at low pH (~2) is due to a fixation of the protonated (charged) state of the nitrogen. In the absence of deprotonation, inversions do not occur and so presumably neither does methyl exchange. Therefore deprotonation rates set a lower limit (rate limiting) on the overall methyl exchange process (which occurs via nitrogen inversion).
We begin by comparing our rate analysis with those of Goux et al. (1984) to set up a description of individual exchange processes governing the overall proton-exchange rate at dimethylamino nitrogens. The kinetic analysis shown here, while conducted under slightly different conditions that those employed by Goux et al. (1984), demonstrates that at pH 6.0 the methyl exchange rate is consistent with predicted rates for hydrogen-exchange (protonation-deprotonation) at the $\text{N}_\alpha$ of tetramethyllysine at that pH. While these results are entirely congruent with those of the prior study, they alone are not sufficient to exclude the hypothesis that methyl exchange is solely dependent on proton-transfer events at the amino nitrogen (in the absence of inversion). Consequently exchange rates (via line shape analyses) were obtained as a function of increasing temperature in order to provide a complete thermodynamic description of the methyl exchange process. Later discussions (in subsequent sections of this thesis) on leucine methyl exchange and DFT-based $^{13}\text{C}$ and $^1\text{H}$ chemical shift predictions for tetramethyllysine and dimethylalanine provide additional support for the deprotonation-inversion hypothesis and also introduce several other elements to the methyl exchange picture including rotameric exchange and the origins of methyl inequivalence.

1.7.3 Methods

**Preparation of tetramethyllysine.** Tetramethyllysine was prepared by reacting a 10 mM solution of lysine in 20 mM phosphate buffer with a 3-fold molar excess of $^{13}\text{C}$ formaldehyde and equimolar (relative to aldehyde) sodium cyanoborohydride. The final NMR sample contained 10% D$_2$O for lock. pH was not corrected for the presence of D$_2$O.

**NMR.** $^{13}\text{C}$ 1D NMR experiments were conducted at 150 MHz on a Varian INOVA NMR spectrometer equipped with a cryogenic probe. Standard 90 degree pulse widths were 18 µsec at a B1 field of 13.8 kHz.

**Line shape analysis and simulations.** All simulations were done using WinDNMR (Reich 1995). Experimental $^{13}\text{C}$ 1D spectra were first imported into the program and simulated using the DNMR function. Rigid limit separation of $V_A$ and $V_B$ (chemical shifts of each unique methyl group) was set to either 4 ppm (600 Hz) or 5 ppm (750 Hz) given that the experimental spectra were acquired at 150 MHz. A natural linewidth of 3 Hz was assigned to both resonances A and B. Given that resonances A and B correspond to each of the methyl groups of a substituted dimethylamine, the populations were 50% A and 50% B. $K_{\text{ex}} (K_{\text{ex}}=k_{\text{ab}}+k_{\text{ba}})$ was assigned arbitrarily until good fit to the experimental data was achieved. These simulated exchange rates were compared to the predicted exchange rate for proton-transfer at the amino nitrogen, determined using models demonstrated by Goux et al. (1984) and Borisenko et al. (2000).

1.7.4 Results and Discussion

The protonation of the side chain amino group in dimethyllysines is a necessity for methyl inequivalence (Wilson et al. 1995, Gerken et al. 1982, Goux et al. 1984). Deprotonated amino groups are known to rapidly undergo nitrogen inversions, for which the rate of exchange can exceed $10^8 – 10^9$ s$^{-1}$. This is clearly far too fast to give rise to the slow-intermediate chemical exchange observed in the $^{13}\text{C}$ spectra of $\text{N}_\alpha$ dimethylamino groups at low pH. Therefore, it is apparent that for slow exchange to
occur, dimethylamine groups must exist in a protonated state for a sufficiently long period of time. Alternatively, if the lifetimes are short the populations must be weighted heavily toward the protonated state. The lifetime of a protonated lysine can be estimated by considering the mechanisms and kinetics behind deprotonation events at the nitrogen including: i) proton transfers to hydroxide, ii) to water, iii) and to buffer (Borisenko et al. 2000, Bell 1973, Eigen 1964). Borisenko et al. (2000) estimated that for a lysine of pKa 10.5, at a pH between 8 to 9, the lifetime of the protonated state is \(~10-100~\mu\text{sec}\), and between 1-10 \(\mu\text{sec}\) at pH 9-10 (Borisenko et al. 2000). Although the lifetimes are indeed short, the equilibrium populations at pH 8 are heavily weighted to the protonated state.

**Modeling N-terminal (Na) amino group protonation-deprotonation Kinetics**

Figure 1.24 (C), shows a \(^{13}\text{C}\) 1D spectra with a rigid-limit \(\Delta\delta\) for the N\(\alpha\) methyls of TML of \(~4\) ppm at pH 2.0 (the peaks are still broadened indicating that the rigid limit separation is somewhat larger). Furthermore, it has been noted that the chemical shift separation for inequivalent methyls in dimethylamino groups is concentration dependent (Wilson et al. 1995). Therefore it is likely that \(\Delta\delta\) for the N\(\alpha\) methyls could be larger at lower concentrations of TML.

![Figure 1.24](Image)

**Figure 1.24** Structure of \(\text{N}^{\text{N}},\text{N}^{\text{N}}\)-tetramethyllysine in the deprotonated and protonated states (left). Arrows identify resonances in the \(^{13}\text{C}\) 1D NMR spectra arising from either the N\(\alpha\) or N\(\epsilon\) dimethyl groups (right). Figure modified from original (addition of TML structures, arrows) published in Biophysical Chemistry 19 (1984) 363-373. Copyright © Elsevier B.V.

The kinetic model employed by Goux et al. (1984) considered the following proton transfer processes occurring at the N\(\alpha\) dimethylamino group of tetramethyllysine:
\[
\begin{align*}
\text{RNH}_3^- + \cdot\text{OH} & \xrightleftharpoons[k_{-1}]{k_1} \text{RNH}_2^- + \text{H}_2\text{O} \quad (1) \\
\text{RNH}_3^- + \text{H}_2\text{O} & \xrightleftharpoons[k_{-2}]{k_2} \text{RNH}_2^- + \text{H}_3^+ \quad (2) \\
\text{RNH}_3^- + \text{RNH}_2^- & \xrightleftharpoons[k_{-3}]{k_3} \text{RNH}_2^- + \text{RNH}_3^- \quad (3) \\
\text{RNH}_3^- + \text{RCO}_2^- & \xrightleftharpoons[k_{-4}]{k_4} \text{RNH}_2^- + \text{RCO}_2\text{H} \quad (4)
\end{align*}
\]

Where the pKa of the Nα for tetramethyllysine (TML) was determined by titration to be 8.5 (Goux et al. 1984). The rate constants for the four deprotonation events were given as: \(k_1 = 10^{9-10^{12}} \text{M}^{-1}\text{s}^{-1}\), \(k_2 = 1-10 \text{M}^{-1}\text{s}^{-1}\), with upper limits for \(k_3\) and \(k_4\) of \(3 \times 10^6 \text{M}^{-1}\text{s}^{-1}\) and \(10^3 \text{M}^{-1}\text{s}^{-1}\), respectively. The rate constant for nitrogen inversions was assumed to be \(~10^8 - 10^9\text{ s}^{-1}\). Interestingly, it was noted that process 4 need not be considered to explain the pH-dependent trends observed for TML. These NMR experiments were conducted in unbuffered \(\text{H}_2\text{O}\) with 60 mM tetramethyl lysine (Goux et al 1984). Therefore, under these conditions no term for buffer-catalyzed deprotonation is needed.

At pH 6.0, given a pKa of 8.5 for the Nα of tetramethyllysine and using the Henderson-Hasselbalch formula:

\[
6.0 = 8.5 + \log(\left[\text{RNH}_2^-\right]/\left[\text{RNH}_3^+\right])
\]

The ratio of protonated amine (\(\text{RNH}_3^+, P\)) to deprotonated amine (\(\text{RNH}_2^-, D\)) is ~316:1. The fractional populations of state P and D (above) are therefore ca.:

\[p_P = \frac{316}{317} \text{ and } p_D = \frac{1}{317}\]

While not stated explicitly in their analysis (Goux et al. 1984), the effective rate constants for each of the deprotonation processes 1-3 are ca.: 10 s\(^{-1}\), 66 s\(^{-1}\), and 568 s\(^{-1}\), respectively (for \([\cdot\text{OH}] = 10^8 \text{ M}, [\text{H}_2\text{O}] = 55 \text{ M}, \text{ and } [\text{RNH}_2^-] = (0.06\text{M})x(p_B)\) at pH 6.0). Because deprotonation via \(\text{RCO}_2^-\) was deemed by the authors to have little effect on their kinetic fit to the data, it has been neglected here.

Assuming state P is the protonated state and state D is the deprotonated state, then: (Figure).
Where $k_{PD}$ is the overall deprotonation rate, whereas $k_{DP}$ is the protonation rate.

The overall deprotonation rate ($k_{PD}$) is the sum of the individual deprotonation rates for processes 1-3 (process 4 neglected):

$$k_{PD} = 10 \text{ s}^{-1} + 66 \text{ s}^{-1} + 568 \text{ s}^{-1} = 644 \text{ Hz}$$

The overall exchange process is $K_{ex}$, where $K_{ex} = k_{PD} + k_{DP}$.

Given:

$P_e/k_{DP}/K_{ex}$ and $P_D = k_{PD}/K_{ex}$.

$K_{ex}$ for the overall proton exchange process for the $\alpha$ of tetramethyllysine is:

$$K_{ex} = (644 \text{ Hz})/(p_b) = 204148 \text{ Hz or ~200 kHz}$$

Clearly, the exchange between protonated and deprotonated states at pH 6.0, is very fast. However, this is entirely reasonable given the coalescence of the methyl resonances with a rigid limit $\Delta\delta$ of ~4-5 ppm as will be shown shortly.

In order to confirm the applicability of the above kinetic analysis by Goux et al. 1984 to a buffered system (useful when modeling dimethylamino exchange in reductively methylated proteins), we sought to replicate the experimental results under conditions of 20 mM phosphate. Phosphate is known to be one of the most efficient and important catalysts of proton exchange under physiological conditions given its large buffering range and its pKa2 being very near neutral pH (Liepinsh and Otting 1996, Luz and Meiboom 1964):

$$\begin{align*}
H_3PO_4 & \rightleftharpoons H_2PO_4^- & \text{pKa}_1 = 2.15 \\
H_2PO_4^- & \rightleftharpoons HPO_4^{2-} & \text{pKa}_2 = 7.2 \\
HPO_4^{2-} & \rightleftharpoons PO_4^{3-} & \text{pKa}_3 = 12.43
\end{align*}$$

We prepared tetramethyllysine by reacting a 10 mM solution of lysine in 20 mM phosphate buffer with $^{13}$C formaldehyde and sodium cyanoborohydride. The final NMR sample contained 10% D$_2$O for lock. pH was not corrected for the presence of D$_2$O. Note that the effect of additional salt on hydrogen exchange rates is often disregarded due to its small contributions (Liepinsh and Otting 1996). It has therefore, been unaccounted for in the following analysis.
Following methylation, standard $^{13}$C 1D NMR with proton decoupling produced the following spectrum at 150 MHz carbon frequency (600 MHz at $^1$H) (Figure 1.25):

![Figure 1.25 $^{13}$C 1D NMR (at 150 MHz carbon frequency) of reductively methylated lysine, affording tetramethyllysine. The broad resonance at 44.1 ppm corresponds to the non-degenerate $N_\alpha$ methyl groups in exchange. The resonance at 45.28 ppm corresponds to the $N_\varepsilon$ methyls whereas the resonances denoted by red asterisks are residual impurities, which are easily identifiable in two-dimensional HSQC spectra (not shown).](image)

Proceeding with a kinetic analysis as above, we must introduce a new term for phosphate buffer-mediated catalysis of hydrogen exchange, so the processes mediating proton exchange are now:

\[
\begin{align*}
R\overset{\ddagger}{\text{NH}}_3 + \overset{\ddagger}{\text{OH}} & \overset{k_1}{\rightleftharpoons} R\overset{\ddagger}{\text{NH}}_2 + H_2O \quad (1) \\
R\overset{\ddagger}{\text{NH}}_3 + H_2O & \overset{k_2}{\longrightarrow} R\overset{\ddagger}{\text{NH}}_2 + H_3\overset{\ddagger}{\text{O}} \quad (2) \\
R\overset{\ddagger}{\text{NH}}_3 + R\overset{\ddagger}{\text{NH}}_2 & \overset{k_3}{\rightleftharpoons} R\overset{\ddagger}{\text{NH}}_2 + R\overset{\ddagger}{\text{NH}}_3 \quad (3) \\
R\overset{\ddagger}{\text{NH}}_3 + \overset{\ddagger}{\text{BH}} & \overset{k_4}{\rightleftharpoons} R\overset{\ddagger}{\text{NH}}_2 + BH \quad (4)
\end{align*}
\]

Where B corresponds to buffer. For phosphate at pH 6.0, we assume that only $H_2PO_4^{2-}$ and $HPO_4^{3-}$ are relevant exchanging buffer species. R.P. Bell (1973) employs a generalization for proton exchange rates, where $H^+$ transfer between nitrogen or oxygen atoms in the thermodynamically favourable direction is assumed to be diffusion-controlled with a second order rate constant of $10^{10}$ M$^{-1}$s$^{-1}$. The rate constant for the reverse (non-thermodynamically-favourable direction) is estimated using the following relationship:
For,

\[
\begin{align*}
\text{RNH}_3^+ + \text{OH}^- & \rightleftharpoons \text{RNH}_2 + \text{H}_2\text{O} \\
\text{k}_1 & = \text{k}_D \\
\text{k}_1 & = \text{k}_D/10^{(\Delta \text{pK}_a)}
\end{align*}
\]

Where \( \text{k}_D = 10^{10} \text{ M}^{-1}\text{s}^{-1} \), and \( \Delta \text{pK}_a \) is the difference in pKa values between the acceptor and donor species.

Using this relationship, we can estimate the rate constants for each of the deprotonation events for processes 1-4 at pH 6.0 for the \( \text{Na} \) of tetrathemalysine (pKa = 8.5). As before \( p_A=316/317 \) and \( p_B=1/317 \). The concentration of tetrathemalysine in solution is 10 mM. The second order rate constants for processes 1 and 2 are therefore:

\[
\begin{align*}
\text{k}_1 & = 10^{10} \text{ M}^{-1}\text{s}^{-1}, \\
\text{k}_2 & = \text{k}_D/10^{(\Delta \text{pK}_a)}
\end{align*}
\]

Using the formula proposed by Bell (1973) to estimate \( \text{k}_2 \), we obtain:

\[
\begin{align*}
\text{k}_2 & = \text{k}_D/10^{(8.5-(-1.75))} = 10^{0.25} = 0.562 \text{ M}^{-1}\text{s}^{-1}. \text{ Given that } \text{pKa H}_2\text{O} \leftrightarrow \text{H}_3\text{O}^+ = -1.75
\end{align*}
\]

This seems comparable to the \( \text{k}_2 \) estimated by Goux et al., of \( \text{k}_2=1.2 \text{ M}^{-1}\text{s}^{-1} \). Both of these estimates yield reasonable fits in the line shape simulations that follow.

Using the Bell (1973) method to estimate exchange between protonated amine and phosphate oxyanions, however, leads to an overestimate of the effective rate constants for process 4 (\( \text{k}_4 \)). Furthermore, the same approach does not yield satisfactory results for estimates of \( \text{k}_3 \). This is perhaps not surprising as the condition specified by Bell (1973) was exchange between nitrogen and oxygen centers whereas process 3 considers proton exchange between two nitrogen centers. For \( \text{k}_3 \), we resort to using the value reported by Goux et al., (1984), where \( \text{k}_3=3\times10^6 \text{ M}^{-1}\text{s}^{-1} \). This is well below the diffusion limit for proton exchange (\( \text{k}_D \)).

No rate constant for process 4 (proton exchange between amine and buffer) was reported by Goux et al., (1984) given the unbuffered conditions used in their study. Assuming phosphate oxyanions are valid for the method described by Bell (1973) would result in \( \text{k}_4\sim10^6 \text{ M}^{-1}\text{s}^{-1} \). Attempts to use this for estimating \( \text{k}_{AB} \) and the overall \( \text{K}_{ex} \) for proton exchange result in a poor fit to the experimental spectrum of the broad \( \text{Na} \) methyl resonance at pH 6.0 (not shown). Consequently, the second order rate constant for phosphate-catalyzed proton exchange must be smaller and the \( \text{k}= \text{k}_D/10^{(\Delta \text{pK}_a)} \) approach is invalid for exchange between nitrogen and phosphate oxyanions. Bell (1973) does address the fact that phosphate does not conform to this generalization and that numerous factors including: electrostatics and hydrogen bonding can result in rate constants smaller than the diffusion limit.
Liepinsh and Otting (1995) conducted a study of buffer-catalyzed proton transfers between several common amino acids and common buffers. For a 20 mM solution of acetyllysine (where N\(_\alpha\) was acetylated, and so considers only N\(_\varepsilon\) proton exchange) in 80 mM phosphate, the rate constant for buffer-catalyzed proton transfer at pH 6.0 was on the order of 5000-7000 s\(^{-1}\) (Figure 1.26).

![Figure 1.26 Proton exchange rates for several amino acids under various solution conditions. (A) Proton exchange rates for side chain amino group protons for acetyl-lysine-NH\(_2\) in aqueous solution. The proton exchange rate constants are plotted versus the pH for data measured at 20°C (circles), 36°C (squares), and at 36°C in the presence of sodium phosphate at a total concentration of 80 mM (triangles). The concentration of acetyl-lysine-NH\(_2\), was 20 mM. (B) Exchange rates for side chain hydroxyl protons of acetyl threonine-OCH\(_3\) as a function of sodium phosphate concentration at 4°C (circles), 10°C (squares), 20°C (diamonds), and 30°C (triangles). The pH was about 7.0 at room temperature. (C) Exchange rates of the side chain hydroxyl proton of acetyl-tyrosine-NH\(_2\) at 4°C at 6, 12, and 28 mM concentration. Reprinted with permission from Magn Reson Med. (1996) 35(1):30-42. Copyright © John Wiley and Sons.

This rate constant was determined at 36 °C whereas the current analysis considers a 10 mM solution of tetramethyllysine in 20 mM phosphate at 25 °C. From Figure 26 C above, a doubling of concentration results in roughly a doubling of the magnitude of the rate constant. The same doubling of rate also occurs when temperature is increased by ~10 °C. Considering the differences between the current study and that of Liepinsh and Otting (1995), it is not unreasonable to expect a phosphate catalyzed proton transfer rate constant of between 300-500 Hz under the current conditions.

In summarizing the rate constants for processes 1-4 above, we have:
\( k_1 = 10^{10} \text{ M}^{-1} \text{s}^{-1}, \quad k_2 = 1.2 \text{ M}^{-1} \text{s}^{-1}, \quad k_3 = 3 \times 10^6 \text{ M}^{-1} \text{s}^{-1} \)

The effective rates at pH 6.0 (for \([\text{OH}]=10^{-8} \text{ M}, [\text{H}_2\text{O}]=55 \text{ M}, \text{ and } [\text{RNH}_2]=(0.01\text{M})x(p_B)\) ) are:

1. \(100 \text{ s}^{-1}\),
2. \(66 \text{ s}^{-1}\),
3. \(105 \text{ s}^{-1}\),
4. \(~320 \text{ s}^{-1}\)

The overall deprotonation rate \(k_{PD}\) is the sum of all of the above rates for the processes mediating deprotonation (~ 590 Hz) and the lifetime of the protonated state is equal to the inverse of this sum:

\[
\text{Lifetime of protonated state} = \frac{1}{k_{PD}} = ~1.7 \text{ millisec.}
\]

Given the \(p_P\) and \(p_D\) for the \(\text{N}_\alpha\) at pH 6.0, \(K_{ex}\) is estimated to be ~190 kHz. Using these values, we obtain a much better fit to the experimental spectrum (Figure 1.27).

![Figure 1.27 Lineshape simulation #1. \(^{13}\text{C}\) 1D spectra of 10 mM tetramethyllysine in 20 mM phosphate, pH 6.0 acquired at 150 MHz on a Varian INOVA NMR spectrometer equipped with an HCN cryogenic probe (red). Line shape analysis for the \(\text{N}_\alpha\) methyl exchange process using the updated proton-exchange rate for phosphate-catalyzed protonation-deprotonation (blue), superimposed with the experimental spectrum. As before, line shape simulations were done in WinDNMR (J. Chem. Educ., 1995, 72 (12), p 1086).](image)

The residual error in the fit is likely due to one or both of the following:

1. An overestimate for the phosphate catalyzed proton exchange rate constant.
2. The \(\Delta\delta\) at the rigid limit of exchange between the inequivalent methyl groups is greater than the 4 ppm (600 Hz) used in this analysis.

As mentioned previously, Goux et al. employed a concentration of 60 mM for TML in solution. Furthermore, a concentration dependence of \(\Delta\delta\) has been mentioned earlier (lower amine concentration gives larger \(\Delta\delta\)) so it is reasonable to assume that the rigid limit \(\Delta\delta\) is > 4 ppm.
Using a $\Delta \delta = \omega_A - \omega_B = 5$ ppm (750 Hz) while keeping all other parameters the same, we obtain the following fit:

The fit to the experimental spectrum using a rigid limit $\Delta \delta$ of 5 ppm is reasonably good. A further consideration as to whether this assumption is valid will be considered below in the computational section of this thesis.

**A general model for methyl exchange tied to proton-transfer events at a dimethylamino nitrogen**

It is very important to distinguish the methyl exchange rate constants (dictated by the inversion rate) from the protonation-deprotonation rate constants ($k_{DP}$ and $k_{PD}$, respectively). For substituted dimethylamines there are two distinct exchange processes (1. methyl exchange via inversion and 2. proton-transfers, which set a rate limit on (1)) that are intrinsically linked given that deprotonations always precede methyl exchange via inversion. For dimethylamines, the global $K_{ex}$ governing methyl exchange is therefore dictated by protonation-deprotonation kinetics. Whereas for methyl exchange kinetics: $p_A = p_B$ and $k_{BA} = k_{AB}$ (where A and B represent each distinct methyl group), in the case of proton exchange: $p_P >> p_D$ and $k_{PD} << k_{DP}$ (See comparison table below).
Methyl exchange kinetics

\[
A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B
\]

Where A and B are methyl groups 1 and 2 on a dimethylamine:

\[
\text{R} - \text{N}^+\text{CH}_3 \rightleftharpoons \text{A} = \text{CH}_3 \#1 \quad \text{B} = \text{CH}_3 \#2
\]

Note that in the case of $^{13}$C methyl resonance inequivalence, the observed NMR spectrum corresponds to the methyl resonances of the protonated dimethylamino group only. The deprotonated state is minimally populated at pH<8-9, according to the Henderson-Hasselbalch relationship and so is not observed.

The origin of this methyl inequivalence will be considered in the next section.

Na Methyl Exchange Rates versus Temperature for Tetramethyllysine: Thermodynamic Description of the Exchange Process

Given,

\[
K_a = [H^+][A^-]/[HA]
\]

\[
pK_a = -\log(K_a)
\]
\[ \Delta G^\circ = -RT \ln(K_{eq}) \]

\( K_a \) is an equilibrium constant so: \[ \Delta G^\circ = -RT \ln(K_a) \]

The free energy change associated with ionization under non-standard conditions is:

\[ \Delta G_{ioniz} = \Delta G^\circ + RT \ln(Q) \]

Where \( Q = [H^+][A^-]/[HA] \).

When \( Q = K_a \) (this occurs at equilibrium), then the free energy of ionization (\( \Delta G_{ioniz} \)) is 0.

However, the above model considers a system under a static equilibrium and also does not consider non-standard conditions. Whereas the methyl exchange process is a dynamic equilibrium and therefore has a \( \Delta G_{DP} \) associated with the deprotonation event, and a second \( \Delta G_{PD} \) associated with the protonation event. Furthermore, the free energy change between the two states (protonated versus deprotonated) must be considered at the pH relevant to the experimental conditions.

A general free energy diagram describing the methyl exchange process for the Nα methyls of tetramethyllysine is shown below (Figure 1.30).

Figure 1.30 General free energy diagram (left) representing energy associated with the methyl exchange process for substituted dimethylamines which are mediated by proton-exchange and inversion events. An alternate free energy diagram is shown (below) which represents how the line shape analysis “sees” the exchange process (with protonation-deprotonation events tied to inversion). This helps to explain why the \( \Delta G_{DP} \) is not simply equal to \( \Delta G^\circ \). The lineshape simulations do not isolate proton transfers from inversion events given that inversion is what results in methyl exchange.

The free energy of deprotonation is given by: \[ \Delta G_{deprot} = -2.303RT(pH-pK_a) \]

Whereas the free energy of protonation is: \[ \Delta G_{prot} = +2.303RT(pH-pK_a) \]

Given a pKa of 8.5 for the Nα of tetramethyllysine and a pH of 6.0, at 298K, the \( \Delta G_{deprot} = 14.3 \text{ kJ/mol} \), whereas the \( \Delta G_{prot} = -14.3 \text{ kJ/mol} \).
Methyl exchange rates for the exchange between the Na CH3 groups for tetramethyllysine were obtained by line shape simulation of 13C NMR spectra recorded at pH 6.0 as a function of increasing temperature (5, 10, 15, 25, and 35 °C) (Table 1.1). Rates obtained as a function of temperature provide insight into the thermodynamics governing methyl exchange as well as the activation energy for the exchange process.

Table 1.1 Kex for methyl exchange for the Na dimethylamino of tetramethyllysine.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>~Kex* (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>278</td>
<td>35000</td>
</tr>
<tr>
<td>283</td>
<td>50000</td>
</tr>
<tr>
<td>288</td>
<td>80000</td>
</tr>
<tr>
<td>298</td>
<td>160000</td>
</tr>
<tr>
<td>308</td>
<td>380000</td>
</tr>
</tbody>
</table>

* obtained by line shape fitting of 13C spectra acquired at each temperature. Note that $K_{ex} = k_{BA} + k_{AB}$

Given the populations of protonated and deprotonated states of the Na of tetramethyllysine at pH 6.0 where:

![Image of protonated and deprotonated states of substituted dimethylamines]

Figure 1.31 Dynamic equilibrium between protonated and deprotonated states of substituted dimethylamines where $p_p > p_D$ when pH $pK_a$ and where $p_D > p_p$ when pH $pK_a$.

As before, according to Henderson-Hasselbalch at pH 6.0,

$p_p \sim 316/317 = 0.9968$ and $p_D \sim 1/317 = 0.0031$

Given, $k_{PD} = p_D(K_{ex})$ and $k_{DP} = p_p(K_{ex})$
Table 1.2 Deprotonation ($k_{PD}$) and protonation ($k_{DP}$) rates associated with the methyl exchange for the $\text{N}^\alpha$ of tetramethyllysine.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>~K_c $^*$ (Hz)</th>
<th>$k_{DP}$</th>
<th>$k_{PD}$</th>
<th>$1/T$</th>
<th>ln($k_{DP}$)</th>
<th>ln($k_{PD}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>278</td>
<td>35000</td>
<td>34888</td>
<td>110.25</td>
<td>0.003597</td>
<td>10.4599</td>
<td>4.7028</td>
</tr>
<tr>
<td>283</td>
<td>50000</td>
<td>49840</td>
<td>157.5</td>
<td>0.003534</td>
<td>10.8166</td>
<td>5.0594</td>
</tr>
<tr>
<td>288</td>
<td>80000</td>
<td>79744</td>
<td>252</td>
<td>0.003472</td>
<td>11.2866</td>
<td>5.2943</td>
</tr>
<tr>
<td>298</td>
<td>160000</td>
<td>159488</td>
<td>504</td>
<td>0.003356</td>
<td>11.9797</td>
<td>6.2226</td>
</tr>
<tr>
<td>308</td>
<td>380000</td>
<td>378784</td>
<td>1197</td>
<td>0.003247</td>
<td>12.8447</td>
<td>7.0876</td>
</tr>
</tbody>
</table>

*Note that the measured rates are not corrected for any temperature-dependent changes in pH.

Given:

\[ k = A \exp^{-\frac{E_a}{R}}. \]

\[ \ln(k) = (-\frac{E_a}{R})(\frac{1}{T}) + \ln(A) \]

The slope of the Arrhenius plot is equal to $-\frac{E_a}{R}$

$E_a = -R(-6771.3) = 56296 \text{ J/mole} = 56.3 \text{ kJ/mol}$

Given that $E_a = \Delta H^\ddagger + RT$, at 25 °C

$\Delta H^\ddagger = 53.8 \text{ kJ/mol}$

![Figure 1.32 Arrhenius plots for the protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates versus temperature given in Table 1. Arrhenius relationship: \( \ln(k) = \ln(A) + (-\frac{E_a}{R})(\frac{1}{T}). \)]
ln(k/T) vs 1/T from Table 1 for the deprotonation and protonation rates yields the following Eyring plots (Figure 1.33)

![Eyring plots](image)

**Figure 1.33** Eyring plots for protonation (k\(_{DP}\)) and deprotonation (k\(_{PD}\)) rates versus temperature given in Table 1. Erying relationship: ln(k/T)=ln(k\(_B\)/h)-(\(\Delta H^\ddagger\)/RT)+(\(\Delta S^\ddagger\)/R)

Given that the slope = -\(\Delta H^\ddagger\)/R, \(\Delta H^\ddagger\) = 53.8 kJ/mol (As confirmed by the Arrhenius plot above)

Also given that the intercept = ln(A) = (\(\Delta S^\ddagger\)/R)+ln((k\(_B\))/h), where k\(_B\) is the Boltzmann constant and h is the Planck constant.

\(\Delta S_{PD}^\ddagger\) = 36.1 Jmol\(^{-1}\)K\(^{-1}\) whereas \(\Delta S_{DP}^\ddagger\) = -11.8 Jmol\(^{-1}\)K\(^{-1}\)

\(\Delta G_{PD}^\ddagger\) at 25 °C is therefore: \(\Delta G_{PD}^\ddagger_{298K}\) = 43.1 kJ/mol

Whereas \(\Delta G_{DP}^\ddagger\) at 25 °C is \(\Delta G_{DP}^\ddagger_{298K}\) = 57.3 kJ/mol

The table below summarizes the kinetic and thermodynamic parameters (Table 1.3).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_a)</td>
<td>56.3 kJmol(^{-1})</td>
</tr>
<tr>
<td>(\Delta H^\ddagger)</td>
<td>53.8 kJmol(^{-1})</td>
</tr>
<tr>
<td>(\Delta S_{DP}^\ddagger)</td>
<td>36.1 Jmol(^{-1})K(^{-1})</td>
</tr>
<tr>
<td>(\Delta G_{DP}^\ddagger_{298K})</td>
<td>43.1 kJmol(^{-1})</td>
</tr>
<tr>
<td>(\Delta S_{PD}^\ddagger)</td>
<td>-11.8 Jmol(^{-1})K(^{-1})</td>
</tr>
<tr>
<td>(\Delta G_{PD}^\ddagger_{298K})</td>
<td>57.3 kJmol(^{-1})</td>
</tr>
</tbody>
</table>
The $\Delta G^\ddagger$s for the deprotonation and protonation events gives insight into the chemical phenomena underlying the methyl exchange process.

The difference between $\Delta G_{PD}^{\ddagger 298K}$ and $\Delta G_{DP}^{\ddagger 298K}$ should be consistent with the free energy difference between the two states at the given pH (pH=6.0).

$\Delta G_{PD}^{\ddagger 298K} - \Delta G_{DP}^{\ddagger 298K} = 57.3 - 43.1 = 14.2 \text{ kJ/mol}$.

This is consistent with the predicted free energy difference between the two states, as determined by:

$\Delta G_{prot} = +2.303RT(pH-pK_a)$ and $\Delta G_{deprot} = -2.303RT(pH-pK_a)$

$\Delta G_{prot} = -14.3 \text{ kJ/mol} \text{ and } \Delta G_{deprot} = +14.3 \text{ kJ/mol}$

With the thermodynamic parameters known, the energy level diagram that describes the proton-exchange process which is the rate-limiting step for the methyl exchange (via nitrogen inversions) is as shown (Figure 1.34).

![Figure 1.34](image)

Figure 1.34 Completed free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines which are mediated by proton-exchange and inversion events. Where 43.1 kJ/mol corresponds to the free energy change associated with nitrogen inversion whereas proton-exchange is associated with a 14.2 kJ/mol change in free energy.

The lower energy protonated state is consistent with DFT-based comparisons between the energy of the protonated dimethylamine Nα vs the deprotonated state.

1.7.5 Conclusions

In summary, Goux et al., (1984) modeled the exchange between diastereomeric methyl groups, and concluded that the exchange process was the result of base-mediated deprotonations followed by rapid nitrogen inversions. However the authors did not allude to the origin of the chemical shift difference between the two methyl populations apart from noting that it is due to the presence of some asymmetry (the origin of this magnetic inequivalence for seemingly equivalent methyls is considered in the next section). Furthermore, solely modeling the proton-exchange at the nitrogen does not provide sufficient evidence to exclude the possibility that deprotonation events alone are entirely responsible for
the methyl exchange observed. We first replicated the proton-exchange model proposed by Goux et al. (1984) with the inclusion of an additional exchange term describing buffer-mediated proton-transfers given the sample conditions used in this study. We find that the model fits well to our experimental $^{13}$C NMR spectra at pH 6.0. Rates obtained as a function of temperature provide thermodynamic insight into the exchange process and support the deprotonation-inversion hypothesis with a free energy of activation of 43 kJ/mol, consistent with known inversion barriers for substituted amines. Note that the assumption that nitrogen inversions are necessary for exchange between methyl states relies on a model where exchange only occurs when there are an odd number of inversions (an even number would return the methyl groups to their original state) (Morgan and Leyden 1970).

1.8 Intramolecular Asymmetry and Methyl Inequivalence: Trimipramine Maleate

A $^{13}$C NMR study on the dimethyl amine side chain of trimipramine maleate, a small molecule tricyclic antidepressant, investigated the exchange between the methyl groups of the terminal dimethylamine. The rate of exchange between the methyl conformers was sufficiently slow in trimipramine, as to be able to resolve each distinct methyl group on the sidechain amine by both $^1$H and $^{13}$C NMR (resonance 15 and 16, Figure 1.35) (Wilson et al. 1995).

![Figure 1.35](image-url)

**Figure 1.35** Structures of trimipramine (1), imipramine (2), and amitriptyline (3) and both $^1$H and $^{13}$C spectra for (1). (A) $^1$H 1D NMR spectra of trimipramine maleate (0.035 M, CDCl$_3$, 303 K, 300 MHz) with resonance assignments. (B) $^{13}$C 1D NMR spectrum of trimipramine maleate (0.035 M, CDCl$_3$, 303 K, 75 MHz). Asterisks indicate resonances from the maleate counter ion. Reprinted with permission from Magnetic Resonance in Chemistry (1995) 33(5):367-374. Copyright © John Wiley and Sons.
Wilson and Munro (1995) attribute the methyl chemical shift inequivalence to a difference in the local environment for each of the two methyl groups (sites 15, 16 – blue arrows in Figure 1.35) with exchange between them occurring at a rate slower than 1.1x10^{-2} Hz (Wilson et al. 1995). They sought to determine the origin of this methyl inequivalence by comparing the $^1$H and $^{13}$C NMR spectra of trimipramine (1) to imipramine (2), which lacks the methyl at position 17 and discovered that in the absence of a chiral carbon at position 13, the $^1$H resonances of methyl 15 and 16 are equivalent by both $^1$H and $^{13}$C NMR. Furthermore, comparison of trimipramine (1) to amitriptyline (3), which lacks the C-17 methyl group but has a C=C bond at position 12, clearly demonstrates that asymmetry along the side chain is critical for inequivalence of the methyl groups. Unlike trimipramine, amitriptyline (3) exhibits only splitting of $^{13}$C resonances for each of the inequivalent methyls, whereas its $^1$H resonances for methyl group 15 and 16 are equivalent. The fact that resonance splitting was observed in $^{13}$C whereas not in $^1$H suggests that carbon nuclei may be more sensitive to conformational asymmetry (though this may simply be a consequence of the carbon center being closer in proximity to the asymmetric center than are the proton nuclei). The authors conclude that methyl inequivalence arises from the presence of a chiral center in the aliphatic chain and that inversion at the nitrogen center is the process mediating chemical exchange. It does seem clear that for methyl inequivalence to occur, there is a need for an asymmetric (chiral) center either along the side chain or at the backbone (for example, at the Cα) as suggested by earlier studies (Kroschwitz et al. 1969).

At this point, it is perhaps appropriate to propose a general hypothesis for the origin of methyl inequivalence for dimethylamine moieties (such as side chains of dimethyllysines and Nα-terminal dimethylamino groups in proteins). It would seem for methyl inequivalence to occur, two pre-requisites must be satisfied: 1) The nitrogen center to which the methyls are bound must be protonated. In the protonated state methyl exchange by inversion is impeded, therefore resulting in a $\Delta \delta$ between the inequivalent methyl groups. 2) There must be a chiral center to induce asymmetry at some point along the side chain (or backbone) and the closer the nitrogen center is to this chiral site, the more pronounced the methyl inequivalence. The subsequent section on leucine side chain methyl exchange will draw similarities to the protonated dimethyl lysine side chain whereas later computational studies on leucine and tetramethyllysine will highlight the importance of both bond rotations and protonation state of the dimethylamino nitrogen.

1.9 The γ-Gauche Effect for leucine side chain δ-methyls.

As described earlier for trimipramine maleate the natural asymmetry induced from the chirality of carbon at position 13 was sufficient to cause chemical shift inequivalence for both $^1$H and $^{13}$C resonances of the terminal nitrogen methyl groups. This effect has also been known to occur for methyl carbon resonances in polymers (Kroschwitz et al. 1969, Brath et al. 2006, Mulder 2009). Similarly, it is known that the chemical shift inequivalence of seemingly equivalent δ methyl groups in leucine side chains arises from magnetic shielding of the methyl carbon that is in a gauche position relative to the chiral Cα (Mulder 2009). This results in an upfield shift for the gauche methyl relative to the methyl that is trans to the Cα and is known as the γ-gauche effect (Figure 1.36).
This upfield shift for sterically perturbed methyl carbon atoms has been attributed to induced charge polarization along the H-C bonds (which are angularly dependent) (Grant and Cheney 1967). A plot between the force applied along the H-C bond to the associated induced chemical shift perturbation demonstrates a high degree of correlation (Figure 1.37). The effect is thought to manifest as a result of steric polarization of valence electrons (Grant and Cheney 1967).

The \( \gamma \)-gauche effect can result in magnetic shielding of up to 6 ppm for the methyl in the gauche plus position relative to the methyl occupying the trans position, though \( \Delta \delta \) values smaller than 5 ppm are also common (Mulder 2009, MacKenzie et al. 1996, Dalling and Grant 1967, Grant and Cheney 1967) (Figure 1.38).
In contrast to the Nα methyl groups of tetramethyllysine (described above), where either increasing temperature or pH can result in resonance coalescence, leucine side chain δ-methyl resonances are expected to be unaffected by pH and have also been shown to exhibit only weak coalescence with increasing temperature (Figure 1.39).

The weak coalescence with temperature (and in some cases – chemical shift divergence with increasing temperature), is somewhat unexpected if one were to assume that the slow exchange giving rise to the large Δδ values represents switching between gauche plus, trans, and gauche minus rotamers for leucine side chain δCH₃ groups. This latter assumption is perhaps incorrect (at least in many cases), even though it would seem reasonable given the distribution of gauche plus, trans, and gauche minus...
conformers sampled by leucine side chains. In most cases, leucine side chains minimally populate the gauche minus state and sample trans:gauche+ in a roughly 67:33 ratio (Figure 1.40).

The trans:gauche+:gauche- ratio presented above are a guideline and many deviations from these distributions have been noted including some where the trans:gauche distribution is 52:43 and others where minimally populated states represent ~10% of the equilibrium conformers (Lovell et al. 2000, Batchelder et al. 1982, MacKenzie et al. 1996, Mulder 2009).

At this time, it is perhaps important to consider the rates associated with bond isomerizations in leucine side chains. For leucine, the Cγ-Cδ bond rotations (methyl rotations) occur within the ps-ns timescale (Lipari and Szabo 1982, Kay and Torchia 1991, Nicholson et al. 1992, Kjaergaard et al. 2012). Similar to Cγ-Cδ methyl bond isomerizations, the timescale for methylene bond rotation is on the order of 10-100 ps for room temperature liquid alkanes (Fayer 2009). Therefore both methylene and methyl group rotations are too fast to give rise to the apparent intermediate-slow exchange observed in the NMR spectra of leucine δCH₃ groups. Leucine side chain gauche-trans rotamer transitions involve isomerizations around the Cβ-Cγ bond. A 2D IR study done on 1-bromo-2-isocyanato ethane demonstrated that no gauche-trans isomerizations occurred on the timescale of 50 ps, indicating that bulky substituents can significantly affect the barrier to re-orientation, thereby slowing bond isomerizations which might otherwise be assumed to be on the ps-ns timescale as for methyl rotations (Zheng et al. 2006). It is perhaps expected then, that exchange between gauche-trans isomers for leucine might also be slowed relative to unhindered methyl rotations (Cγ-Cδ isomerizations). Yet, while there is a barrier to Cβ-Cγ bond isomerizations, Cδ-Cα couplings have often suggested rapid rotameric averaging between gauche-trans conformers (MacKenzie et al. 1996). These isomerizations are therefore likely still on the picosecond to low nanosecond timescale in most cases; even though most leucine side chains are buried, and would perhaps exhibit increased steric barriers toward reorientation. Indeed, although MacKenzie et al. (1996) showed via 3JCδ-Cα couplings that leucine side chain rotamers for a glycoporphin A transmembrane peptide undergo motional averaging, they recognized from earlier studies on staphylococcal nuclease by Vuister et al. (1993) that some leucines are highly ordered and exhibit 3JCδ-Cα consistent with all-trans or all-gauche rotameric states (MacKenzie et al. 1996, Vuister et al. 1993). Regardless of fast or slow reorientations for leucine side chains, one methyl (gauche to Cα) tends to always exhibit an upfield chemical shift relative to the methyl trans to Cα. What is important
for the understanding of the methyl exchange process is the average chemical shift attributed to either methyl group across all gauche-trans rotameric states.

In summary, it is likely that many, if not most, gauche-trans isomerizations are outside of the high microsecond to millisecond timescales needed to fully explain the apparent slow methyl exchange suggested by NMR spectra of leucine side chain δCH₃ groups, and that rapid rotameric averaging occurs frequently. This leaves the question as to why the ¹³C NMR of leucine side chains seem to exhibit slow chemical exchange and weak temperature-dependent coalescence between the inequivalent δ methyl groups, even for those exhibiting fast rotameric averaging.

It would seem that the slow exchange for the δCH₃ groups observed in standard ¹H,¹³C NMR spectra does not simply represent slow switching between gauche plus, trans, and gauche minus rotamers but instead, a more appropriate model is one that assumes changes in the distribution of gauche -, trans, and gauche + rotamers, leading to modulation of ∆δ between the δ methyl groups. As mentioned earlier, the only way to obtain a “true” exchange between methyl groups at a tetrahedral center would be to “invert” the chiral center (as in the case of nitrogen inversions) as shown in Figure 1.22. Given that tetrahedral carbon centers cannot undergo inversions, it is not unreasonable to assume that the average chemical shifts for either of the δ-methyls in a given leucine side chain would be different under all rotameric conditions (∆δ>0). This hypothesis is consistent with the weak temperature-dependence of δCH₃ resonance coalescence shown in Figure 1.39. Employing the latter hypothesis, as temperature increases, there is only a change in the distribution or “sampling” of gauche plus, trans, and gauche minus conformers due to thermally-induced altered Boltzmann populations and rates of rotamer switching. Such changes in lifetimes and populations associated with the gauche +, trans, and gauche - states leads to an overall change in ∆δ. In some cases, a change in this average rotameric sampling could indeed result in a larger ∆δ, as exhibited by Leu30, whereas using a model that simply assumes slow exchange between gauche and trans conformations would not account for larger ∆δs with increasing temperature (Figure 35).

In summary, the important points regarding the exchange between leucine side chain δ-methyl groups are:

i) Rather than slow exchange between trans and gauche conformers for leucine side chain δCH₃ groups, there is (generally) fast rotameric averaging across all side chain conformers. While there are perhaps few exceptions to this rule, there will typically always be one methyl gauche and another trans to Cα (with a minor population of gauche/gauche), resulting in a measurable ∆δ.

ii) In the absence of inversions at the tetrahedral Cγ center, the (weighted) average rotameric state of Cδ1 and Cδ2 are different (∆δ>0 under most conditions) and “true” exchange of the Cδ1 and Cδ2 populations does not occur. (Though both Cδ1 and Cδ2 can occupy the gauche -, gauche +, and trans positions at different times through rapid bond isomerizations, the tetrahedral Cγ ensures that they effectively remain magnetically distinct under all conditions – see Figure 1.22 for analogous description).

42
The $\Delta \delta$ in $^{13}$C NMR spectra may give the impression of slow exchange under all circumstances, whereas changes in $\Delta \delta$ between Cδ1 and Cδ2 chemical shifts as a function of temperature are simply due to changes in the population distributions and rates of exchange between $\text{trans}$, $\text{gauche}$, and minor conformer states.

In order to provide support to this model, the next section describes computational studies which were conducted on geometry-optimized models of a small molecule L-leucine analog (5-methyl-3-(methylamino)hexan-2-one). Fast motional averaging (rotameric averaging) will be assumed for purposes of explaining the exchange phenomena, though the model is still applicable to those leucine side chains exhibiting conformational restriction to either all-$\text{gauche}$ +, all-$\text{gauche}$ – (unlikely), or all-$\text{trans}$. It is important to note that the model of leucine methyl exchange considered here has direct relevance toward understanding the methyl inequivalence for substituted dimethylamines such as dimethyllysines in reductively methylated proteins.

1.10 DFT-based calculations of $^1$H and $^{13}$C chemical shifts for the $\delta$CH$_3$ groups of leucine side chains

1.10.1 Abstract

DFT-derived $^{13}$C and $^1$H chemical shifts for 9 side chain conformers of (5-methylamino)hexan-2-one were used to model methyl proton and carbon chemical shift inequivalence for the $\delta$CH$_3$ groups of leucine side chains. Chemical shift differences ($\Delta \delta_{C_1-C_2}$) between the non-degenerate methyl groups are consistent with the known $\text{gauche}$ effect whereby the methyl group in a $\text{gauche}$ position exhibits an upfield shift of 4-5 ppm relative to the methyl in the $\text{trans}$ position, whereas the average $\Delta \delta_{C-C}$ across all rotameric states was 1.7 ppm, consistent with experimental spectra of leucine side chain methyls. The average proton chemical shift difference ($\Delta \delta_{\text{H1-H2}}$) between inequivalent methyl groups across all rotameric states was shown to be on the order of 0.2-0.3 ppm, also consistent with NMR. The magnitude of $\Delta \delta_{C-C}$ compared to $\Delta \delta_{\text{H-H}}$ is congruent with the greater chemical shift sensitivity of carbon nuclei.

1.10.2 Methods:

Quantum-mechanical (QM) methods. All calculations were performed using Spartan’10 (Wavefunction Inc). The structure of (5-methylamino)hexan-2-one was first energy minimized using Spartan’s built-in molecular mechanics minimization method. Its equilibrium conformer was then predicted via $ab$ initio methods using the B3LYP hybrid functional with the 6-31G(d) basis set, which has been shown to yield reasonably accurate molecular geometries (Bauschlicher and Partridge 1995). The side chain $\chi$ angles were adjusted for three values of $\chi_1$ and three values of $\chi_2$, for a total of 9 side chain conformations, where all conformations were staggered (anti- or $\text{gauche}$, no eclipsed). The Cα-Cβ-Cγ-Cδ dihedrals were: $\text{gauche}$ minus (+60°,-60°), $\text{gauche}$ plus (+60°,180°), and $\text{trans}$ (-60°,180°), one dihedral per $\delta$CH$_3$. DFT calculations using B3LYP and 6-31G(d,p) were used to obtain isotropic NMR chemical shifts for each of the structures studied. $^{13}$C chemical shifts across the non-degenerate
δCH₃ groups were differenced for all gauche +, trans, and gauche – conformers (3 for each χ¹ angle). The same was done for ¹H chemical shifts.

1.10.3 Results and Discussion:

To determine whether averaging of ¹³C and ¹H chemical shifts for the Cδ₁ and Cδ₂ of leucine side chains across several rotameric states would still yield residual Δδ>0, isotropic ¹³C, and ¹H chemical shifts were predicted via DFT for a small molecule leucine analog (5-methyl-3-(methylamino)hexan-2-one). The isotropic ¹³C chemical shifts for the Cδ₁ and Cδ₂ were differenced for the side chain in the trans, gauche plus, and gauche minus states across three different Cβ-Cγ (χ¹) rotamers (Figure 1.41 B). The same was done for ¹H shifts. In total, ¹³C and ¹H shifts for 9 different side chain conformers were predicted (Figure 1.41 C).
The shift differences between Cδ1 and Cδ2 (∆δ), provide some indication for the expected rigid limit chemical shift separation between the two methyl group 13C resonances and should accurately reflect observed shift differences in experimental 13C NMR spectra. The isotropic chemical shifts for Cδ1 and Cδ2 were averaged across all three trans conformers. The same was done for the three gauche plus and three gauche minus conformers as well (Table 1.4).

<table>
<thead>
<tr>
<th></th>
<th>Average isotropic 13Cδ1 shift (ppm, unreferenced*)</th>
<th>Average isotropic 13Cδ2 shift (ppm, unreferenced*)</th>
<th>∆δ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average gauche -</td>
<td>170.5054</td>
<td>170.0566</td>
<td>0.45</td>
</tr>
<tr>
<td>Average gauche +</td>
<td>170.1020</td>
<td>165.2460</td>
<td>4.86</td>
</tr>
<tr>
<td>Average trans</td>
<td>169.0965</td>
<td>169.3735</td>
<td>0.28</td>
</tr>
</tbody>
</table>

* Note that “unreferenced” refers to the fact that computationally-derived chemical shifts are often compared to a reference compound of known chemical shift. Given that we are only interested in the relative difference in shift between Cδ1 and Cδ2, comparing unreferenced shifts is valid.

The calculated ∆δs clearly indicate that only in the gauche + conformer is there a large difference between the 13C chemical shifts of Cδ1 and Cδ2. Furthermore, the predicted difference matches quite well to the expected magnetic shielding of ~5 ppm for a methyl carbon in the gauche position relative to trans to the Ca (consistent with the gauche effect). In the case of the gauche + conformer above, it is indeed the Cδ2 methyl that is gauche to the Ca, as expected given its more upfield shift. Interestingly, without consideration of the average chemical shifts across the different χ1 bond rotamers, much larger ∆δ values are obtained. This is because for higher energy conformations, chemical shift differences between Cδ1 and Cδ2 are larger. This speaks to the importance of predicting shifts across all major side chain conformations. Arguably, the side chain could populate a much greater number of high energy conformers (for every given degree of rotation about the χ1 and χ2 bonds) but these become less significant as contributors to the ∆δ as they are minimally populated under all relevant temperature regimes (eclipsed conformations would presumably be minimally populated).
Note that the ideal Cα-Cβ-Cγ-Cδ dihedral angles for each of the three conformers are: \textit{gauche} minus (+60°,-60°), \textit{gauche} plus (+60°,180°), and \textit{trans} (-60°,180°). There are two dihedral angles per conformer – one for either δ-CH$_3$ group. In practice, the dihedral angles will often deviate to some extent from the ideal values of 60° or -60° for methyls \textit{gauche} to the Cα, and 180° for methyls \textit{trans} to the Cα, when constructing optimized structures. This occurs because of slight deviations from perfect tetrahedral carbon centers (bond angles adjust to steric strain and bond lengths to various substituents). However, even with variation in bond angles by several degrees off the ideal, the results seem consistent with the trend that the \textit{gauche} plus conformer exhibits the largest Δδ with differences of up to 4-5 ppm, whereas much smaller Δδs occur for \textit{gauche} minus and \textit{trans} rotamer conformers.

The “fast rotamer averaging” hypothesis discussed in the previous section assumed that Cδ1 and Cδ2 do not slowly exchange as might otherwise be expected given the $^1$H,$^{13}$C spectra of leucine side chain methyl resonances. The evidence that has been presented up to this point seems to suggest that Cδ1 experiences an average chemical shift for rapid sampling of \textit{gauche+}, \textit{trans}, and \textit{gauche} – states across all major conformations of the leucine side chain. Similarly, Cδ2 also experiences an average shift dictated by rapid sampling of these three states. However, Cδ1 and Cδ2 generally exhibit distinct chemical shifts given that they never “truly” exchange. It was shown earlier that for “true” exchange to occur, the Cγ would need to undergo an inversion or experience a loss of asymmetry. The result is a difference in chemical shift between Cδ1 and Cδ2 (Δδ > 0) across all rotamers. If the latter hypothesis is correct, then the average shifts presented in Table 1.4, must be averaged across \textit{gauche} -, \textit{trans}, and \textit{gauche} + for Cδ1 and Cδ2 (Table 1.5). Note that this average Δδ across all rotameric states is not weighted toward the \textit{gauche} + and \textit{trans} conformers (~33% and ~66%, respectively) and therefore likely underestimates the true experimental $^{13}$C Δδ.

**Table 1.5** DFT-based average $^{13}$C chemical shifts for the δ1 and δ2 methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) across all nine major side chain conformers (three $\chi^1$ and three $\chi^2$ rotamers).

<table>
<thead>
<tr>
<th>Average rotamer state (across all 9 conformers shown in Figure 1.41)</th>
<th>Average isotropic $^{13}$Cδ1 (ppm, unreferenced)</th>
<th>Average isotropic $^{13}$Cδ2 (ppm, unreferenced)</th>
<th>Δδ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>169.2208</td>
<td>167.4987</td>
<td>1.72</td>
</tr>
</tbody>
</table>

The average shifts for Cδ1 and Cδ2 across the 9 side chain conformers shown above, result in a Δδ of ~1.72 ppm. This $^{13}$C shift difference between the two δ-methyls for a leucine side chain is not unreasonable and is consistent with Δδ values observed in experimental $^1$H,$^{13}$C spectra (Figures 1.36-1.39). However, as mentioned above, this “unweighted” model assumes that each of the 9 leucine side chain conformers is equivalent in terms of lifetime (or sampling distribution is evenly spread across all 9 conformers), and is likely the reason for deviations from experimentally observed Δδ. Weighting toward \textit{gauche} + and \textit{trans} conformers would increase the value of Δδ. Indeed, for the most extreme cases, a Δδ of up to ~5 ppm is possible assuming a distribution of side chain conformations that maximizes time spent in rotameric states with large $^{13}$Cδ1-$^{13}$Cδ2 differences (such as \textit{gauche} +).
Similarly, the \( \Delta \delta \)s for the average proton chemical shifts for the C\( \delta_1 \) and C\( \delta_2 \) methyl groups also indicate that the largest shift difference occurs when the leucine side chain is in the gauche plus conformation (Table 1.6).

**Table 1.6** DFT-based average \(^1\)H chemical shifts for the \( \delta_1 \) and \( \delta_2 \) methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) for each of the three rotameric states (gauche +, trans, and gauche -). Each chemical shift is an average across the three C\( \beta \)-C\( \gamma \left( \chi^1 \right) \) side chain reorientations.

<table>
<thead>
<tr>
<th></th>
<th>Average isotropic ( \delta_1 ) methyl (^1)H shift (ppm, unreferenced)</th>
<th>Average isotropic ( \delta_2 ) methyl (^1)H shift (ppm, unreferenced)</th>
<th>( \Delta \delta ) (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average gauche -</td>
<td>29.3234</td>
<td>29.2488</td>
<td>0.08</td>
</tr>
<tr>
<td>Average gauche +</td>
<td>31.2043</td>
<td>28.6689</td>
<td>2.54</td>
</tr>
<tr>
<td>Average trans</td>
<td>29.4244</td>
<td>31.2396</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Averaging across all 9 conformers as was done earlier for the C\( \delta_1 \) and C\( \delta_2 \) \(^{13}\)C shifts shows that the average separation between the leucine \( \delta \)-methyls in \(^1\)H is \(~0.26 \text{ ppm}\) (Table 1.7).

**Table 1.7** DFT-based average \(^1\)H chemical shifts for the \( \delta_1 \) and \( \delta_2 \) methyl groups of the leucine analog (5-methyl-3-(methylamino)hexan-2-one) across all nine major side chain conformers (three \( \chi^1 \) and three \( \chi^2 \) rotamers).

<table>
<thead>
<tr>
<th></th>
<th>Average isotropic ( \delta_1 ) methyl (^1)H shift (ppm, unreferenced)</th>
<th>Average isotropic ( \delta_2 ) methyl (^1)H shift (ppm, unreferenced)</th>
<th>( \Delta \delta ) (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average rotamer state (across all 9 conformers shown in Figure 1.41)</td>
<td>29.9841</td>
<td>29.7191</td>
<td>0.26</td>
</tr>
</tbody>
</table>

While the predicted \(^1\)H shift differences seems somewhat higher than would be expected given that leucine side chain methyl spectra often demonstrate identical \(^1\)H resonant frequencies for C\( \delta_1 \) and C\( \delta_2 \), it is still within the limits of what can be observed experimentally. Leucine 75 in Glycophorin A (Figure 1.38) is a good example of a large \(^1\)H \( \Delta \delta \) between the inequivalent methyls. Comparing the predicted \(^1\)H and \(^{13}\)C \( \Delta \delta \)s, it is apparent that the carbon nuclei are more sensitive to conformational sampling than are the \(^1\)H nuclei. This observation fits within expectations for nuclear sensitivities to chemical environment given that \(^{13}\)C exhibits a much larger effective range of chemical shifts (~220 ppm) than does proton (~14 ppm).

Up to this point, the DFT-based shift predictions have been restricted to leucine side chain conformers. The intent has been to demonstrate the precise origin of the chemical shift differences observed between inequivalent methyl groups in side chains under conditions of asymmetry for both \(^1\)H and \(^{13}\)C nuclei. The model that has been laid out will now be applied to substituted dimethylamines.
1.11 DFT-based calculations of \(^1\)H and \(^{13}\)C chemical shifts for the Na amino methyl resonances of protonated tetramethyllysine

1.11.1 Abstract

\(^{13}\)C and \(^1\)H chemical shifts for the N\(\alpha\) methyls of tetramethyllysine and dimethylalanine (in the protonated states) were predicted computationally via DFT using the B3LYP hybrid functional with the 6-31G(d,p) basis set in order to predict the magnitude of carbon and proton chemical shift differences (\(\Delta \delta_{C1-C2}\) and \(\Delta \delta_{H1-H2}\) respectively) between non-degenerate methyl groups. The results are consistent with the gauche effect, where one methyl of a dimethylamine experiences a large upfield chemical shift relative to the other due to greater magnetic shielding in one of three staggered rotameric states. These results are highly suggestive of the fact that one of the mechanisms underlying methyl inequivalence in the \(^{13}\)C and \(^1\)H NMR spectra of dimethylamines is the gauche effect.

1.11.2 Introduction

In comparing dimethylamino methyl inequivalence to leucine methyl inequivalence, there are a number of important distinctions to make. For small molecule dimethylamines such as trimipramine, which was considered earlier, the primary distinction is that the nitrogen can adopt both protonated and deprotonated (free base) states. In the protonated state, the nitrogen is a tetrahedral center just as is the C\(\gamma\) carbon of leucine. However, in the deprotonated state, although the nitrogen exhibits near tetrahedral geometry (it becomes trigonal pyramidal with bond angles of \(\sim 107^\circ\)), it can undergo inversions, which are fast on the NMR timescale (\(10^7-10^9\) s\(^{-1}\)) and lead to exchange of the methyl populations. Earlier, it was shown that the methyl exchange for the N\(\alpha\) methyls of tetramethyllysine occurs via base-mediated deprotonation events followed by inversions. In the protonated state, however, the nitrogen is locked in a tetrahedral geometry and the methyl groups are inequivalent due to the gauche effect. Asymmetry is induced by the presence of a chiral or asymmetric center along the side chain (Wilson et al. 1995). For the N\(\alpha\) of tetramethyllysine in particular, this asymmetric center is the C\(\alpha\) the backbone. The DFT results presented here support the hypothesis that methyl inequivalence for substituted dimethylamines does indeed result from the gauche effect, where the largest \(\Delta \delta\) occurs for one of the three major rotameric states.

1.11.3 Methods

Quantum-mechanical (QM) methods. All calculations were performed using Spartan’10 (Wavefunction Inc). The structures of dicationic N\(\alpha\)N\(\alpha\)N\(\varepsilon\)N\(\varepsilon\)-tetramethyllysine and cationic N\(\alpha\)N\(\alpha\)-dimethylalanine were first energy minimized using Spartan’s built-in molecular mechanics minimization method. The backbone equilibrium conformer of tetramethyllysine and the total equilibrium conformer of dimethylalanine were then predicted via ab initio methods using the B3LYP hybrid functional with the 6-31G(d) basis set. For tetramethyllysine the side chain was conformationally optimized independently of the backbone due to the large number of possible conformations for total residue optimization. DFT calculations using B3LYP and 6-31G(d,p) were used to obtain isotropic NMR chemical shifts for the N\(\alpha\) methyl \(^{13}\)C and \(^1\)H nuclei across each of the structures studied. \(^{13}\)C
chemical shifts across the non-degenerate δCH₃ groups were differenced for all three staggered reorientations of the Cε-Nα bond (χ¹). The same was done for the methyl ¹H chemical shifts.

1.11.4 Results and Discussion

DFT was used to obtain isotropic chemical shift estimates for the inequivalent methyl ¹³C and ¹H nuclei for the Nα methyls of tetramethyllysine. As with the leucine δ-methyls, one particular conformer exhibited an unusually large ∆δC₁-C₂, consistent with the gauche effect (Figure 1.42 C1). The estimated shifts for the Nα methyls were averaged across all three side chain conformers (given the assumption of rapid rotomer sampling on the NMR timescale) (Table 1.8). Note that unlike the leucine side chain where bond rotations about χ¹ and χ² must be considered, for Nα dimethyls, only a single Cα-Nα bond rotation must be accounted for (χ¹, Figure 1.42 A). Therefore, rather than 9 conformers as was the case for leucine, only three were assumed relevant to the analysis of the Nα methyl inequivalence.

<table>
<thead>
<tr>
<th>Table 1.8</th>
<th>DFT-based average ¹³C and ¹H chemical shifts and shift differences between inequivalent methyls for the Nα of tetramethyllysine across all three χ¹ rotamers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average rotamer state (across all 3 conformers)</td>
<td>Average isotropic ¹³C1 shift (ppm, unreferenced)</td>
</tr>
<tr>
<td>Average rotamer state for C1 (ppm, unreferenced)</td>
<td>202.0427</td>
</tr>
<tr>
<td>Average rotamer state for C2 (ppm, unreferenced)</td>
<td>25.8967</td>
</tr>
</tbody>
</table>

Figure 1.42 Three major rotamer conformations for Nα of tetramethyllysine. (A) Three rotameric conformers for the inequivalent methyls of the Nα of tetramethyllysine (TML). The relevant CH₃ groups are colored (blue and red). (B) Structure of tetramethyllysine with select bond and atom assignments relevant to the Newman projections in (A). (C) Structures for the three Cα-Nα rotamers for which isotropic ¹³C and ¹H chemical shifts were predicted via DFT using the B3LYP level of theory and the 6-31(d,p) basis set.
As was the case for the inequivalent leucine methyls, the $^{13}$C nuclei of the $\text{N}_\alpha$ methyls exhibit a greater $\Delta\delta$ than do the $^1$H nuclei, as expected. What is perhaps unexpected is the magnitude of $\Delta\delta_{C1-C2}$ of ~10 ppm. What must be taken to account, however, is that this $\Delta\delta$ corresponds to the maximum rigid limit separation. With regard to leucine, it was shown earlier that while the predicted rigid limit separation is ~4-5 ppm, experimentally observed chemical shift differences for $\text{C}_\delta\text{1}$ and $\text{C}_\delta\text{2}$ can range between 0-5.5 ppm (Figures 1.36-1.39). The same is therefore expected for $\text{N}_\alpha$ dimethyls; that the largest experimentally observed shift differences would be on the order of ~10 ppm but the average would tend to be smaller.

In order to determine whether this was a valid assumption, DFT was performed (employing the same level of theory and basis set as described above) on $\text{N}_\alpha\text{N}_\alpha$-dimethylalanine (Figure 1.43).

The $^{13}$C shift difference between the inequivalent methyls is substantially smaller in the case of dimethylalanine than for tetramethyllysine (Table 1.9).

<table>
<thead>
<tr>
<th>Average rotamer state (across all 3 conformers)</th>
<th>Average isotropic $^{13}$C1 shift (ppm, unreferenced)</th>
<th>Average isotropic $^{13}$C2 shift (ppm, unreferenced)</th>
<th>$\Delta\delta$ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average isotropic methyl $^1$H shift for C1 (ppm, unreferenced)</td>
<td>172.5889</td>
<td>170.1272</td>
<td>2.46</td>
</tr>
<tr>
<td>Average isotropic methyl $^1$H shift for C2 (ppm, unreferenced)</td>
<td>27.3655</td>
<td>28.6652</td>
<td>1.30</td>
</tr>
<tr>
<td>Average rotamer state (across all 3 conformers shown in Figure 1.43)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results lend support to the assumption that $\Delta\delta_{C1-C2} < 10$ ppm are reasonable even for the $\text{N}_\alpha$ methyls of tetramethyllysine. It also indicates that for tetramethyllysine, the side chain conformation...
likely influences the magnitude of $\Delta\delta$ for the $\text{N} \alpha \text{ methyls}$ across the three rotamer conformations. For TML, DFT was performed on a structure for which the side chain conformation was set to its equilibrium conformer. However, it is important to note that in solution, the side chain will generally not be fixed rigid in one conformation, but will experience fast rotations around all methylene carbon-carbon bonds (according to a fast rotamer exchange hypothesis). This is expected to significantly affect the magnitude of $\Delta\delta$ and similar bond-orientation effects have been reported previously (Grant and Cheney 1967). Therefore, it is perhaps not surprising that the experimental $\Delta\delta$ for TML at the rigid limit was $\approx 5$ ppm less than that predicted by DFT. Conditions, which lock the side chain into one conformer would perhaps be expected to give rise to larger $\Delta\delta$s depending on the bond orientations.

1.12 Methyl exchange and $\Delta\delta$ for protonated vs deprotonated states of dimethyl amines

1.12.1 Abstract

DFT-derived $^{13}\text{C}$ and $^1\text{H}$ chemical shifts for protonated and deprotonated states for the $\text{N} \alpha \text{ methyls}$ of dimethylalanine show that upon deprotonation, the magnitude of $\Delta\delta_{\text{C1-C2}}$ is significantly increased whereas $\Delta\delta_{\text{H1-H2}}$ decreases slightly. This is indicative that deprotonation events alone (in the absence of inversions) are not sufficient to explain experimentally observed methyl resonance coalescence (decreasing $\Delta\delta_{\text{C-C}}$) with increasing pH for substituted dimethylamines (such as TML). If deprotonation alone were responsible for the methyl exchange, one would expect a reduction in $\Delta\delta_{\text{C-C}}$ upon removal of a proton from the amino nitrogen. Interestingly, a comparison of DFT calculations performed using basis sets lacking diffuse functions report an opposite trend but can be excluded given their inconsistency with observed chemical shift changes with pH.

1.12.2 Methods

Quantum-mechanical (QM) methods. All calculations were performed using Spartan’10 (Wavefunction Inc). The structure of cationic $\text{N} \alpha \text{N} \alpha$-dimethylalanine was first energy minimized using Spartan’s built-in molecular mechanics minimization method. Three sets of DFT calculations were performed using 6-31G(d,p) and three using the 6-31+(d) basis set for both the protonated and deprotonated states of dimethylalanine in three staggered rotamer conformations (Figure 42). The inclusion of diffuse functions was deemed necessary in these DFT calculations given the anionic nature of the deprotonated amine. Comparisons between each basis set were made to demonstrate the importance of diffuse functions for predicting $\Delta\delta_{\text{C1-C2}}$ and $\Delta\delta_{\text{H1-H2}}$ for the deprotonated state. Isotropic NMR chemical shifts were obtained for the $\text{N} \alpha$ methyl $^{13}\text{C}$ and $^1\text{H}$ nuclei for the protonated and deprotonated state of the three major $\chi^1$ rotamers (Figure 1.45).
1.12.3 Results and Discussion

DFT was conducted on Na-protonated and Na-deprotonated states of dimethylalanine (TML) for the three $\chi^1$ rotamers (Figures 1.44, 1.45).

**Figure 1.44** Gauche plus, trans and gauche minus rotamers for Na dimethylamino groups. Unlike leucine side chain conformers where the gauche-trans is defined relative to the Cα, for N-terminal dimethylamino groups, the gauche-trans convention employed here is defined by the position of the terminal methyl groups relative to the Cβ.

**Figure 1.45** Gauche plus, trans and gauche minus rotamers for NaNα-dimethylalanine in the protonated (top) and deprotonated (bottom) states.

As in the case of leucine methyls, where the gauche plus conformer results in the largest $^{13}$C shift difference ($\Delta\delta_{C1-C2}$) between the methyl carbons, the protonated conformer exhibiting the largest $\Delta\delta_{C1-C2}$ for the NaN methyls of TML is that for which one methyl is in a “gauche plus” orientation relative to the
Cβ of the side chain (Figure 1.44). This is perhaps surprising given that the Cβ is not considered an asymmetric center (not a chiral carbon as is the Cα). However, the anisotropy induced by fixing the side chain into its equilibrium conformation is likely sufficient to induce asymmetry at the Cβ. Note that only one methyl rotamer conformation for Cβ was considered here, whereas in solution, the Cβ methyl would be an average of all rotamers through rapid (nano-picosecond) Cα-Cβ bond isomerizations.

Considering only the DFT performed using the 6-31+G* basis set, it is surprising that upon deprotonation, there is a very large change in the magnitude of ∆δ (~80 ppm) between the two Nα methyl 13C resonances (Table 1.10).

Table 1.10 DFT predicted dimethylalanine methyl carbon and methyl proton ∆δ values for the protonated and deprotonated states, using the 6-31+G* basis set which includes diffuse functions.

<table>
<thead>
<tr>
<th>6-31+G* basis set</th>
<th>Average 13C1 shift (ppm, unreferenced)</th>
<th>Average 13C2 shift (ppm, unreferenced)</th>
<th>∆δ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Protonated</td>
<td>145.9713</td>
<td>143.7931</td>
<td>2.18</td>
</tr>
<tr>
<td>Average Deprotonated</td>
<td>261.215</td>
<td>180.5393</td>
<td>80.68</td>
</tr>
</tbody>
</table>

*Note that as before, reported isotropic shifts are unreferenced.

Comparing the above results to those obtained using the 6-31G** basis set, it is apparent that the inclusion of diffuse functions changes the results substantially for the deprotonated state conformers (Table 1.11).

Table 1.11 DFT predicted dimethylalanine methyl carbon and methyl proton ∆δ values for the protonated and deprotonated states, using the 6-31G** basis set which lacks diffuse functions. Compare the results to Table 1.10.

<table>
<thead>
<tr>
<th>6-31G** basis set</th>
<th>Average 13C1 shift (ppm, unreferenced)</th>
<th>Average 13C2 shift (ppm, unreferenced)</th>
<th>∆δ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Protonated</td>
<td>172.5889</td>
<td>170.1272</td>
<td>2.46</td>
</tr>
<tr>
<td>Average Deprotonated</td>
<td>110.6412</td>
<td>110.3489</td>
<td>0.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6-31G** basis set</th>
<th>Average 'H1 shift (ppm, unreferenced)</th>
<th>Average 'H2 shift (ppm, unreferenced)</th>
<th>∆δ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Protonated</td>
<td>27.36548</td>
<td>28.66517</td>
<td>1.30</td>
</tr>
<tr>
<td>Average Deprotonated</td>
<td>34.96824</td>
<td>35.07777</td>
<td>0.11</td>
</tr>
</tbody>
</table>
The $\Delta \delta$ between methyl carbons is reasonably consistent between either basis set for the protonated dimethylalanine structures (6-31G* is also consistent with the results for the protonated state, not shown). However, the magnitude of $\Delta \delta_{C-C}$ is surprisingly large for the average deprotonated state as determined via 6-31+G*. At first, this might seem to be an indication of error but it should be noted that diffuse functions are important for accurate DFT calculations for anions. An additional check for whether the result is sensible or not is to compare the direction of experimentally observed changes in the $^{13}$C and $^1$H shifts of dimethyl groups with increasing pH. Whereas the 6-31G** dataset predicts decreasing (upfield shift) $^{13}$C chemical shifts upon deprotonation, the 6-31+G* predicts the opposite. Experiment shows that as pH increases, $^{13}$C shifts for dimethyllysine resonances increase (move downfield), suggesting that the latter basis set is a more accurate predictor of $^{13}$C shifts for methyl groups bonded to an anionic nitrogen. As for the direction of proton chemical shifts upon deprotonation, neither basis set seems to accurately predict the direction. However, whereas 6-31G** would not be expected to accurately predict shifts for anionic dimethylalanine, perhaps the lack of additional polarization p-functions on hydrogen for the 6-31+G* basis set makes it less suitable for predicting accurate $^1$H shifts. Overall, the result indicates that deprotonation effectively increases the magnitude of $\Delta \delta$ relative to the protonated state. It is therefore readily apparent that deprotonations in the absence of nitrogen inversions would not give rise to the coalescence behavior seen with increasing pH. Nitrogen inversion is therefore critical for the methyl exchange process for substituted dimethylamines.

1.12.4 Conclusions

A summary of all DFT data presented above provides strong support for the hypothesis that methyl inequivalence in substituted dimethylamines is due to the gauche effect, as it is for leucine $\delta$-CH$_3$ groups. Simulations of the leucine analog, (5-methylamino)hexan-2-one accurately reproduce the large $\Delta \delta$ between the $\delta$-methyl $^{13}$C resonances expected for the gauche plus rotamer. The results suggest that under all rotamer conditions, leucine $\delta$-methyls (C$\delta$1 and C$\delta$2) will generally exhibit a non-zero $\Delta \delta$. Furthermore, the apparent slow chemical exchange between leucine $\delta$-methyl groups is a manifestation of this non-zero $\Delta \delta$, even though rotamer sampling between gauche-trans isomers (via C$\beta$-C$\gamma$ bond isomerizations) is generally fast (picosecond to nanosecond). The weak temperature dependent coalescence of leucine $\delta$-methyls can be explained by simply assuming temperature-dependent changes in the population distributions of gauche+, gauche-, trans, and other minor conformers – which thereby modulate $\Delta \delta$. Under this model, redistribution of the populations toward the gauche + conformer with increasing temperature would indeed result in an increase in $\Delta \delta$ as is seen experimentally for some select leucine side chains. This is in contrast to a model where there is initially slow exchange between rotameric states. Under such a model there would be a pronounced reduction in $\Delta \delta$ with increasing temperature as the rate of rotamer switching between gauche+, gauche-, and trans increases (especially given that the largest $\Delta \delta$ occurs for gauche+). It is expected that the fast rotameric exchange hypothesis holds true for dimethylamino groups as DFT performed on gauche-trans rotamers of the protonated state of tetramethyllysine show similar results.
Earlier it was shown via NMR and line shape analyses that the Nα methyls of tetramethyllysine exhibit substantial coalescence as a function of temperature. This coalescence is also highly sensitive to pH as well. This behavior indicates a second exchange process that is unique to substituted dimethylamines (does not occur for leucine side chain methyls). This additional exchange process involves base-mediated deprotonations at the nitrogen center followed by rapid inversions ($10^7$-$10^9$ s$^{-1}$) which exchange the methyl populations. DFT performed on the protonated and deprotonated states of dimethylalanine indeed show that deprotonation alone is unlikely to account for experimentally observed methyl resonance coalescence, given the increase in $\Delta\delta$ predicted upon proton removal. Proton-transfers therefore establish the rate limiting step for methyl exchange even though the actual exchange process involves nitrogen inversion.

The section that follows will attempt to relate the computational results for Nα methyl inequivalence to Nε methyl exchange for dimethyllysine side chains. Experiments conducted on reductively methylated hen egg white lysozyme demonstrate that the methyl inequivalence occurring for Nε dimethylamino groups is also consequence of gauche like effects. However, for hydrogen bonded and salt-linked lysines, methyl exchange is much slower than that seen via earlier experiments on Nα dimethylamines. The reasons for the slowed methyl exchange will be explained.

1.13 $^{13}$C resonance inequivalence for Nε dimethylamino methyls of lysine side chains in reductively methylated proteins

1.13.1 Abstract

$^{13}$C NMR spectra of the K1ε dimethylamino methyl resonances (of hen egg white lysozyme) were acquired as a function of increasing temperature (5, 10, 15, 20, 25, 30, 35 °C) in order to obtain rates for the methyl exchange process in the presence of a known salt-link interaction. Unlike Nα methyl exchange considered earlier in this thesis, which was in the intermediate-fast NMR time scale, the salt-linked Nε dimethylamine exhibits slow-intermediate exchange. The slowed exchange is presumably due to a salt-bridge interaction with the E7 side chain carboxylate. Line shape simulations of the experimental $^{13}$C NMR spectra provide insight into the thermodynamics of the methyl exchange process.

1.13.2 Introduction

Earlier it was shown that the methyl resonances of lysine residues in $^{13}$C-reductively methylated proteins occasionally exhibit inequivalence (splitting) in $^{13}$C 1D NMR spectra. This inequivalence is also evident in proton-detected heteronuclear NMR spectra ($^1$H,$^{13}$C HMQCs and HSQCs) (Figures 1.15-1.17). This is perhaps surprising given the earlier study on trimipramine, imipramine, and amitriptyline maleate which demonstrated that dimethylamino group proximity to an asymmetric center (Cα in the case of TML) is critical for methyl group inequivalence (Wilson et al. 1995). Unlike Nα-dimethylamino groups, the terminal methyls for Nε-dimethyllysine are 6 bonds away from the chiral Cα center and would therefore be unlikely to experience sufficient asymmetry to cause methyl inequivalence (Kroschwitz et al. 1969, Lindeman and Adama 1971). Indeed, most Nε-dimethyllysines typically do not exhibit $^{13}$C methyl resonance inequivalence. A number of earlier studies have shown that those lysines
exhibiting such $^{13}$C methyl resonance inequivalence are distinct given that they are involved in salt-bridge interactions with adjacent glutamate, aspartate, and C-terminal carboxylate groups (Goux et al. 1984, Jentoft et al. 1981, Zhang and Vogel 1993, Gerken et al. 1982, Larda et al. 2012). Indeed, the presence of a salt-link might be expected to slow proton-transfer at dimethyllysine Nε amino groups by increasing the pKa of the tertiary amine. The hydrogen bond between a salt-bridged amino group and an adjacent carboxylate most likely shifts the equilibrium toward the protonated state and/or extends the lifetime of the protonated state. As has been shown in previous sections/chapters, protonation-deprotonation kinetics are primarily responsible for dimethylamine methyl exchange. It would therefore be expected that any process which slows the overall $K_{ex}$ for protonation-deprotonation at the amino nitrogen would cause line-broadening and in extreme cases, full resonance “splitting”. Earlier studies brought attention almost exclusively to those dimethyllysines involved in salt-bridges because of the extreme cases of methyl inequivalence observed for these residues. Several more recent investigations have presented $^1$H, $^{13}$C NMR spectra where a multitude of resonances exhibit extensive $^{13}$C and $^1$H line broadening.

From the work presented in previous sections/chapters of this thesis, it is reasonable to assume that those lysines exhibiting $^{13}$C and $^1$H line broadening but not full resonance splitting, are involved in weaker hydrogen bonding interactions (sidechain-backbone, sidechain-sidechain). Strong hydrogen-bonding interactions such as salt-bridges tend to cause full methyl resonance splitting in the carbon dimension. It might be possible, therefore, to correlate the extent of methyl $^{13}$C $\Delta\delta$ to the strength of hydrogen bonding interactions between a dimethyllysine and an adjacent functional group in the protein. The strongest interactions would exhibit the largest $^{13}$C $\Delta\delta$s. Though more realistically, it would be necessary to account for differences in solvent accessibility as well as other factors that could affect the protonation-deprotonation kinetics at a given dimethylamino nitrogen.

### 1.13.3 Methods

**Preparation of reductively dimethylated hen egg white lysozyme.** A hen egg white lysozyme was dissolved in 20 mM phosphate pH 6.0 to a concentration of 0.5 mM. Aqueous $^{13}$C formaldehyde was added to a stoichiometry of 5 moles per mole of reactive amino groups (number of lysines + number of n-terminal amines). After 30 minutes at room temperature, solid sodium cyanoborohydride was added at a stoichiometry of 2:1 per moles of aldehyde. Removal of byproducts was achieved by dialysis. The final NMR sample contained 10% D$_2$O for lock. pH was not corrected for the presence of D$_2$O.

**NMR.** $^{13}$C NMR experiments were conducted at 150 MHz on a Varian INOVA NMR spectrometer equipped with a cryogenic probe tunable to $^1$H, $^{13}$C, $^{15}$N and $^{19}$F. Standard 90 degree pulse widths were 18 $\mu$sec at a B1 field of 13.8 kHz for $^{13}$C.

**Line shape analysis and simulations.** All simulations were done using WinDNMR (Reich 1995). Experimental $^{13}$C 1D spectra were imported into WinDNMR and simulated using the DNMR function. A natural linewidth of 6 Hz was assigned to both resonances A and B. Given that resonances A and B correspond to each of the methyl groups of a substituted dimethylamine, the populations were set to 50% A and 50% B. $K_{ex}$ ($K_{ex} = k_{ab} + k_{ba}$) was assigned arbitrarily until good fit to the experimental data
was achieved. The rigid limit separation between resonances A and B \((v_A - v_B = 38 \text{ Hz})\) was defined by low temperature \((5 \, ^\circ\text{C})\) experiments where the methyls were found to exhibit slow exchange.

### 1.13.4 Results and Discussion

Methyl exchange rates for the exchange between the K1ε methyl groups for reductively dimethylated lysozyme were obtained by line shape simulation of \(^{13}\text{C}\) NMR spectra as a function of increasing temperature \((5, 10, 15, 20, 25, 30, 35 \, ^\circ\text{C})\) (Figure 1.46, Table 1.12).

![Figure 1.46](image)

Earlier studies by Larda et al. (2012) demonstrated that the pKa for the K1ε dimethyllysine was >10.4. It was difficult to obtain an accurate estimate of the pKa for this residue due to substantial line broadening at intermediate to low pH, as has been noted by others (Gerken et al. 1982). As will be shown, a pKa of 10.5 is a reasonable assumption for this particular dimethylated lysine residue.

The general free energy diagram for the methyl exchange process for K1ε is identical to that for the Nα methyl exchange of tetramethyllysine (Figure 1.47):

![Figure 1.47](image)
Yet, given the differences in pKa between the Nα dimethylamino group of tetramethyllysine (pKa = 8.0) and the salt-linked Nε dimethylamino group of K1ε in lysozyme (HEWL), it is expected that the thermodynamic parameters will be different.

The free energy of deprotonation is given by: \[ \Delta G_{\text{deprot}} = -2.303RT(pH-pK_a) \]

Whereas the free energy of protonation is: \[ \Delta G_{\text{prot}} = +2.303RT(pH-pK_a) \]

Given a pKa of 10.5 and a pH of 6.0, at 298K the \[ \Delta G_{\text{deprot}} = 25.7 \text{ kJ/mol} \], whereas the \[ \Delta G_{\text{prot}} = -25.7 \text{ kJ/mol} \].

Table 1.12 Methyl exchange (Kex) and both protonation (kDP) and deprotonation (kPD) rate constants versus temperature for the dimethylamino of K1ε in reductively methylated hen egg white lysozyme.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Kex* (Hz)</th>
<th>ln(kDP/T)</th>
<th>ln(kPD/T)</th>
<th>1/T (K⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>278</td>
<td>15</td>
<td>-2.9196</td>
<td>-13.2812</td>
<td>0.0036</td>
</tr>
<tr>
<td>283</td>
<td>25</td>
<td>-2.4266</td>
<td>-12.7882</td>
<td>0.0035</td>
</tr>
<tr>
<td>288</td>
<td>53</td>
<td>-1.6927</td>
<td>-12.0543</td>
<td>0.0035</td>
</tr>
<tr>
<td>293</td>
<td>90</td>
<td>-1.18039</td>
<td>-11.542</td>
<td>0.0034</td>
</tr>
<tr>
<td>298</td>
<td>215</td>
<td>-0.32649</td>
<td>-10.6881</td>
<td>0.0034</td>
</tr>
<tr>
<td>303</td>
<td>500</td>
<td>0.500844</td>
<td>-9.86079</td>
<td>0.0033</td>
</tr>
<tr>
<td>308</td>
<td>1200</td>
<td>1.359945</td>
<td>-9.00169</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

*estimates obtained from line shape simulations. Note, once again that \( k_{DP} = (p_D)(K_{ex}) \) and \( k_{PD} = (p_P)(K_{ex}) \), where \( p_D \) and \( p_P \) are the fractional populations of the deprotonated and protonated states of the K1ε dimethylamino group (pKa = 10.5) at pH 6.0.

As before, given:

\[ k = A \cdot \exp^{E_a/RT} \]

\[ \ln(k) = (-E_a/R) \cdot (1/T) + \ln(A) \]

The slope of the Arrhenius plot (not shown) is equal to \(-E_a/R\)

\[ E_a = -R \cdot (-12234) = 104143 \text{ J/mole} = 104.1 \text{ kJ/mol} \]
Given that $E_\theta = \Delta H^\ddagger + RT$, at 25 °C

$\Delta H^\ddagger = 101.7 \text{ kJ/mol}$

Eyring plots for the protonation ($k_{DP}$) and deprotonation ($k_{PD}$) rates provide additional thermodynamic insight into the methyl exchange for the K1ε dimethyllysine (Figure 1.48).

Given,

$\ln(A) = (\Delta S^\ddagger / R) + \ln((k_B) / h)$, where $k_B$ is the Boltzmann constant and $h$ is the Planck constant.

$\Delta S_{PD}^\ddagger = 55.9 \text{ Jmol}^{-1}\text{K}^{-1}$ whereas $\Delta S_{DP}^\ddagger = 142.1 \text{ Jmol}^{-1}\text{K}^{-1}$

$\Delta G_{PD}^\ddagger$ at 25 °C is therefore: $\Delta G_{PD}^\ddagger_{298K} = 85.0 \text{ kJ/mol}$

Whereas $\Delta G_{DP}^\ddagger$ at 25 °C is $\Delta G_{DP}^\ddagger_{298K} = 59.4 \text{ kJ/mol}$
The table below summarizes the kinetic and thermodynamic parameters (Table 1.13).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_a$</td>
<td>104.1 kJmol$^{-1}$</td>
</tr>
<tr>
<td>$\Delta H^\ddagger$</td>
<td>101.7 kJmol$^{-1}$</td>
</tr>
<tr>
<td>$\Delta S_{DP}^\ddagger$</td>
<td>142.1 Jmol$^{-1}$K$^{-1}$</td>
</tr>
<tr>
<td>$\Delta G_{DP}^{\ddagger}298K$</td>
<td>59.4 kJmol$^{-1}$</td>
</tr>
<tr>
<td>$\Delta S_{PD}^\ddagger$</td>
<td>55.9 Jmol$^{-1}$K$^{-1}$</td>
</tr>
<tr>
<td>$\Delta G_{PD}^{\ddagger}298K$</td>
<td>85.0 kJmol$^{-1}$</td>
</tr>
</tbody>
</table>

The difference between $\Delta G_{PD}^{\ddagger}298K$ and $\Delta G_{DP}^{\ddagger}298K$ should be consistent with the free energy difference between the protonated and deprotonated states at the given pH ($pH=6.0$).

From table 1.13 above: $\Delta G_{PD}^{\ddagger}298K - \Delta G_{DP}^{\ddagger}298K = 85.0 - 59.4 = 25.6$ kJ/mol.

This is consistent with the predicted free energy difference between the two states, as determined by:

$\Delta G_{prot} = +2.303RT(pH-pK_a)$ and $\Delta G_{deprot} = -2.303RT(pH-pK_a)$

$\Delta G_{prot} = -25.7$ kJ/mol and $\Delta G_{deprot} = 25.7$ kJ/mol

With the thermodynamic parameters known, the complete energy level diagram for the methyl exchange process (rate limited by proton-exchange) is as follows:

![Figure 1.49](image-url) Completed free energy diagram representing energy associated with the methyl exchange process for substituted dimethylamines which are mediated by proton-exchange and inversion events. Where 43.1 kJ/mol corresponds to the free energy change associated with nitrogen inversion whereas proton-exchange is associated with a 14.2 kJ/mol change in free energy.

The following table compares the thermodynamics for the methyl exchange process for both the $N\alpha$ of tetramethyllysine and the salt-linked $K1\varepsilon$ of dimethyllysine (Table 1.14).
Table 1.14 Comparison between the thermodynamic parameters associated with the N\textalpha\ methyl exchange in tetramethyllysine to those associated with the methyl exchange for the salt-linked K1\epsilon dimethyllysine in HEWL.

<table>
<thead>
<tr>
<th></th>
<th>N\textalpha\ tetramethyllysine</th>
<th>K1 N\epsilon dimethylamino in HEWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_a$</td>
<td>56.3 kJmol$^{-1}$</td>
<td>104.1 kJmol$^{-1}$</td>
</tr>
<tr>
<td>$\Delta H^\ddagger$</td>
<td>53.8 kJmol$^{-1}$</td>
<td>$\Delta H^\ddagger$</td>
</tr>
<tr>
<td>$\Delta S_{DP}^\ddagger$</td>
<td>36.1 Jmol$^{-1}$K$^{-1}$</td>
<td>$\Delta S_{DP}^\ddagger$</td>
</tr>
<tr>
<td>$\Delta G_{DP}^\ddagger_{298K}$</td>
<td>43.1 kJmol$^{-1}$</td>
<td>$\Delta G_{DP}^\ddagger_{298K}$</td>
</tr>
<tr>
<td>$\Delta S_{PD}^\ddagger$</td>
<td>-11.8 Jmol$^{-1}$K$^{-1}$</td>
<td>$\Delta S_{PD}^\ddagger$</td>
</tr>
<tr>
<td>$\Delta G_{PD}^\ddagger_{298K}$</td>
<td>57.3 kJmol$^{-1}$</td>
<td>$\Delta G_{PD}^\ddagger_{298K}$</td>
</tr>
</tbody>
</table>

Note that it is entropically favourable for deprotonation of K1\epsilon but entropically unfavourable for deprotonation of N\textalpha\ of tetramethyllysine. This is likely due to K1’s participation in a salt link with E7 in HEWL. Presumably, the salt link reduces the entropy of system thereby making deprotonation entropically favourable. The N\textalpha\ of the small molecule tetramethyllysine does not participate in strong hydrogen bonding under the conditions used and so deprotonation would perhaps be expected to be entropically unfavourable.

Furthermore, unlike the N\textalpha\ methyl exchange for tetramethyllysine where the activation energy was 56.3 kJmol$^{-1}$, the $E_a$ for the K1 sidechain N\epsilon methyl exchange is almost two-fold larger. Whereas the N\textalpha\ dimethylamino of tetramethyllysine is highly solvent exposed and presumably only H-bonds to solvent, the N\epsilon dimethylamino of K1\epsilon in reductively methylated HEWL is salt linked with E7. Salt bridge interactions have been shown to contribute ~12 kJmol$^{-1}$ of stability to proteins (Anderson et al. 1990). This is consistent with the difference in $\Delta G_{DP}^\ddagger$ between the N\textalpha\ (43.1 kJ/mol) and K1 N\epsilon (59.4 kJ/mol) inversion processes. The additional energy in the case of $\Delta G_{DP}^\ddagger$ for K1 N\epsilon inversion is therefore likely due to salt link breakage.

1.14 Average $^1$H resonance inequivalence for N\epsilon dimethylamino methyls of lysine side chains in reductively methylated proteins

In most cases, average proton chemical shifts differences ($^1$H $\Delta\delta_{H-H}$) for inequivalent methyls are small, even in the case of salt linked dimethylamino groups; such that only resonance broadening is observed. However, in very rare occasions, $^1$H resonance “splitting” may also occur. The next section attempts to provide a description of the underlying mechanisms behind $^1$H resonance splitting for dimethylamino groups in reductively methylated proteins.
Several recent investigations of $^{13}$C-reductively methylated proteins by $^1$H,$^{13}$C heteronuclear NMR have revealed dimethyllysine methyl group protons exhibiting large chemical shift differences ($\Delta\delta_{H-H}$) (Zhang and Vogel 1993, Larda et al. 2012, Hattori et al. 2013) (Figure 1.50).

Typically, however, methyl $^1$H resonances are either coalesced or line broadened whereas complete $^1$H resonance splitting is relatively uncommon for reductively methylated lysine residues in proteins (Abraham, Larda). As shown previously, $^{13}$C methyl carbon resonance inequivalence for leucines and dimethylamines can be sufficiently large such that distinct methyl resonances are observed for a given dimethyllysine side chain; with chemical shift differences being on the order of ~0-10 ppm by DFT. In contrast, $^1$H chemical shift differences are typically much smaller, on the order of 0-2 ppm by DFT for N$\alpha$ dimethyls but usually <0.2 ppm for N$\varepsilon$ dimethylamino groups. Asymmetry, like that experienced by leucine $\delta$-methyls or N$\alpha$ dimethyls is expected to be much less for N$\varepsilon$ dimethyls given their relative distance to the Ca chiral center (six intervening bonds for N$\varepsilon$ vs two for N$\alpha$). Indeed, NMR spectra of reductively methylated proteins show that line broadening is generally greater for N$\alpha$ methyl protons than for N$\varepsilon$ methyl protons in the absence of hydrogen-bonding interactions (Figure 1.51).
Therefore, it is reasonable to assume that asymmetry due to proximity to the Cα is far less for Nε dimethyls than for Nα dimethyls. Thus, whereas extensive 1H and 13C line broadening is common for Nα dimethylamino resonances, only slight line broadening occurs for Nε dimethyl resonances. What then, is responsible for cases of extreme 1H inequivalence for certain lysine Nε dimethylamino groups (Figure 1.50 (A) K13 in CaM, (B) K27 in Ubiquitin, (C) K1ε in HEWL)?

Unlike 13C resonance splitting for Nε dimethyllysines, which arises from strong hydrogen bonding interactions, 1H resonance splitting does not seem to occur simply via salt bridge or H-bond formation. 13C methyl resonance inequivalence often occurs even in the absence of 1H inequivalence. Insight into what causes 1H resonance inequivalence for seemingly equivalent protons was found in a 2006 study on isotopically labeled lysine residues of DNA Polymerase γ (Gao et al. 2006). In this study, the protein was enriched with lysines labeled with 13C at the ε carbon. The authors found that at neutral pH, the methylene protons of the Cε were inequivalent for only three of ten lysines (K287, K291, and K312), resulting from immobilization of their side chains (Figure 1.52). The side chain immobilization was attributed to a combination of steric and strong hydrogen-bonding interactions.
Due to the chirality of proteins, one might expect that many seemingly magnetically equivalent nuclei by NMR (such as the methylene protons of a lysine side chain) should naturally exhibit some degree of inequivalence. Normally, however, fast side chain dynamics (such as C-C and C-N bond rotations) as well as rapid molecular tumbling create an effective isotropic condition such that protein chirality effects become somewhat negligible. Thus, for non-immobilized lysine side chains, the methylene Cε protons are effectively magnetically equivalent provided rapid reorientation and bond isomerizations along the methylene chain (Gao et al. 2006). For K287, K291, and K312 of the Pol γ lyase domain, however, the authors note that these residues are involved in either stable salt-bridge or H-bonding interactions, which immobilize the methylene side chain. The fixation of the side chain prevents rapid reorientations and leads to a condition of relatively long-lived asymmetry. Furthermore, K312 is somewhat de-solvated (exhibits the lowest solvent accessibility of all lysines), which is thought to result in constrained packing with adjacent residues (Gao et al. 2006). Similar to K312 Cε methylene protons, the Nε methyl protons of reductively dimethylated K1 in hen egg white lysozyme (Figure 1.50 C) and K27 in dimethylated-ubiquitin (Figure 50 B) exhibit methyl proton resonance inequivalence. Predicted side chain and total residue solvent exposure (solvent-accessible surface) for these two residues shows that both K1 and K27 are two of the most buried lysines in HEWL and UBQ (Figure 1.53).
Figure 1.53 Normalized side chain and total residue solvent exposure estimates for two lysine residues (K1 of Lysozyme (193L) and K27 of Ubiquitin (1UBQ)) that exhibit methyl \(^1^H\) resonance inequivalence. Solvent exposure values were predicted using four different software packages: molmol (Koradi et al. 1996), getarea (Fraczkiewicz and Braun 1996, 1998), surfaceracer (Tsodikov et al. 2002), and DSSP (Joosten et al. 2011). Surface Racer and DSSP predict total residue exposure only, whereas both total and side chain exposure can be predicted with molmol and getarea.
Furthermore, B-factor estimates for K1 (hewl) and K27 (ubq) indicate that these residues are two of the least dynamic of the lysines in either structure (Figure 1.54).

As expected, K13 (involved in salt bridge) and K96 (doubly H-bonded) also exhibit low B-factors consistent with their involvement in strong H-bonding interactions (Gerken et al. 1982). However, neither exhibit $\text{Ne}$-methyl $^1\text{H}$ resonance inequivalence, indicating that even in the presence of salt-links and strong H-bonds, lysine methylene side chain immobilization does not always occur. It may be that the prerequisites to $\text{Ne}$ methyl proton inequivalence are a combination of both strong H-bonding and reduced solvent exposure. However, K287 and K291 for DNA Polymerase $\gamma$ both have relatively large B-factors and high solvent exposure indicating that additional factors such as packing with adjacent residues are likely important for determining whether a lysine side chain is immobilized or not.

A final note of interest is the distinct upfield chemical shift (in both $^1\text{H}$ and $^{13}\text{C}$) of the K1$\varepsilon$ dimethyl resonances for reductively methylated HEWL (Figure 1.50 C) and for the $\text{Ce}$ methylene resonances of K291 in DNA Polymerase $\gamma$ Lyase Domain (Figure 1.52). Whereas both K1 and K291 are nearby phenylalanines and given that ring current effects have been known to perturb carbon chemical shifts by $>5$ ppm, it is perhaps reasonable to suggest that the unique upfield $^{13}\text{C}$ shifts for the K1$\varepsilon$ dimethyls and the K291 $\text{Ce}$ methylene protons are the result of ring currents (Blanchard et al. 1997). However, there are several lysines proximal to aromatics in crystal structures of HEWL and the Lyase Domain of DNA Poly $\gamma$ (a number of which are also predicted to be involved in cation-pi interactions) that exhibit normal chemical shifts, thereby suggesting otherwise (Figure 1.55).
London et al. (2008), describe the dependence of amino acid side chain $^{13}$C chemical shifts on dihedral angles (London et al. 2008). In particular, they provide an explanation for the upfield $^{13}$C$\varepsilon$ shift of K291 for the DNA Polymerase $\gamma$ Lyase Domain. As mentioned previously, both K291 and K287 were shown to be immobilized as a result of salt bridge interactions. The particular conformation adopted by a fixed lysine methylene side chain seems to significantly influence the chemical shift of $^{13}$C nuclei within the side chain (in an identical manner to the upfield shifts for leucine side chain $\delta$-methyls). In the case of K291, the side chain adopts a gauche conformer whereas K287 (which has a relatively normal $^{13}$C chemical shift) adopts the trans conformer (London et al. 2008). These conformations are depicted in the figure below (Figure 1.56).
A comparison of the side chain conformation of K1 in the crystal structure of HEWL to the side chain of K291 for DNA Pol γ (Figure 51), demonstrates that K1 also adopts a non-trans \( g^+ \)-like conformer. This is entirely consistent with the upfield shifts of the terminal methyl (dimethylamino) \(^{13}\)C resonances for K1 in HEWL.

1.15 Summary

The pH-dependent line broadening effects and methyl \(^1\)H and \(^{13}\)C resonance inequivalence in dimethylamine NMR spectra are directly related to proton exchange at the nitrogen. Whereas existing hypotheses postulate that base-mediated deprotonations followed by nitrogen inversions are responsible for methyl exchange, it has been shown throughout this thesis that proton-exchange alone results in resonance coalescence (and therefore methyl exchange). Nitrogen inversions are much faster than the timescale associated with the process giving rise to the line broadening effects, and in comparing protonated dimethylamines to leucine side chain δ-methyls, it is apparent that the \( \text{gauche} \) effect is the primary reason for large \( \Delta\delta \) between seemingly equivalent methyl groups. Protonation-deprotonation at the nitrogen is the dominant exchange process underlying pH-dependent resonance coalescence, and any factors which alter either the protonation-deprotonation rates and/or the populations of protonated:deprotonated amine will alter the chemical shift separation \( \Delta\delta \) between inequivalent methyl \(^1\)H resonances and between inequivalent methyl \(^{13}\)C resonances. In most cases, \( \text{Ne} \) methyl groups only exhibit line broadening in part due to the distance from the chiral \( \text{Ca} \), and exchange on the fast-intermediate timescale. \( \text{Na} \) methyls are much closer to the chiral \( \text{Ca} \) carbon and therefore experience greater asymmetry due to the \( \text{gauche} \) effect. They exhibit extensive broadening or complete splitting of resonances in a manner similar to leucine δ methyls. However, \( \text{Ne} \) methyls can also exhibit complete resonance splitting (methyl resonance inequivalence) under conditions of slowed proton exchange at the nitrogen. This primarily occurs when the nitrogen is hydrogen bonded or salt linked to other residues (usually backbone carbonyl oxygens and acidic residue side chains). Under these conditions, the methyls are fixed in an asymmetric condition due to the tetrahedral geometry of the nitrogen and the \( \text{gauche} \) effect. The increased pKa of the dimethyllysine arising from the stable hydrogen bonding...
interaction also presumably results in slowed deprotonation and a shifted equilibrium toward the protonated state. Some have noted that salt-linked lysines do not always exhibit increased pKas and that proton-exchange is faster for salt-linked amino groups (Advances in Protein Chemistry and Structural Biology, Volume 93). The timescale for these proton-exchange events at salt-linked amino groups was reported on the scale of tens of picoseconds. This exchange is outside the window of NMR experiments whereas the protonation-deprotonation events considered here refer to those in which protonated and deprotonated states persist on a microsecond to millisecond timescale.

1.16 References


Bauschlicher CW Jr., Partridge H (1995) The Sensitivity of B3LYP Atomization Energies to the Basis Set and a Comparison of the Basis Set Requirements for CCSD(T) and B3LYP. NASA Ames Research Center, Moffett Field, CA 94035, USA DOI: 10.1016/0009-2614(95)91855-R


69


Gancarz R (1995) Nucleophilic addition to carbonyl compounds: competition between hard (amine) and soft (phosphite) nucleophile. Tetrahedron 51(38):10627–10632


70


Lundblad RL. Techniques in Protein Modification. 1995 CRC Press LLC.


Chapter 2

Lysine Methylation Strategies for Characterizing Protein Conformations by NMR

2.1 ABSTRACT

In the presence of formaldehyde and a mild reducing agent, reductive methylation is known to achieve near complete dimethylation of protein amino groups under non-denaturing conditions, in aqueous media. Amino methylation of proteins is employed in mass spectrometric, crystallographic, and NMR studies. Where biosynthetic labeling is prohibitive, amino $^{13}$C-methylation provides an attractive option for monitoring folding, kinetics, protein-protein and protein-DNA interactions by NMR. Here, we demonstrate two improvements over traditional $^{13}$C-reductive methylation schemes: 1) By judicious choice of stoichiometry and pH, ε-aminos can be preferentially monomethylated. Monomethyl tags are less perturbing and generally exhibit improved resolution over dimethyllysines, and 2) By use of deuterated reducing agents and $^{13}$C-formaldehyde, amino groups can be labeled with $^{13}$CH$_2$D tags. Use of deutero-$^{13}$C-formaldehyde affords either $^{13}$CHD$_2$, or $^{13}$CD$_3$ probes depending on choice of reducing agent. Making use of $^{13}$C-$^2$H scalar couplings, we demonstrate a filtering scheme that eliminates natural abundance $^{13}$C signal.

The work presented in this chapter was originally published as: “Larda ST, Bokoch MP, Evanics F, Prosser RS. (2012) Lysine methylation strategies for characterizing protein conformations by NMR. J Biomol NMR. 54(2):199-209”. Copyright © Springer Science+Business Media B.V.

My contributions to this project involved preparing all methylated protein samples (except those for Figures 2 and S8), working out conditions for enhanced monomethylation yield, assignment of K1α and K1ε, the CH$_2$D labeling chemistry, conducting linewidth simulations using MEXICO, study on deuteromethyl-Calmodulin in PVP, and pH titrations. Most methyl resonance assignments were done by Dr. Ferenc Evanics. Dr. Michael Bokoch contributed via methylation studies on the β2-Adrenergic receptor. Dr. R. S. Prosser conducted mono and dimethyl deprotonation rate analyses as well as both temperature and formaldehyde concentration experiments shown in Figures 2 and 8, respectively.

75
2.2 INTRODUCTION

NMR studies of proteins generally require $^2$H, $^{13}$C, and $^{15}$N isotopic enrichment or, at the very least, incorporation of precursors for $^{13}$C-methyl labeling (Sprangers et al. 2007). Isotope labeling is necessary to accomplish assignments and perform detailed analyses of structure and dynamics. Although enrichment through minimal media is easily achieved with proteins expressed in E. coli, many proteins require eukaryotic expression systems to provide the proper protein folding machinery and accomplish the requisite post-translational modifications. Under such circumstances, biosynthetic isotopic enrichment may be unfeasible. As an alternative, $^{13}$C-methyl tagging of lysine residues provides domain specific information on conformation and dynamics. NMR applications have primarily made use of cysteine- (Kalbitzer et al. 1992; Klein-Seetharaman et al. 1999; Luchette et al. 2002; Oxenoid et al. 2002) and lysine-specific tags (Means and Feeney 1968, Zhang and Vogel 1993; Macnaughtan et al. 2005; Gerken et al. 1982), although other residues may be tagged under aqueous, non-denaturing conditions (Means and Feeney 1968; Joshi et al. 2004; Antos and Francis 2006; Ojida et al. 2005; Means and Feeney 1995; Abe et al. 1997).

Lysines are particularly convenient labeling sites due in part to their abundance in proteins of 6-7% (Yokoyama 2003). They are often found in structurally and functionally important regions including binding interfaces and salt linkages. Both N-terminal $\alpha$-NH$_2$ groups and lysine $\varepsilon$-NH$_2$ groups may be easily labeled under non-denaturing conditions through reductive methylation using $^{13}$C-formaldehyde and a suitable reducing agent. The resulting methyl probes can be resolved in a $^{13}$C,$^1$H HSQC spectrum, allowing the study of structure and dynamics of specific protein domains. Assignments are made possible by preparing lysine to arginine substitution mutants, should more traditional NMR assignment approaches fail. A significant body of research has been published on the use of $^{13}$C-dimethyl lysine probes for NMR studies of protein conformational changes through intermolecular interactions, metal or small molecule binding, or changing physiological conditions (Gerken et al. 1982; Abraham et al. 2009; White and Rayment 1993).

Lysine methylation is generally non-perturbing because only a small exogenous chemical moiety is added to the native protein and charge is maintained, unlike acylation. The majority of spectroscopic and crystallographic studies have found that lysine dimethylation results in relatively minor structure perturbations (Zhang et al. 1994; Moore et al. 1998; Rayment 1997; Rypniewski et al. 1993; White and Rayment 1993), although crystallization properties of proteins may be significantly altered (Rayment...
Hen egg white lysozyme has long served as a model system for protein biochemistry and structural biology. X-ray crystallography studies of fully dimethylated HEWL reveal that the protein backbone is superimposable with that of the unmethylated protein to an RMSD of 0.4 Å, with the exception of two surface loops (Palmer et al. 2001). The effect of reductive methylation on protein activity is more case specific. For example, dimethylation slows the steady-state rate of MgATP hydrolysis under saturating actin concentrations for the 20 kDa myosin subfragment-1 enzyme while the catalytic activities of trypsin and chymotrypsin are unchanged (Korzhnev et al. 2005). Conversely, reductively methylated horse liver alcohol dehydrogenase was reported to have enhanced catalytic activity (Tsai et al. 1974) whereas dimethylation of the β2 adrenergic receptor was shown to produce no significant alteration of the GPCR’s structure, ligand binding or G-protein coupling relative to the unlabeled receptor (Bokoch et al. 2010).

$^{13}$C-methyl lysine probes generally serve as excellent NMR reporters due to favorable chemical shift dispersions, relatively long transverse relaxation times, and high sensitivity (Abraham et al. 2009; White and Rayment 1993). One drawback, however, is that dimethyl resonances are often line broadened by intermediate or slow conformational exchange between unique methyl environments (Gerken et al. 1982; Abraham et al. 2009). This exchange broadening in combination with resonance proximity/overlap with other CH-containing species in solution makes interpretation of $^{13}$C,$^1$H spectra difficult, especially for very large systems.

We describe a methylation protocol that favors monoalkylation of protein amino groups and demonstrate that monomethyl tags circumvent problems typically associated with exchange broadened dimethylamino. We also discuss modifications to reductive alkylation protocols which result in the production of $^{13}$CH$_2$D, $^{13}$CHD$_2$, or $^{13}$CD$_3$ tags. The presence of deuterium permits the use of NMR pulse sequence schemes that serve to filter out natural abundance $^{13}$C signal, thereby making $^{13}$C,H spectroscopy a viable option in situations where the labeled fraction is very small. Furthermore, deuterium is an excellent probe for the investigation of dynamics by NMR (Tugarinov and Kay 2006).

**2.3 MATERIALS AND METHODS**

Lysozyme from hen egg white (HEWL), bovine serum albumin (BSA), sodium cyanoborohydride, and dimethamine (DMA) borane were obtained from Sigma-Aldrich. Aqueous $^{13}$C formaldehyde (20% w/w) was purchased from Cambridge Isotope Laboratories Inc (Andover, MA). *Xenopus* calmodulin was expressed in *Escherichia coli* and purified according to previously described methods (Ikura et al. 1990). Polyvinylpyrrolidone (PVP-40) was purchased from Sigma-Aldrich.
NMR. All NMR experiments were performed on a 600 MHz Varian Inova spectrometer using an HCN/FCN cryogenic probe. Typical $^{13}$C and $^1$H pulse lengths were 14 µs and 8.5 µs, respectively. Standard $^{13}$C,$^1$H HSQC spectra were obtained using 8 scans and 128 increments. A spectral width of 15000 Hz was used in the indirect ($^{13}$C) dimension with WALTZ-16 $^{13}$C-decoupling. For $^2$H-edited $^{13}$C,$^1$H HSQC spectra of apo-calmodulin in 300 mg/mL polyvinylpyrroldione (PVP), standard pulse widths for $^1$H, $^{13}$C, and $^2$H were 8.2 µs, 13.8 µs, and 515 µs, respectively. Spectra in PVP were acquired with 64 scans and 96 increments at a temperature of 20°C. A spectral width of 6000 Hz was used in the indirect dimension. Spectral analyses were performed using NMRPipe (Delaglio et al. 1995) and NMRViewJ (Johnson and Blevins 1994). All sample volumes were between 500 and 550 µL and contained 5-10% D$_2$O. Assignments of lysozyme methyl lysine spectra were determined by corroborating those of (Macnaughtan et al. 2005) and by comparing aromatic NOEs and line broadening from dissolved nitroxide spin-labels (4-Hydroxy-TEMPO) to the predicted effects from the X-ray crystal structure. The N-terminal amino group resonances of both the mono and dimethyl species were confirmed by: 1) the unique pH dependence of the chemical shift and 2) the pronounced intensity of the dimethyl species and relatively weak intensity of the monomethyl species of K1α in agreement with its reactivity across a range of formaldehyde concentrations.

**Typical methylation reaction.** HEWL, BSA, or apo-calmodulin was dissolved in 0.05-0.1 M phosphate, pH 7.5 to a concentration of 0.3-1.5 mM. $^{13}$C formalin (aqueous formaldehyde) was added directly to the protein solution at a concentration equal to that of the reactive amines (N-terminal residue plus lysines). The solution was then briefly mixed and left to sit at room temperature for 1 hour without stirring. Solid sodium cyanoborohydride (3 to 5-fold excess over total aldehyde) was then added directly to the mixture followed by several inversions to mix. The sample was left at room temperature for 4-6 hours, followed by dialysis or exchange into appropriate buffer using centrifugal concentrators. This scheme gives roughly a 1:1 ratio of mono:dimethyl lysine.

**Dimethyl-selective reaction.** Lyophilized protein (HEWL or BSA) was dissolved in 0.1 M phosphate pH 7.0 to a concentration of 0.3-1.3 mM. $^{13}$C formalin was added at a ratio of five equivalents aldehyde relative to the number of reactive amines. The reaction mixture was allowed to equilibrate for a minimum of 1 hour at room temperature before addition of solid sodium cyanoborohydride (~2 equivalents relative to aldehyde). After 6 hours at room temperature, the above scheme was repeated (for a total of 10 equivalents of $^{13}$C formaldehyde).

**Monomethyl-selective reaction.** Lyophilized protein was dissolved in 0.1 M phosphate pH 5.7 to a
concentration of 1.3 mM. $^{13}$C-formalin was added to achieve equimolar stoichiometry between aldehyde and the number of reactive amines. The mixture was left to equilibrate for at least 1 hour at room temperature before addition of roughly 15 equivalents of solid sodium cyanoborohydride. Reactions were left to proceed for a minimum of 6 hours prior to extensive dialysis.

**Aldehyde stoichiometry reaction series:**

**HEWL.** Lyophilized HEWL was dissolved to a concentration of 2 mM in 20 mM HEPES, pH 7.5. Separate reactions were conducted with the following ratios of $^{13}$C-formaldehyde to reactive primary amine: 0.2:1, 0.5:1, 1:1, 2:1, 5:1, and 10:1. Each sample was reduced with 100 mM DMA-borane. After 12 hours reacting at 20°C, samples were dialyzed extensively into 20 mM HEPES, pH 7.5 and protein was then concentrated to ~0.5 mM prior to NMR experiments.

**BSA.** A stock solution of 300 µM BSA in 50 mM phosphate, pH 6.0 was prepared. Overnight reactions were conducted at 23°C, each with one of following aldehyde:reactive amine stoichiometries: 0.1:1, 0.25:1, 0.5:1, 1:1, 2:1, and 5:1. Samples were then extensively dialyzed back into 50 mM phosphate, pH 6.0. Final concentration of protein in each sample was 250 µM.

**Lineshape estimation for exchange-broadened resonances:** We used the non-iterative MEXICO program (MEXICO software suite v.3, Alex Bain, McMaster University) to predict line shapes for exchange-broadened mono and dimethyl-lysines. Deprotonation rates were calculated according to previously established methods (Borisenko et al. 2000a) and populations of protonated and deprotonated species were determined from pKas determined by pH titration of hewl mono- and dimethyl lysines (Online Resource 1-7).

**Deuteromethylation of $^{15}$N-CaM:** 400 µM apo-$^{15}$N-calmodulin in 50 mM phosphate buffer, pH 7.5 was reacted with $^{13}$C formalin at a stoichiometry of 3 aldehydes per reactive amine. The mixture was left at room temperature for 1 hour before mixing with 5 molar equivalents of sodium cyanoborodeuteride relative to aldehyde. The reaction was left to proceed at room temperature for 6 hours before repeating the entire sequence a second time. In total, 6 molar equivalents of $^{13}$C-formaldehyde and 10 equivalents of sodium cyanoborodeuteride, relative to moles of reactive amine, were added to the solution of protein. The solution was purified by dialysis into 50 mM phosphate, pH 7.5.

**Calmodulin in PVP** – A 400 µM solution of partial mono/di-$^{13}$CH$_2$D deuteromethylated calmodulin in phosphate was added to solid PVP. The mixture was gently shaken until the solid was fully solvated. 400 µL of the viscous mixture was combined with 50 µL D$_2$O and loaded into a salt tolerant
susceptibility matched slot shaped Shigemi tube. The final concentration of PVP was 300 mg/mL.

2.4 RESULTS AND DISCUSSION

1. A comparison of mono- and dimethyl lysine spectroscopic features. Using a typical reductive methylation scheme, described in the Materials and Methods section, hen egg white lysozyme was labeled at pH 7.4 so as to produce a roughly equivalent quantity of monomethyl and dimethyl lysines. Figure 2.1 shows the mono and dimethyl regions and assignments for each of the methyl resonances. Note that K1(ε) residue is represented by two distinct peaks at physiological pH and temperature in the dimethyl and monomethyl regions. This is consistent with the X-ray crystal structure of the lysine dimethylated protein, which reveals that the K1 ε-NH$_3$ is involved in a salt linkage with E7 (Kumar and Nussinov 1999; Kumar et al. 2000; Sarakatsannis and Duan 2005). A similar dimethyl lysine resonance splitting is observed for K13, which is thought to be involved in a salt link with the carboxy terminal residue, Leu129 (Imoto et al. 1972). Salt linkages presumably slow the otherwise rapid interconversion of non-degenerate methyl conformers, such that each state of the K1(ε) and K13 dimethyl aminos gives rise to a distinct resonance.

Figure 2.2 compares the $^{13}$C,$^1$H HSQC spectra of both mono- and dimethyl lysines as a function of temperature. Resonances associated with the dimethyl lysines generally exhibit a significant degree of line broadening between 0 °C and 40°C. In particular, resonance K1(ε) is represented by one motionally averaged resonance at 55°C and two distinct resonances at 0°C, as is also the case for K13. In contrast, the monomethyl resonances are generally well-resolved over the same temperature range.

![Figure 2.1](image-url) (13C,1H) HSQC spectra at 37°C and pH 7.5 of partially methylated HEWL showing assignments for mono- (top) and dimethyl-lysine resonances (bottom)
Of note is that monomethylation (specifically, partial mono/dimethylation) does not appear to give rise to inhomogeneous broadening, which might be expected for a superposition of protein conformers with differing methylation levels. This further suggests that methylation is minimally perturbing to the structure of HEWL, although we do find that monomethylation of the protein leads to reduced solubility at pH 8.5-9.0 relative to the unmethylated or fully dimethylated form; an observation that has been corroborated by others (Gerken et al. 1982).

While the broadening of the K1(ε) and K13 resonances can be explained in part by their participation in salt-linkages (with breakage and reformation giving rise to two unique environments), we consider below, possible sources for broadening of non salt-linked dimethyl species. These include: 1) hydrogen bonding of the amino group to other residues or solvent (Moult et al. 1976), 2) cation-π interactions 3) contacts between the amino-methyl moiety and hydrophobic regions of the protein, and 4) deprotonation-controlled nitrogen inversions which reorient methyl groups between distinct, local chemically inequivalent sites (Goux et al. 1984).

Figure 2.2 ($^{13}$C,$^1$H) HSQC spectra of reductively methylated HEWL, as a function of temperature at pH 7.4. The dimethyl lysine corresponding to K1(ε) exists as two distinct resonances at lower temperatures as is the case for K13; many resonances exhibit coalescence and line broadening at physiological temperatures. The latter two residues are implicated in salt linkages to E7 and L129, respectively. Note that the $^1$H and $^{13}$C referencing have been altered to maximize overlap of peaks in order to emphasize the changes in linewidth that occur with temperature.
The stark difference in resolution between mono- and dimethylamino resonances can be explained by the fact that the above processes are fast on the NMR timescale for the mono species, but sufficiently slow for dimethyls so as to give rise to exchange broadening. To investigate the influence of protonation-deprotonation of the ε-amino groups on methyl dynamics, 13C,1H HSQC spectra were recorded as a function of pH for both mono- and dimethyl species. Increasing pH dramatically reduces line broadening of dimethyl resonances (Abraham et al. 2009) as shown by the pH-dependent spectra in Figure 2.3. In contrast, pH has a small effect on the line widths associated with monomethyl lysine resonances, whose reorientations are likely in the fast motional limit in the majority of cases. In addition, the size of the dimethyl moieties exacerbates local steric crowding effects whereas the smaller monomethyl lysine tags, experience greater freedom with regard to N-C bond isomerizations and nitrogen inversions. Our observations suggest that monomethyl lysine reorientations are sufficiently fast that the variety of environments sampled average to a single isotropic state, whereas the slow reorientations accompanying dimethyl lysines result in sufficiently long lived sampling of local heterogeneity (Abraham et al. 2009; Gerken et al. 1982).

2. Lysine side chain dynamics and deprotonation rates. An analysis of the line broadening at 25 °C, reveals that the dimethyls of K1(ε) undergo two-site exchange between similarly populated states with a correlation time of roughly 1.2 ms (kex = 10^1 s^-1 at 25 °C, 2x10^2 s^-1 at 40 °C). A similar degree of broadening is observed for K13. In general, slow lysine side chain reorientations likely depend in part on the charge state of the ε-amino group since an increase in pH to within roughly 1.5 pH units of the pKa results in dramatically improved resolution. Transient deprotonation events are expected to be very fast near the dimethyl lysine pKa, as discussed by others (Borisenko et al. 2000b; Tieleman et al. 2003). Lysine protonation/deprotonation kinetics are governed by the pKa of the lysine residue and the buffer, B (Borisenko et al. 2000b).
Upon deprotonation, salt linkages and interactions with negatively charged sites on the protein surface become weakened allowing for more frequent reorientations of lysine side chains. Furthermore, deprotonation by either solvent or adjacent titrating species in a protein is a critical step preceding nitrogen inversion (Goux et al. 1984). Deprotonation kinetics are well known to occur through interactions with water or the buffer base as described by the following reactions:

\[
\text{Lys}^- \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} \text{Lys}^- \overset{k_2}{\rightarrow} \text{Lys}^+ + \text{H}_2\text{O} + \text{H}^+ \quad (1)
\]

\[
\text{Lys}^- \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} \text{Lys}^+ + \text{H}_2\text{O} \quad (2)
\]

\[
\text{Lys}^- \overset{k_1}{\rightarrow} \text{Lys}^+ + \text{B} \quad (3)
\]

Following the approach of Borisenko et al, we assume proton transfer to be diffusion-controlled in the thermodynamically favorable direction with a rate constant \( k = 10^{10} \text{ M}^{-1} \text{ s}^{-1} \), and smaller by a
factor $\Delta pK_a$ (i.e. the difference in the pKa values of the partners exchanging a proton) in the reverse direction (Borisenko et al. 2000b). It then becomes possible to evaluate the average deprotonation rate associated with Lys-$NH_3^+$ for a given pKa. Note that the dimethyl lysines of lysozyme possess an average pKa around 10.5 while the average pKa of monomethyl lysines is estimated to be 11.9 (Table S2.1). Figure 2.4 graphs the estimated molecular deprotonation rates for mono- and dimethyl lysines, based on the sum of the above three mechanisms, and assuming a buffer concentration of 50 mM and buffer pKa of 7.2. Exchange with either hydroxide ions or the buffer tend to make the greatest contribution to the overall deprotonation rate, which is predicted to be on an intermediate NMR timescale at physiological pH (*vida infra*).

Using the predicted kinetic rates and weighted pH-dependent populations of protonated and deprotonated species, line shape simulations could not account for the exchange broadening observed for the dimethyl lysine resonances near physiological pH. Moreover, given that mono- and dimethyl lysines exhibit comparable pKa’s and deprotonation kinetics, we conclude that protonation kinetics alone cannot account for the line broadening. In other words, while deprotonation events are frequent, the populations are heavily weighted towards the protonated states. Therefore any broadening due to sampling of the deprotonated state would be negligible.

Goux et al. observed broadening of $^{13}$C-dimethyl amino resonances in one-dimensional carbon spectra for reductively methylated concanavalin (ConA). They employed model compounds to show that the pH-dependence of line broadening could be adequately explained by base-catalyzed nitrogen inversions. At pH < 5.6, the $^{13}$C methyl resonances of tetramethyllysine (TML) are represented by two
distinct peaks, attributable to two distinct methyl conformers related through inversion at the nitrogen center. Above pH 5.6, the resonances coalesce, yet there still exists substantial broadening at near-physiological pH (Goux et al. 1984). At pH 10, the $^{13}$C linewidths were sharp suggesting isotropic conditions or fast-regime exchange between methyl conformers. They explain that while they could not resolve $^{13}$C resonances for each of the $\varepsilon$-dimethyllysyl residues in ConA, they expect these to exhibit the same pH-dependent line broadening trend as for the N-terminal $\alpha$-dimethylamino group whose inversion rate is on the order of $10^7$ s$^{-1}$ (Goux et al. 1984). With nitrogen inversions being as fast as they are, exchange between methyl conformers alone would be too rapid to give rise to line broadening. It is therefore apparent that transient protonation/deprotonation events serve to lock and unlock inversion at the nitrogen. The populations are heavily weighted toward the protonated species but at neutral pH, roughly once every millisecond, solvent-mediated deprotonation of the amine allows for sampling of the other methyl conformer(s). Protonation extends the lifetimes of the different methyl conformers, thereby putting the exchange between states in the intermediate NMR timescale regime. In addition to inversions, differences between each of the dimethylamino pKas, their solvent-exposure, and steric environments, serve to explain the observed variability in line broadening as a function of pH between the 8 resonances in the dimethyl spectra (Figure 2.3).

### 4. Reaction schemes to favor monomethylated lysines.

The scheme shown in Figure 2.5 illustrates the key intermediates involved in the reductive methylation of lysine residues. The reaction mechanism can be described in terms of the formation of imine with subsequent conversion to the methylated species upon reduction (Means and Feeney 1968; Means and Feeney 1995). The first pass through the cycle generates monomethyl while the second round produces dimethyl lysine.

Reductive methylation of lysine residues is strongly biased toward complete conversion to the dimethyl species. Following monoalkylation, the resulting secondary amine is considerably more nucleophilic than unlabeled, primary amine. Free formaldehyde, therefore, preferentially reacts with monomethyl lysines over unlabeled lysines, favoring dimethylation. Consequently, it becomes difficult to maintain a large population of monomethyl lysines while ensuring labeling at every reactive site. Our approach to monoalkylation is to employ a pH at which imine formation is most favorable and a low ratio of formaldehyde to amine, followed by reduction with a large excess of cyanoborohydride. Experimentally, methylation at pH $\sim$6 enhances mono to dimethyl ratios substantially over reactions at higher pH (Figure 2.6). When labeling at pH 7.0-7.5 with a stoichiometry of 1:1 to 1.5:1 (formaldehyde:amine), the mono to dimethyl peak integrals are roughly 1:1 in a given sample; with
effective peak intensities of 1:2 (mono:dimethyl) given the two-fold greater number of dimethyl protons. At higher formaldehyde:amine stoichiometry, as per the dimethylation protocol, mono to dimethyl ratios are substantially reduced. For the monoalkylation scheme we describe, roughly half of the signal arises from monomethyl derivatives (3 proton each) and half from dimethyl lysines (6 proton each); giving roughly a 2:1 ratio of mono:dimethyl.

**Figure 2.5** Reaction scheme for $^{13}$C-reductive methylation of lysines. H refers to any hydride donor suitable for reducing the iminium ion to the corresponding secondary (monomethyl) or tertiary amine (dimethyl); a variety of reducing agents can be employed. R represents the rest of the protein chain. Note that while the amine nucleophilicity does not necessarily correlate with pKa, the monomethyl species is considerably more nucleophilic than the primary, unlabeled amine

The effect of aldehyde stoichiometry on the mono:dimethyl ratios can be seen in Figures S8, S9, and S10, for methylation of HEWL (S8) and BSA (S9/S10) (Figures S2.8-2.10), respectively. The labeling efficiency at a given lysine is dependent on pKa and as would be expected, the K1α of HEWL most readily incorporates $^{13}$C, often to the point of complete dimethylation even when the monoalkylation
scheme is employed.

The reducing agent of choice also influences the dimethyl to monomethyl ratio. While use of large molar excess of cyanoborohydride does improve monomethyl yield, it also enhances byproduct formation; we observe direct reduction of formaldehyde to methanol. Overall, monomethylation is less efficient than dimethylation but gives substantial improvements in spectral quality.

5. $^{13}\text{C}$-deuteromethylation of protein amino groups.

While $^{13}\text{C}$-monomethylation of lysine residues provides improved resolution at physiological pH over that of dimethyl lysines, our experience has shown that dimethyl lysines serve as useful reporters of protein conformational exchange and dynamics. We sought to further improve the tags by incorporating a deuterium nucleus in the methyl groups. $^{13}\text{CH}_2\text{D}$ deuteromethyl probes provide a means to observe methyl lysines in macromolecular systems where natural abundance $^{13}\text{C}$ signal may otherwise obscure detection. Deuteromethylation can be achieved by modification of the standard reductive alkylation protocol whereby pre-formed imine is reduced with deuteride rather than hydride (Figure 2.7).

Alternatively $^{13}\text{CHD}_2$-labeling is afforded by using $^{13}\text{C},d_2$ formaldehyde and a suitable mild reducing agent such as sodium cyanoborohydride or dimethylamine borane, whereas reduction with sodium cyanoborodeuteride would generate the corresponding $^{13}\text{CD}_3$ tag (Figure 8). A $\text{CD}_3$ dimethyl lysine tag should prove equally effective in mass spectrometry applications, where isotope coded affinity tagging (ICAT) is utilized (Goshe and Smith 2003). In this case, tagging by a protonated or perdeuterated dimethyl tag would result in a difference of mass to charge ratios ($\Delta m/z$) of 6. Recently, Kashai et al...
outlined an approach in which multiple ligand-specific conformations of the β2-adrenergic receptor could be identified by monitoring the extent of reactivity of specific lysine residues on the membrane protein through ICAT labeling by a protonated and deuterated version of succinic anhydride (Kahsai et al. 2011). Reductive alkylation has the advantage that charge is preserved by the reaction, while the dimethyl tag is minimally perturbing, in which case the delicate ensemble of membrane protein conformers should be faithfully reproduced.

The versatility of reductive alkylation lies in the diversity of aldehydes, ketones and reducing agents that can be employed. For example, primary and secondary amines can also be tagged with ³H via tritiated cyanoborohydride, thereby generating radionuclide labeled methyl groups. Figure 2.9 presents an overlay of spectra obtained from a standard ¹³C,¹H HSQC versus that from a deuterium-edited version for ¹³CH₂D-deuteromethyl labeled apo-calmodulin in polyvinyl pyrrolidone (PVP-40). Deuterium editing was achieved by making use of the C-²H coupling as described by Muhandiram et al (Muhandiram et al. 1995). The PVP concentration was such that the macromolecular crowding approximates that which is found within cells (300 mg/mL).

![Figure 2.9](image)

*Figure 2.9* Comparison of a standard ¹³C,¹H HSQC spectrum of ¹³CH₂D-deuteromethylated apo-calmodulin in 300 mg/mL polyvinylpyrrolidone (cyan) versus a ²H-edited ¹³C,¹H HSQC spectrum (black) acquired on the same sample. Deuterium filtering eliminates all background PVP polymer signal and allows for unambiguous identification of the signal arising from the ¹³C-deuteromethyl probes. Expansions of mono and dimethyl lysine regions show minor methyl lysine species that are well above baseline for the ²H-edited spectrum but within the noise regime for the unedited HSQC.
The deuteromethyl tag and corresponding $^2$H filtering schemes also provide improved water suppression while allowing for more detailed dynamics studies, as discussed by others, particularly with regard to CHD$_2$ probes (Tugarinov et al. 2005). Note from Figure 9 that resonances for the methyllysines seem to be nearly as well resolved under conditions of high viscosity as they are under standard solution conditions.

2.5 CONCLUSIONS

Reductive methylation of primary amino groups with $^{13}$C-formaldehyde is an effective means to generate site-specific methyl reporters on proteins. Aminomethyl probes can be used for the study of protein folding, dynamics, ligand binding and enzyme kinetics by NMR. However, without significantly raising the pH, it is found that dimethyl lysine resonances of HEW lysozyme exhibit pronounced line broadening, due to millisecond timescale reorientations at physiological temperatures. These intermediate timescale reorientations can be attributed primarily to salt-link breakage and reformation, and deprotonation-mediated nitrogen inversions. Other possible contributing factors to methyllysine resonance broadening include: intra- and intermolecular hydrogen bonds, steric, and hydrophobic interactions.

Significant line narrowing of dimethyllysines is often observed at a pH of 8.5 or greater where deprotonation rates, and consequently, nitrogen inversions, are fast on an NMR timescale. This rapid side chain reorientation is also found to be the case, at any pH, for monomethyl lysines since the single methyl is less restrictive toward bond rotations or inversions. More rapid interconversion between local anisotropic environments results in sharper line widths. Accordingly, certain non-salt-linked dimethyllysine resonances are sharper than others, presumably because of differences in their pKas and local environment.

While typical schemes for reductive alkylation of proteins with $^{13}$C-formaldehyde generate dimethyl-lysines, monomethylation can be favored by suitable stoichiometry and pH. Use of roughly 1 to 1.2 equivalents of aqueous formaldehyde relative to primary amine is crucial to prevent over-modification of sites while at 2-3 fold equivalents or higher, the reaction is heavily weighted toward dimethylation. Nearly selective monomethylation can be achieved under conditions that favour imine formation and minimal labeling. While higher monomethyl yields are readily achieved using the protocol described above, it becomes inevitable that some sites will be dimethylated (especially for the $\alpha$-NH$_2$ which has a significantly lower pKa than the $\varepsilon$-NH$_2$ species). We have demonstrated a
modification to conventional reductive alkylation and presented applications of $^{13}$C-deuteromethylation to studies of proteins by NMR. Benefits of $^{13}$C-deuteromethylation in combination with $^2$H-editing of $^{13}$C,$^1$H HSQCs include: 1) enhanced spectral quality via elimination of any and all background $^{13}$C-coupled proton signal, 2) unambiguous identification of signal arising from the aminomethyl probes, 3) substantial minimization of water signal breakthrough, and 4) the possibility of studying side chain dynamics via deuterium relaxation and $^2$H CPMG dispersion measurements. Finally, $^{13}$C-methyl probes on lysines are less susceptible to line-broadening associated with slower molecular tumbling given the high degree of rotational correlation of the sidechain. Methyl lysines should therefore serve useful as probes for in cell spectroscopy studies of proteins, where the protein of interest is either injected or transported into the cell. We emphasize that reductive $^{13}$C methylation is a robust, mild labeling protocol that can be applied under non-denaturing conditions to virtually any protein. Reaction conditions can be tailored to preferentially monoalkylate while isotopically labeled formaldehyde and or reducing agents provide further utility to the tags, in the case of $^{13}$CH$_2$D, $^{13}$CHD$_2$, and $^{13}$CD$_3$ deuteromethyls.

ACKNOWLEDGMENTS. RSP acknowledges NSERC (grant number 261980) for a research discovery award. MPB acknowledges support from the Stanford Medical Scientist Training Program.

ONLINE RESOURCES/SUPPORTING INFO.

Table S1 reports $^1$H and $^{13}$C chemical shifts and linewidths for methylated HEWL. Figures S2-S7 detail the effect of pH on chemical shifts of mono- and di-methyl lysines. Figure S8 shows the effect of aldehyde stoichiometry on the resulting ratio of mono- to dimethyl lysines for reductively methylated HEWL. Figure S9 and S10 show the effect of aldehyde stoichiometry on the ratio of mono- to dimethyl lysines for reductively methylated BSA.

2.6 SUPPLEMENTARY INFO

3. A comparison of the pKa of monomethyl and dimethyl lysines. Previous studies have reported estimates for the pKa’s of both the $\varepsilon$- and $\alpha$-amino group of free unmethylated, mono-, and dimethyl lysine$^{32}$ as well as for dimethyl lysine residues in methylated proteins such as egg white lysozyme$^{10}$. For free, unlabeled lysine at 20 mM, the $\varepsilon$-amino has a pKa of around 10.9, while mono- and dimethyl derivatives have pKa’s of roughly 11.2 and 10.3, respectively $^{10,33}$ These data correlate relatively well with estimates for lysine residues in proteins and prior estimates of the pKa’s of solvent-exposed dimethyl lysines, which are between 9.5 and 10.1$^{10}$. Yet pKa estimates for monomethyl lysines in egg white lysozyme have proved more difficult to assess – a consequence of poor solubility of the monomethylated derivative of the protein above pH 9.0. By taking advantage of enhanced stability of
monomethyl HEWL at lower temperatures and improved resolution of resonances afforded by two-dimensional NMR, estimates of monomethyl pKa’s could be attained. Our results indicate that monomethyl pKa’s are, on average, 11.9 for ε-aminos and between 7.9 and 8.1 for the N-terminal α-amino, as summarized in Table 1. The dimethyl pKa estimates are within the expected range and average 10.5 for ε- and 7.5 for the α-amino species.

<table>
<thead>
<tr>
<th>Residue</th>
<th>1H chemical shift [ppm]</th>
<th>13C chemical shift [ppm]</th>
<th>pKa</th>
<th>1H line width [Hz]</th>
<th>13C line width [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomethyl Lysines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K96/ K97</td>
<td>2.76</td>
<td>35.83</td>
<td>11.74</td>
<td>10.17</td>
<td>15.82</td>
</tr>
<tr>
<td>K116</td>
<td>2.72</td>
<td>35.44</td>
<td>11.79</td>
<td>8.50</td>
<td>17.23</td>
</tr>
<tr>
<td>K13</td>
<td>2.67</td>
<td>36.02</td>
<td>11.72</td>
<td>8.58</td>
<td>14.64</td>
</tr>
<tr>
<td>K96/ K97</td>
<td>2.62</td>
<td>35.18</td>
<td>12.13</td>
<td>8.75</td>
<td>16.06</td>
</tr>
<tr>
<td>K33</td>
<td>2.57</td>
<td>35.46</td>
<td>12.04</td>
<td>8.86</td>
<td>14.26</td>
</tr>
<tr>
<td>K1(α)</td>
<td>2.48</td>
<td>34.72</td>
<td>7.91-8.06</td>
<td>11.44</td>
<td>11.93</td>
</tr>
<tr>
<td>K1(ε)</td>
<td>2.45, 2.44</td>
<td>34.08, 34.28</td>
<td>11.98</td>
<td>8.57, 13.33</td>
<td>17.27, 17.40</td>
</tr>
<tr>
<td>Dimethyl Lysines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1(ε)</td>
<td>2.57, 2.57</td>
<td>43.74, 44.12</td>
<td>&gt;10.4</td>
<td>13.94, 17.13</td>
<td>17.56, 25.99</td>
</tr>
<tr>
<td>K116</td>
<td>2.87</td>
<td>45.41</td>
<td>10.65</td>
<td>16.51</td>
<td>15.38</td>
</tr>
<tr>
<td>K13</td>
<td>2.92, 2.90</td>
<td>45.57, 45.29</td>
<td>10.77</td>
<td>14.04, 11.89</td>
<td>15.73, 49.62</td>
</tr>
<tr>
<td>K1(α)</td>
<td>2.44</td>
<td>45.15</td>
<td>7.47</td>
<td>21.90</td>
<td>17.69</td>
</tr>
<tr>
<td>K33</td>
<td>2.70</td>
<td>45.49</td>
<td>10.68</td>
<td>9.95</td>
<td>23.39</td>
</tr>
<tr>
<td>K97</td>
<td>2.72</td>
<td>45.32</td>
<td>10.31</td>
<td>15.34</td>
<td>17.63</td>
</tr>
</tbody>
</table>

**Table S2.1** 1H and 13C chemical shifts and line widths for mono- and dimethyl lysines. Chemical shift and linewidth data is reported for partially labeled HEWL in phosphate buffer, pH 7.4. Where peaks are split, data are reported for both resonances. pKa measurements were determined from chemical shift vs pH titration curves (supplementary) for spectra acquired at 15 ºC.
Figure S2.2 Change in averaged chemical shift for each dimethyl lysine of reductively methylated HEWL.

Figure S2.3 Change in $^1$H chemical shift for each dimethyl lysine of reductively methylated HEWL.
Figure S2.4 Change in $^{13}$C chemical shift for each dimethyl lysine of reductively methylated HEWL.
Figure S2.5 Average weighted change in $^1$H and $^{13}$C chemical shift for each monomethyl lysine of reductively methylated HEWL.

Figure S2.6 Change in $^1$H chemical shift for each monomethyl lysine of reductively methylated HEWL.
Figure S2.7 Change in $^{13}$C chemical shift for each monomethyl lysine of reductively methylated HEWL.

Figure S2.8 Comparison of $^{13}$C,$^1$H HSQC spectra of reductively methylated HEW lysozyme, as a function of increasing $^{13}$C-formaldehyde concentrations. Panels A-F were obtained after reacting 2 mM HEW lysozyme (in 20 mM HEPES pH 7.5) with 100 mM DMA borane and 2.9, 7.2, 14.3, 28.6, 71.5, and 143 mM $^{13}$C-formaldehyde, corresponding to formaldehyde to primary amine ratios of 0.2:1, 0.5:1, 1:1, 2:1, 5:1, and 10:1, respectively. Panels A and B clearly show evidence of pronounced monomethylated species, whose $^{13}$C chemical shifts reside between 30 and 35 ppm, while panels C-F reveal increasingly intense resonances associated with dimethylated species. Each reaction mixture was gently stirred for 12 hours at 20°C, then dialyzed extensively overnight in H$_2$O. Final purified products were then concentrated to ~ 0.5 mM and prepared in pH 7.5 HEPES buffer for NMR analyses.
Figure S2.9 Comparison of $^{13}$C, $^1$H HSQC spectra of monomethyl lysine resonances for reductively methylated BSA, as a function of increasing $^{13}$C-formaldehyde concentrations. Labeling conditions were as follows: 300 µM BSA in 50 mM phosphate, pH 6.0. Panels A-F were obtained by reacting with 0.1, 0.25, 0.5, 1, 2, and 5X molar equivalents of $^{13}$C formaldehyde relative to number of reactive amino groups. Mild reduction of each sample with ~10X molar equivalents of sodium cyanoborohydride relative to moles of aldehyde. Following overnight reactions, samples were dialyzed extensively into 50 mM phosphate, pH 6.0. Final concentration of labeled protein in each sample is ~250 µM.
Figure S2.10 Comparison of $^{13}$C,$^1$H HSQC spectra of dimethyl lysine resonances for reductively methylated BSA, as a function of increasing $^{13}$C-formaldehyde concentrations. Labeling conditions were as follows: 300 µM BSA in 50 mM phosphate, pH 6.0. Panels A-F were obtained by reacting with 0.1, 0.25, 0.5, 1, 2, and 5X molar equivalents of $^{13}$C formaldehyde relative to number of lysines. Mild reduction of each sample with ~10X molar equivalents of sodium cyanoborohydride relative to moles of aldehyde. Following overnight reactions, samples were dialyzed extensively into 50 mM phosphate, pH 6.0. Final concentration of labeled protein in each sample is ~250 µM.
2.7 REFERENCES


Abraham SJ, Kobayashi T, Solaro RJ, Gaponenko V (2009) Differences in lysine pKa values may be used to improve NMR signal dispersion in reductively methylated proteins. J Biomol NMR 43:239-246


CHAPTER 3

Dynamic Equilibria between Monomeric and Oligomeric Misfolded States of the Mammalian Prion Protein Measured by $^{19}$F NMR*

3.1 ABSTRACT

The assembly of misfolded proteins is a critical step in the pathogenesis of amyloid and prion diseases, although the molecular mechanisms underlying this phenomenon are not completely understood. Here, we use $^{19}$F NMR spectroscopy to examine the thermodynamic driving forces surrounding formation of β-sheet-rich oligomers early in the misfolding and aggregation pathway of the mammalian prion protein. We show that initial assembly of a small octameric intermediate is entropically driven, while further assembly to putative prefibrillar aggregates is driven by a favorable change in enthalpy. Kinetic data suggest that formation of the β-octamer represents a rate-limiting step in the assembly of prion aggregates. A disease-related mutation (F198S) known to destabilize the native state of PrP was also found to stabilize the β-octamer, suggesting that it can influence susceptibility to prion disease through two distinct mechanisms. This study provides new insight into the misfolding pathway leading to critical oligomers of the prion protein and suggests a physical basis for increased assembly of the F198S mutant.


My role in this project involved running all $^{19}$F NMR experiments utilized in the main text (temperature, pressure, diffusion, saturation transfer) as well as data analyses for said NMR experiments. Karen Simonetti prepared all PrP samples and conducted DLS and SEC. Sameer Al-Abdul-Wahid ran preliminary $^{15}$N/$^1$H and $^{19}$F NMR vs pressure.
3.2 INTRODUCTION

A broad range of neurodegenerative diseases arise from the misfolding of endogenous proteins, leading to aggregation and protein deposition as amyloid fibril-rich deposits.\textsuperscript{1,2} For example, misfolding of the mammalian prion protein (PrP) leads to the onset of Creutzfeldt–Jakob disease (CJD), Gerstmann–Straussler–Scheinker syndrome (GSS), fatal familial insomnia, and kuru in humans, as well as bovine spongiform encephalopathy in cattle, scrapie in sheep, and chronic wasting disease in deer and elk.\textsuperscript{2–5} PrP is an endogenous glycosylphosphatidylinositol (GPI)-anchored protein associated largely with neuronal membranes, where it adopts a monomeric and predominantly \(\alpha\)-helical structure in the nondiseased state (PrP\textsuperscript{C}).\textsuperscript{6–8} Upon misfolding, which can occur sporadically in response to a destabilizing mutation, or through contact with its misfolded form (PrP\textsuperscript{Sc}), PrP converts to a \(\beta\)-sheet-rich, aggregation-prone state capable of forming a wide range of oligomeric structures, including the fibrillar form commonly associated with prion disease pathology.\textsuperscript{9–13} While there is overwhelming evidence that misfolding and aggregation of PrP is the cause of both neuronal cell death and the infectious nature of prion diseases,\textsuperscript{2,9,10,14,15} it is not yet clear exactly how cell death is induced or which oligomeric states of PrP are primarily responsible for disease pathogenesis. In vitro, misfolding of PrP into a \(\beta\)-sheet-rich conformation can be initiated by high pressure, low pH, or addition of denaturant.

In addition, hydrogen/deuterium (H/D) exchange data for PrP\textsuperscript{C} indicate the presence of residual structure in the unfolded state, within a region critical for pathological misfolding, suggesting that the process of PrP misfolding may proceed through unfolded intermediates in which regions containing defined structural elements are key to defining the PrP\textsuperscript{C}-to-PrP\textsuperscript{Sc} transition pathway.\textsuperscript{16,17} Kinetic studies have indicated the presence of partially ordered monomeric intermediates populated during PrP folding or misfolding, although the dominant secondary structure observed is dependent on solution conditions.\textsuperscript{18–21} There is also substantial evidence for the formation of several types of nonfibrillar, \(\beta\)-sheet-rich oligomers during the misfolding of PrP.\textsuperscript{22–27} Recently, both in vitro and in vivo cytotoxicities have been demonstrated to arise primarily from nonfibrillar species,\textsuperscript{9,25,28,29} rather than fibrils or monomeric protein. Similar cytotoxic nonfibrillar oligomers are also formed by the proteins implicated in several amyloid diseases and are suggested to be responsible for disease pathogenesis.\textsuperscript{30–35} Several
distinct nonfibrillar oligomers have been reported to form during in vitro misfolding of PrP, ranging in size from 10 to 20 nm to over 50 nm in diameter. Octameric or decameric oligomers have been identified as the minimum stable assembly formed during PrP misfolding, and are typically reported as spherical or discoidal. While there remains debate if these small oligomers exist on- or off-pathway to amyloid fibril formation, a recent study has linked a specific β-sheet-rich misfolding intermediate (the β-state) to prion disease susceptibility. The β-state is an early misfolded state populated under conditions of low pH and mild denaturant and comprises β-sheet-rich oligomers in equilibrium with a small population of non-native monomers, reminiscent of the molten globule monomer–octamer equilibrium suggested by the work of Gerber et al. In order to develop a complete understanding of the mechanism of PrP misfolding and assembly, in particular the early events surrounding formation of an initial stable nucleus, it is important to elucidate the forces driving the oligomerization of PrP. To date, studies have primarily focused either on the thermodynamics and kinetics of PrP folding/unfolding, or have tracked parameters associated with formation of oligomers from PrP. In this paper, we examine the dynamic equilibrium that exists in the β-state misfolding intermediate between a non-native β-monomer, a stable β-octamer, and larger prefibrillar aggregates formed by the both wild type (WT) Syrian hamster PrP(90–231) (ShaPrP(90–231)) and an F198S mutant. The F198S mutation in humans is linked to familial prion disease (GSS), and destabilizes the native form of the protein, thus promoting misfolding. ShaPrP(90–231) exhibits a high propensity to populate the β-state during unfolding, facilitating studies of this misfolding intermediate. PrP(90–231) is also commonly used in studies of PrP misfolding, since residues 23–90 are inherently disordered in PrPC and are not found within the protease-resistant core of PrPSc. Importantly, β-state samples of PrP(90–231) have previously been shown to maintain a stable equilibrium, even at elevated temperatures, for a period of weeks. Herein, we make use of 19F NMR to distinguish between assembled states present in β-state ShaPrP(90–231). 19F is a highly sensitive NMR nucleus and exhibits a sizable chemical shift dispersion in response to changes in conformation or electrostatic environment. Substitution of specific aromatic residues by their monofluorinated equivalents (in this case, 3-fluorophenylalanine) generally causes little or no structural perturbation. By selectively introducing 19F into either WT PrP(90–231) or its F198S mutant, it is possible to spectroscopically quantify monomeric and oligomeric species coexisting within
the β-state. The temperature and pressure dependence of the equilibrium then provides a thermodynamic perspective of the interactions which drive the formation of the octamer and prefibrillar states. In particular, enthalpy, entropy, and heat capacity differences between monomer, octamer, and prefibrillar aggregate states may be obtained by studying the equilibrium as a function of temperature, while pressure dependence provides information on specific volume and compressibility differences. At the same time, the equilibrium kinetics of exchange between states provides a perspective on the relative barriers involved in PrP oligomerization. This approach has allowed us to characterize early assembly events occurring during PrP misfolding. We distinguish the entropically favored formation of octamers from enthalpically driven assembly into larger species, and suggest that octamer formation from misfolded monomers is rate limiting. Additionally, an F198S mutation associated with GSS was shown to promote octamer formation, providing a new mechanism through which this mutation might lead to a pathogenic phenotype.

3.3 METHODS

Sample Preparation. Wild-type and F198S Syrian hamster PrP(90–231) were heterologously expressed in E. coli, using M9 minimal media enriched with $^{15}$NH$_4$Cl. Incorporation of 3-fluorophenylalanine was achieved through induced auxotrophy as described previously,$^{52}$ using 1 g/L glyphosate (to inhibit aromatic synthesis) and 75 mg/L of unlabeled tyrosine and tryptophan. When cells reached an OD of 0.8, 3-fluorophenylalanine was added at a concentration of 70 mg/L, and induced with 1 mM IPTG at 25 °C, overnight. Inclusion bodies were dissolved in 6 M guanidinium chloride (GuHCl), and His6-tagged ShaPrP(90–231) was then purified by Ni$^{2+}$ affinity chromatography. Refolding was achieved via rapid dilution into 55 mM Tris pH 8.2, 21 mM NaCl, 0.88 mM KCl, 1.1 M GuHCl, 400 mM L-arginine, 1 mM EDTA, 1 mM reduced glutathione, and 1 mM oxidized glutathione. Upon removal of the His-tag with TEV protease, natively structured PrP was converted to the β-state through dialysis in 10 mM sodium acetate pH 3.6, 150 mM NaCl, and 2 M urea for 48 h.$^{23}$ The F198S mutant was purified using a similar protocol, with thrombin for tag removal. Formation of the β-state was monitored using circular dichroism, and oligomeric β-state PrP was purified by gel filtration chromatography (Superdex 75). The equilibrium population of oligomer and β-state monomer was confirmed by gel
filtration after incubation overnight. Samples were dialyzed into 10 mM sodium acetate, pH 3.6, concentrated, and doped with 7–8% D2O for NMR.

NMR Experiments. All NMR experiments were performed on a 600 MHz (1H Larmor frequency) Varian Inova spectrometer (Agilent Technologies, Santa Clara, CA) using a cryogenic 5 mm HCN probe, capable of being tuned to either 1H or 19F. Standard 90° pulse widths were 17.8–18.5 µs. All spectra were referenced to the internal lock. Generally, spectra were obtained using 4096 scans and a repetition time of 1.3 s. For variable-temperature studies, each experiment was separated by a 20-min equilibration period. The spectrum at 28 °C served as the reference for all deconvolutions, since it exhibited the largest M1 peak intensity. Initial line width estimates at the reference temperature were derived from T2-filtered diffusion experiments where the oligomer peak could be fit to a single line shape, indicating the elimination of the large oligomeric component. Initially, Lorentzian line widths were fixed, while peak heights and chemical shifts were optimized. Chemical shifts and peak intensities were then fixed, and line widths were then optimized in the deconvolution, as a function of temperature. Successive rounds of fitting were applied until changes in line width, shift, and intensity were minimal.

Saturation Transfer Experiments. An array of saturation pulses (0.03, 0.06, 0.2, 0.5, 0.7, 1.1, 1.4, 1.8 s for F198S and 0.03, 0.06, 0.1, 0.2, 0.5, 0.7, 1, 1.1, 1.4, 1.8 s for WT PrP), corresponding to a B1 field of 32 Hz, was applied to the M1 monomer peak (−108.9 ppm). An off-resonance saturation pulse was applied at the transmitter offset (−119 ppm), and its duration varied according to the on-resonance presaturation length in order to maintain a constant saturation period of 2 s. A control experiment was run with the saturation pulse at an equidistant frequency upfield from the oligomer peak; no saturation of the oligomer resonances was detected in the case of the control. The data was fit to a two-site exchange as described by Helgstrand et al. 53 A global fit was applied, allowing variation in M(0), k, and T_1. T_1 was also separately measured using an inversion recovery experiment.

Pressure Experiments. 400 µL of WT or F198S ShaPrP(90–231)β at 3 mM were loaded into a ceramic high-pressure NMR tube pressurized using an Extreme-60 syringe pump (Daedalus Innovations). Light paraffin oil was used as the transducing fluid. Spectra were obtained as
above, using 8192 scans. All pressure experiments were conducted at 28 °C. The sample was equilibrated for 20 min at each pressure prior to acquisition. Deconvolutions were conducted in a manner similar to that for the temperature experiments, using initial parameters derived from T$_2$-filtered diffusion experiments at 28 °C, 1 bar.

**Diffusion Studies.** A pulsed-gradient spin–echo 19F diffusion experiment with variable gradient and t1 mixing period (Δ) was employed. The gradient periods served as a T$_2$-filter which eliminated large oligomer intensity at values above 1 ms. Monomer and octamer line widths were estimated from an experiment with a gradient period of 1.5 ms and a t1 mix time of 400 ms. The natural logarithm of normalized peak intensities were plotted as a function of $-\gamma^2 G^2 \delta^2 (\Delta - (\delta/3)) \times D$ according to the Stejskal–Tanner equation,$^{54}$ and the apparent diffusion coefficient (D$_{app}$) of each species was determined from the slope of the plot. Similarly, an experiment was optimized for detection of large oligomer intensity. Due to exchange between states, the apparent diffusion coefficients were not suitable for quantitative analysis. Thus, the relative D$_{app}$ values were used solely to support the assignment of monomer, octamer, and large oligomer resonances.

**Thermodynamic Analysis.** The equilibrium constant, k$_{ij}$, between two states, i and j, can be directly obtained from the ratio of corresponding peak areas in a one-dimensional 19F NMR spectrum, assuming that states are not in intermediate or fast exchange with each other. The free energy associated with this equilibrium is given by

$$\Delta G_{ij} = -RT \ln k_{ij}$$

(1)

where R represents the gas constant. As it is difficult to obtain an accurate size for the higher-order aggregation states of PrP, we confine the present analysis to the use of effective equilibrium constants which do not specifically account for the aggregation number.$^{55}$ The temperature dependence of the free energy is in turn expressed in terms of the enthalpy, entropy, and heat capacity differences between the states, such that:

$$\Delta G_{ij} = \Delta H_{ij} - T \Delta S_{ij} + \Delta C_{p,ij}[T - T_0 - T \ln(T/T_0)]$$

(2)

where T$_0$ refers to a reference temperature at which the enthalpy and entropy are defined. Assuming the enthalpy can be approximated as a constant over the full temperature range, the equilibrium constant can be expressed in terms of inverse temperature:
\[ \ln k_{ij} = -\Delta H_{ij}/RT + \Delta S_{ij}/R \] (3)

Similarly, the pressure dependence of the free energy provides a measure of the differences in both specific volume, \( \Delta V_{ij} \), and isothermal compressibility, \( \Delta \kappa_{ij} \), between states:

\[ \Delta G_{ij} = \Delta G_0 + \Delta V_{ij}(P - P_0) - 0.5\Delta \kappa_{ij}(P - P_0)^2 \] (4)

where \( \Delta G_0 \) refers to the free energy difference at the reference temperature and pressure.

Dynamic Light Scattering. DLS experiments were performed using a 5 mm × 5 mm quartz cuvette in a PDDLS/Cool Batch 90T detector (Precision Detectors Inc., Bellingham, MA, U.S.A.). 300 µL aliquots of 20 µM \( \beta \)-state ShaPrP(90–231) in 10 mM sodium acetate pH 3.6 were used for all measurements. Precision Deconvolve\textsuperscript{32} software was used for data acquisition and analysis. Sampling time was 10 or 15 µs, and 10–12 scans were averaged for each data set. Scattering data were obtained at several temperatures between 20 and 80 °C, allowing the sample to equilibrate before each acquisition. Reversibility was confirmed by acquiring low-temperature data after each series. Literature values for viscosity were used for data analysis.

3.4 RESULTS

\textsuperscript{19}F NMR Spectra Reveal the Coexistence of \( \beta \)-Monomers, \( \beta \)-Octamers, and Higher-Order Oligomers. Samples of \( \beta \)-state ShaPrP(90–231) were prepared following previously reported protocols and their secondary structure and oligomeric state confirmed using gel filtration, DLS, and CD spectroscopy.\textsuperscript{23} Figure 3.1 presents \textsuperscript{19}F NMR spectra of WT and F198S ShaPrP(90–231), labeled with 3-fluorophenylalanine.

\textbf{Figure 3.1} \textsuperscript{19}F NMR spectra of 3-fluorophenylalanine-labeled ShaPrP(90–231). Spectra (black lines) are shown for both WT (left) and F198S (right) samples containing 3 mM protein, 10 mM sodium acetate, pH 3.6, and 8% D\textsubscript{2}O. The deconvolved peaks, shown in blue, are assigned the \( \beta \)-monomer (one for each 3-fluoro-Phe residue, designated as \( M_1 \), \( M_2 \), and \( M_3 \)), \( \beta \)-octamer, and larger oligomers. Only two monomer peaks are observed for the F198S mutant, which lacks a third Phe residue. The residual error from peak fitting is shown in red. Spectra represent 4096 transients, obtained at 28 °C, at a 600 MHz \textsuperscript{1}H field strength.
Previously reported gel filtration and sedimentation velocity data of PrP(90−231)β prepared at a lower concentration, using an otherwise identical sample preparation protocol, are consistent with the dominant species being octameric with only a small fraction of monomeric PrP present in the β-state.23 Likewise, at temperatures below 60 °C, DLS measurements of our ShaPrP(90−231)β samples show a single dominant species corresponding to the expected hydrodynamic radius (~6 nm) of an octamer (Figure S3.1 in Supporting Information [SI]). We therefore ascribe the well-resolved minor peaks in the 19F NMR spectra to the non-native β-monomer for both WT and F198S ShaPrP(90−231). This assignment is supported by these resonances exhibiting longer T2 relaxation times (Table S3.1 in SI) and faster NMR diffusion rates (Figure S3.2 in SI) than the broader peaks attributed to oligomeric PrP. The monomer resonances are designated M1, M2, and M3 for WT PrP, corresponding to each of the fluoro-Phe residues (residues 139, 173, and 198) present in the WT sequence. While we cannot unambiguously assign F139 and F173, the absence of the M3 peak in the spectrum of the F198S mutant allows us to assign that resonance to F198. Thus, only M1 and M2 are present in spectra of the F198S ShaPrP(90−231). F139 is expected to lie within the disordered region of the protein, while both F173 and F198 are within ordered regions of both the native and fibrillar forms of PrP(90−231). Therefore, M1 is likely to correspond to F139, while the two peaks with similar shifts correspond to F173 and F198. Note that site-specific assignment is not required for the present analysis of equilibrium populations of monomeric and oligomeric ShaPrP(90−231)β. Rather, as long as each state can be spectroscopically resolved over a sufficient range of temperatures and pressures, the thermodynamic and volumetric parameters defining the differences between the states can then be determined, regardless of the location of the probe. The upfield resonance associated with the β-state oligomer is significantly broader, such that each of the phenylalanine resonances appears equivalent, within the NMR line width. While the broad component of the 19F spectrum was initially considered to arise solely from the β-octamer observed by DLS, spectral deconvolution reproducibly indicated the presence of two peaks, as shown in Figure 3.1. The 19F spectrum cannot be accurately fit using a simple two-state (monomer–octamer) deconvolution. Moreover, both broad resonances are consistently observed over the entire temperature range of our NMR experiments (4–43 °C) and exhibit an inversely correlated change in intensity as a function of temperature. This suggests that these components
represent an equilibrium between two distinct oligomeric states of PrP. NMR diffusion measurements, utilizing a $^{19}$F NMR stimulated echo pulse sequence,\textsuperscript{54,57} confirm that the broad resonances exhibit distinct diffusion rates, consistent with the presence of both an octamer and a larger oligomer (Figure S3.2 in SI). Furthermore, the component assigned to larger oligomers is completely abolished in Hahn-echo spectra with long delays to filter out fast-relaxing species, leaving only the longer-lived octamer and monomer peaks. Such T$_2$-filtered spectra can be fit with a single line shape to the octamer. As shown in Figure S3.1 in SI, DLS measurements obtained at elevated temperatures also show the reversible formation of a larger oligomeric species, consistent with the NMR diffusion data. The relatively small population of the larger oligomer seen by DLS can be attributed to those measurements being obtained at a concentration approximately 100 times lower than required for NMR. Robust deconvolution of the $^{19}$F NMR oligomer peak to obtain populations for the octamer and the larger oligomers relies on accurate estimates of the line widths for each component. In stimulated echo diffusion experiments with gradient duration longer than 1.5 ms, no large oligomer intensity could be detected, allowing determination of the octamer line width (Figure S3.2 in SI). These diffusion-filtered spectra provide reliable estimates of the line widths for both oligomeric species present in β-state PrP, allowing the accurate deconvolution of the $^{19}$F spectra shown in Figure 1. T$_2$ relaxation (Hahn-echo) experiments also distinguish the octamer from the prefibril.\textsuperscript{58} Using deconvolution of spectra obtained with varied echo times, and line width estimates from the T$_2$-filtered and diffusion-filtered spectra, it is possible to obtain T$_2$ relaxation times for each oligomeric species within the β-state, as reported in Table S3.1 in SI. The significantly shorter T$_2$ measured for the larger oligomer relative to octamer agrees with our assignment of these two species.

**Thermodynamic Analysis of Monomer–Oligomer Equilibria of ShaPrP(90–231)$^\beta$.**

The identification of resolved resonances arising from β-state monomer and oligomeric PrP provides an opportunity to investigate the equilibria between these species as a function of temperature (16 - 43 °C) and pressure (1–1400 atm). Figure 3.2 shows the temperature dependence of the β-monomer and oligomer fractions for WT and F198S ShaPrP(90–231)$^\beta$, based upon the relative normalized peak areas derived from spectral deconvolution. Monomer fractions remain very small over this temperature range but consistently decrease with temperature, with a corresponding increase in the total oligomer fraction.
For both PrP sequences, temperatures above 31 °C favor the β-octamer, whereas lower temperatures favor the larger oligomers. Given the overlap between the broadened resonances corresponding to octamer and large oligomers, we have performed the analysis in two ways: 1) two-states, considering only the monomer and total oligomer (the sum of octamer and larger oligomer), and 2) three-states, where we distinguish monomer, octamer and larger oligomer based upon spectral deconvolution. In either case, the ratios of any two states directly relates to the free energy difference between states. By making use of eq 2, it is possible to then fit the temperature dependence of the resulting free energy terms (Figure 3.3), thereby providing an estimate of the enthalpy, entropy, and heat capacity differences between monomer and oligomer (octamer or large oligomer) or between β-octamer and large oligomer.

**Figure 3.2** Equilibrium populations of β-state monomer, octamer, and large oligomers formed by ShaPrP(90−231) as a function of temperature. Relative normalized peak volumes were obtained from deconvolution of 19F NMR spectra recorded at temperatures from 4−43 °C. The total oligomer fraction (calculated as the sum of the octamer and large oligomer populations) is also shown for both PrP sequences. In the case of F198S, the monomer profile was obtained by doubling the relative integral of the M₁ peak (the M₁ integral was unreliable as an estimate of the monomer population due to its broadened resonance). For the WT data, the monomer fraction was obtained using \((3/2)×(M_{2} + M_{3})\) to obtain an estimate on the basis of the average over the two monomer peaks. Curves are shown to guide the eye.
The results of the thermodynamic analysis, shown in Table 3.1, reveal that the formation of oligomeric species from monomer is strongly driven by entropy, while the enthalpy change associated with this initial aggregation step is unfavorable.

**Table 3.1** Thermodynamic Parameters Associated with the Temperature-Induced Transition from Monomer to Oligomer and Octamer to Large Oligomer for WT and F198S PrP(90-231)

<table>
<thead>
<tr>
<th>Transition</th>
<th>WT</th>
<th>F198S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomer to total oligomer (WT)</td>
<td>ΔH: 5.83 ± 1.15</td>
<td>ΔH: 8.83 ± 1.98</td>
</tr>
<tr>
<td></td>
<td>ΔS: 40.00 ± 4.00</td>
<td>ΔS: 50.00 ± 6.00</td>
</tr>
<tr>
<td></td>
<td>ΔC_p: −2.00 ± 0.40</td>
<td>ΔC_p: −3.20 ± 0.71</td>
</tr>
<tr>
<td>Monomer to octamer (WT)</td>
<td>ΔH: 55.00 ± 2.30</td>
<td>ΔH: 45.70 ± 6.42</td>
</tr>
<tr>
<td></td>
<td>ΔS: 198.00 ± 7.55</td>
<td>ΔS: 170.10 ± 21.46</td>
</tr>
<tr>
<td></td>
<td>ΔC_p: −4.30 ± 0.51</td>
<td>ΔC_p: −2.97 ± 1.45</td>
</tr>
<tr>
<td>Octamer to large oligomer (WT)</td>
<td>ΔH: −75.50 ± 7.17</td>
<td>ΔH: −65.50 ± 5.53</td>
</tr>
<tr>
<td></td>
<td>ΔS: −250 ± 24.00</td>
<td>ΔS: −220 ± 18.00</td>
</tr>
</tbody>
</table>

*Changes in enthalpy (ΔH), entropy (ΔS), and heat capacity (ΔC_p) are reported. In the three-state model, formation of the large oligomer from octamer did not show a change in heat capacity (no second-order term is present when fitting the temperature-dependent free energy profile of this transition).
The favorable entropic term for oligomerization presumably arises from burial of hydrophobic residues and decreased hydration of misfolded PrP upon formation of the octamer. On the other hand, equilibrium data associated with the transition between octamers and larger oligomers suggests a process that is entropically unfavorable although strongly enthalpically favorable. As shown in A and B of Figure 3, the F198S mutant exhibits significantly lower $\Delta G$ values associated with oligomer formation from monomeric protein. This mutation appears to favor the monomer–octamer transition, whereas little difference is seen for subsequent assembly of octamers into larger oligomers (Figure 3.3C).

**Pressure-Induced Changes in Specific Volume and Isothermal Compressibility for the Monomer–Oligomer and Octamer–Large Oligomer Transitions of ShaPrP(90–231)$^\beta$.**

The equilibrium between $\beta$-state PrP monomers and oligomers was also studied as a function of pressure. The monomer, octamer, large oligomer, and total oligomer fractions were obtained through deconvolution of $^{19}$F NMR spectra acquired at pressures from 1 to 1400 atm. As shown in Figure 3.4, WT ShaPrP(90–231)$^\beta$ exhibits a reduction in total oligomer as pressure is increased, with a concurrent increase in the monomer population. Similarly, there is an apparent pressure-induced dissociation of the large oligomers, leading to an increase in octamer.

![Figure 3.4](image-url)

**Figure 3.4** Normalized populations of monomeric and oligomeric ShaPrP(90–231)$^\beta$ as a function of pressure. The normalized peak areas obtained from deconvolution of $^{19}$F spectra recorded as a function of pressure are shown for WT (top) and F198S (bottom) ShaPrP(90–231)$^\beta$. Peak fitting was performed for each sample as described in Figure 2. The total oligomer fraction represents the sum of the octamer and large oligomer populations. Curves are shown to guide the eye.
Similar analysis of the F198S β-state revealed a significantly weaker dependence of oligomerization on pressure, preventing accurate determination of free energies of association for the mutant protein. The free energies calculated for the pressure-induced monomer–total oligomer and octamer–large oligomer transitions of WT ShaPrP(90–231)β are provided in Figure 3.5.

The changes in partial specific volume (ΔV) and isothermal compressibility (Δκ_m) for each transition were obtained using eq 4 (Table S3.2 in SI). While compressibility was largely unaffected, the negative value for ΔV indicates that the system undergoes a collapse upon formation of oligomers from monomers, consistent with burial of hydrophobic surface suggested by the favorable entropy of this event. On the other hand, formation of large oligomers from octameric ShaPrP(90–231) is associated with a positive change in specific volume, possibly due to conformational rearrangements accompanying further assembly.

**Exchange of ShaPrP(90–231) between β-Monomer and Oligomer States Occurs on a Time Scale of 4–10 s.**
The chemical shift separation between one of the β-monomer resonances (M1) and the oligomer peaks makes it feasible to employ a $^{19}$F NMR saturation transfer experiment to monitor the kinetics of ShaPrP(90–231) monomer–oligomer exchange within the β-state. As shown in Figure 3.6 for the WT sequence, there is a pronounced decay of the NMR signal arising from the oligomer peaks with increasing time of the saturation pulse selectively applied to the M1 peak, while the other monomer peaks (M2 and M3) remain constant. A similar response is observed for the F198S mutant, as shown in Figure S3.3 in SI.

![Figure 3.6](image)

**Figure 3.6** $^{19}$F NMR saturation transfer measurements on ShaPrP(90–231)$^\beta$. Magnetization decay profiles of total oligomer (purple), octamer (red), and large oligomer (blue) components for WT ShaPrP(90–231)$^\beta$ (A) and F198S (B) after selective saturation of the M1 β-monomer resonance for a series of mix times. No effective decay of the M2 or M3 β-monomer resonances (Figure S3) is observed. Control experiments, in which the saturation pulse was applied at an upfield frequency equidistant from the oligomer peaks, showed no signal decay. Here, $k_{ex}$ and $T_1$ represent the fitting parameters associated with the decay profile. Error bars represent the RMS noise in the original spectra.

Fitting each decay profile to an equation corresponding to a two-state exchange model which incorporates the measured spin–lattice relaxation times of monomer and oligomer, and an exchange rate constant $k$, gave rate constants for monomer–total oligomer exchange of ~5.5-6.6 Hz for both F198S and WT ShaPrP(90–231) (Table 3.2).$^{53}$ Deconvolution of the oligomer peak as described above allows estimation of the monomer–octamer and monomer–large oligomer exchange rates, giving 1.26-1.82 Hz and 4.29-4.34 Hz, respectively.
Table 3.2 Exchange Rate Constants Observed for WT and F198S ShaPrP(90-231)β

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th></th>
<th>F198S</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>forward rate $k_{BA}$ (s$^{-1}$)</td>
<td>reverse rate $k_{AB}$ (s$^{-1}$)</td>
<td>forward rate $k_{BA}$ (s$^{-1}$)</td>
<td>reverse rate $k_{AB}$ (s$^{-1}$)</td>
</tr>
<tr>
<td>Large Oligomer-M</td>
<td>0.24</td>
<td>4.34</td>
<td>0.19</td>
<td>4.29</td>
</tr>
<tr>
<td>Octamer-M</td>
<td>0.09</td>
<td>1.26</td>
<td>0.08</td>
<td>1.82</td>
</tr>
<tr>
<td>Total Oligomer-M</td>
<td>0.16</td>
<td>5.55</td>
<td>0.14</td>
<td>6.63</td>
</tr>
</tbody>
</table>

The forward rate ($k_{BA}$) corresponds to oligomer exchange to monomer whereas the reverse rate constant ($k_{AB}$) represents exchange in the direction of monomer to oligomer.

While there is an additional source of error introduced by the deconvolution process, which is compounded by the presence of octamer–oligomer exchange, our results suggest a faster exchange of monomer with the larger oligomers than with the octamers. Our kinetic analysis, based on the knowledge of the equilibrium constants, allows us to assemble both forward and reverse exchange rates. This is illustrated in Figure 3.7, which summarizes the estimated equilibrium populations and exchange rates between each of the three observed oligomeric states coexisting within the misfolded ShaPrP(90–231)β.

**Figure 3.7** Model for equilibrium exchange between assembled states within ShaPrP(90–231)β. The exchange processes at 28 °C and ambient pressure, observed between monomers, octamers, and larger oligomers of ShaPrP(90–231)β are indicated by arrows connecting the different species, with the corresponding rate constants (s$^{-1}$) and equilibrium populations (fraction of total, bold) given in blue (WT) and red (F198S mutant). The curved arrow indicates the possibility of monomer–oligomer exchange occurring only through an octameric intermediate, although our kinetic data support a direct addition of monomers to growing large oligomers.
Consistent with a role for F198S mutation in stabilizing the β-octamer, the rate of dissociation of octamer to monomer is approximately 30% that of the WT protein. The reversibility of the temperature and pressure data and the lack of aggregation in our samples over extended timeframes (months) suggest that further perturbation, such as high salt or agitation, is required to promote fibril formation from the β-state. Therefore, only the three states identified above are presented in Figure 3.7.

3.5 DISCUSSION

The formation of intermediates along the misfolding pathway of PrP is well established and occurs in response to unfolding or destabilization of native PrPC by temperature, pressure, or denaturant. Importantly the exact nature of the intermediate states is sensitive to the PrP sequence being studied and, in particular, to the solution conditions of the experiment. This has led to the identification of partially ordered monomeric folding intermediates\textsuperscript{19−21,42} as well as β-sheet-rich oligomers.\textsuperscript{21,22,26,36−40,59} While one report identifies a putative molten globule monomer with significant α-helical secondary structure,\textsuperscript{21} the majority of the data in the literature supports the formation of non-native structures in the early stages of PrP unfolding or misfolding. In particular, several reports have shown that there is a minimum stable oligomer, generally identified as octameric, which is populated during misfolding of both full-length PrP and PrP(90–231).\textsuperscript{21,22,26,36,37,39} While some groups claim this is a key intermediate populated on the pathway to fibrillization,\textsuperscript{21,26,39} others suggest that these small oligomers are off-pathway intermediates, requiring dissociation and conformational rearrangement of the monomer before fibril assembly can take place.\textsuperscript{22,37,38,47}

Irrespective of their potential role in formation of PrPSc or other amyloid fibrils, misfolding of PrPC to cytotoxic β-sheet-rich assemblies remains important. The β-state intermediate in this study shares many features with reported β-sheet octamers and non-native monomers and has been shown to correlate with disease susceptibility and to exhibit cytotoxicity in vitro, making it a potentially important player in prion disease pathogenesis.\textsuperscript{23,28} Additionally, the similarity of nonfibrillar PrP oligomers to those observed in amyloid and protein misfolding diseases\textsuperscript{30,31,60} strongly suggests that an understanding of early events in PrP misfolding and assembly will also shed light on the mechanisms underlying the onset of these diseases. Here we
have examined the thermodynamic and kinetic parameters governing exchange between a misfolded monomer of ShaPrP(90–231), a relatively stable β-octamer, and a larger oligomer, using $^{19}$F NMR, which is highly sensitive to chemical exchange and molecular dynamics. These misfolded states were populated under conditions previously shown to promote formation of a β-state ensemble containing non-native monomer and octamer in equilibrium. Sokolowski et al. have also reported formation, under similar conditions of low pH and denaturant, of an ensemble of oligomers, with an octamer as the minimal stable size.

Our thermodynamic analysis of the equilibria existing within the β-state support the concept that formation of a stable β-octamer represents a key step in assembly of misfolded PrP. The favorable entropic term and unfavorable enthalpy associated with oligomerization of the monomer (Table 3.1) is reminiscent of the energetics determining initiation of amyloid fibril formation, recently described by Buell et al. Following this model, the enthalpic barrier to formation of oligomers is indicative of a loss of stable intramolecular contacts within the non-native monomer. The barrier is too low to prevent oligomer formation, likely due to the destabilizing conditions under which the β-state is formed. Native intramolecular contacts have already been disrupted and replaced with a presumably less favorable set of interactions stabilizing a partially ordered non-native monomer.

Assembly from the monomeric intermediate is then entropically driven, suggesting a burial of hydrophobic groups that have been exposed following the loss of native structure and a concurrent desolvation of the oligomer. Formation of the larger oligomers is then enthalpically driven, suggesting that this process represents addition of octamers and monomers to a stable β-sheet core. We cannot rule out the possibility that conformational rearrangement is required for further assembly of the octamer into large oligomers, since the time scales involved are slow (seconds). Interestingly, exchange of monomer with the large oligomers occurs at a 3 to 4x higher rate than exchange with octamers, again supporting the notion of the octamer as a stable intermediate, whose formation may represent the rate-limiting step for assembly of misfolded PrP. Our thermodynamic model is supported by both the significant decrease in heat capacity for the oligomeric states and by the pressure data, which suggest a more tightly packed structure relative to monomeric protein. It is important to note that the dissociation observed at increased pressures in our studies is consistent with previous investigations of PrP assembly under
pressure.\textsuperscript{38,45} Significantly higher pressures (~4000 atm) have been shown to reverse this trend, promoting aggregation. Similarly, at the temperatures used here, heating induces a dissociation of the large oligomer, favoring the octameric intermediate. Higher temperatures might be expected to promote higher-order aggregation, as suggested by our DLS data (Figure S3.1 in SI).

Overall, the model presented in Figure 3.7 is consistent with that of Sokolowski \textit{et al.},\textsuperscript{26} as well as with the general models presented by Gerber \textit{et al.}\textsuperscript{21,39} The latter propose assembly of a putative decamer from a partially folded monomeric intermediate of human PrP(90–231), followed by subsequent stacking of these small oligomers into larger aggregates. Sokolowski \textit{et al.} also present data suggesting that an octameric intermediate of ShaPrP(90–231) exists on-pathway to fibril formation, although fibrillization was extremely slow (55 days), and beyond the time scale of our current experiments. The kinetic data summarized in Figure 7 also suggest that formation of a stable small oligomer (i.e., β-octamer) from misfolded monomers may be a rate-limiting step in assembly, consistent with the concept that nucleation events are required to initiate the growth of amyloid fibrils in vitro. While the relationship of the β-state oligomers discussed here with assembly of amyloid fibrils or PrP\textsuperscript{Sc}-like structures remains unclear, the evidence for further assembly of the observed octamers into larger oligomers suggests the potential for nucleation of fibrillization. Further work will be required to define the on-/off-pathway nature of the β-state intermediates. In addition to obtaining insight into the exchange between misfolded states for WT PrP(90–231), we examined the impact of a GSS-related mutation on the β-state ensemble. F198S is known to destabilize the native state of PrP, leading to spontaneous conversion to protease-resistant fibrils.\textsuperscript{18} Apetri \textit{et al.}\textsuperscript{41} have previously demonstrated that certain prion disease related mutations can alter the denaturant folding kinetics of human PrP(90–231), promoting the formation of a partially folded (monomeric) intermediate. The F198S mutation in particular had the largest effect of the nine mutant proteins studied, leading to a 100x increase in the population of this folding intermediate in the absence of denaturant. Here, we report a significant decrease in the free energy associated with oligomer formation by the F198S mutant of ShaPrP(90–231), favoring formation of the octameric intermediate. This is the first indication that the F198S mutation may play a role in promoting assembly of misfolded PrP, in addition to its previously characterized role in destabilizing the native state. Our data also provide a proof-of-principle that examining the equilibria between
misfolded states may shed light on the role of mutations and other sequence alterations in prion misfolding.

3.6 CONCLUSIONS

In conclusion, we have presented a detailed analysis of the thermodynamic and kinetic parameters governing the exchange between three distinct misfolding intermediates of ShaPrP(90–231). This supports an entropically driven assembly of an initial octameric structure, and provides some evidence for further assembly of this minimal unit into larger oligomers, through an enthalpically favorable process. Under the conditions used here, exchange between states is slow, occurring on the time scale of seconds. Surprisingly, a GSS-associated F198S mutation was found to promote oligomerization within the misfolded state, suggesting a dual role for this mutation in initiation of prion disease. Overall, our $^{19}$F NMR approach has demonstrated the ability to obtain thermodynamic and kinetic parameters that define early aggregation events occurring during protein misfolding. This should be an increasingly useful tool for studying amyloid aggregation, especially when combined with recent developments in structural studies of nonfibrillar oligomers.
3.7 ASSOCIATED CONTENT

$^{19}F$ T2 relaxation times and estimated line widths (Table S3.1), and analysis of the pressure-induced monomer–oligomer transitions (Table S3.2). DLS measurements (Figure S3.1), diffusion edited $^{19}F$ NMR spectra (Figure S3.2), and negative controls for the NMR saturation transfer experiments (Figure S3.3). This material is available free of charge via the Internet at http://pubs.acs.org.

Table S3.1 Estimated line widths ($\Delta v$) at 28°C for the most downfield monomeric species (M1), the octamer, and the prefibrillar aggregate, based upon spectral deconvolution and T2 relaxation measurements of F198S ShaPrP(90-231)$^6$.

<table>
<thead>
<tr>
<th>$^{19}F$ NMR Peak assignment</th>
<th>Actual $^{19}F$ T$_{2\text{eff}}$ for F198S based on spectral deconvolution</th>
<th>Calculated $^{19}F$ T$_2$ for F198S based on Hahn echo experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta v$ (Hz)</td>
<td>T$_{2\text{eff}}$ (s)</td>
</tr>
<tr>
<td>M1</td>
<td>48.5</td>
<td>0.0066</td>
</tr>
<tr>
<td>M2</td>
<td>72.2</td>
<td>0.0044</td>
</tr>
<tr>
<td>PFA</td>
<td>497.5</td>
<td>0.0006</td>
</tr>
<tr>
<td>Oct</td>
<td>190.5</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

Table S3.2 Estimates of the thermodynamic parameters associated with the pressure-induced transition from monomer to oligomer and octamer to large oligomer for WT PrP(90-231)$^6$. The estimated changes in specific volume ($\Delta V$) and isothermal compressibility ($\Delta K_m$) were derived from $^{19}F$ spectra recorded as a function of pressure. The changes are defined with respect to a reference pressure, $P_0$, of 1 atm. Note that only data for WT protein is shown, since the lack of prominent changes in populations of the various forms of $\beta$-state PrP for F198S as a function of pressure (Figure 4) prevents reliable fitting to equation 4. As observed for the temperature data (Table 1), there was no second order term for the free energy of forming the large oligomer as a function of pressure. Thus, no change in compressibility was measurable for this transition.

<table>
<thead>
<tr>
<th>Monomer to Total Oligomer</th>
<th>$\Delta V$ (mL·mol$^{-1}$)</th>
<th>$\Delta K_m$ (mL·mol$^{-1}$·bar$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.4710 ± 0.1320</td>
<td>-0.0016 ± 0.0002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monomer to Octamer</th>
<th>$\Delta V$ (mL·mol$^{-1}$)</th>
<th>$\Delta K_m$ (mL·mol$^{-1}$·bar$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.0070 ± 0.1589</td>
<td>-0.0015 ± 0.0002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Octamer to Large Oligomer</th>
<th>$\Delta V$ (mL·mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.4200 ± 0.0070</td>
</tr>
</tbody>
</table>
Figure S3.1 Dynamic light scattering analysis of ShaPrP(90-231)β. Example analyses of DLS data obtained for β-state ShaPrP(90-231)F198S, at 20° and 70°C are shown in (A) and (B) or (C) and (D), respectively. Samples contained 20 µM protein in 10 mM sodium acetate, pH 3.6. In each case, the distribution of Rh is plotted (A) and (C), as well as the distribution function (B) and (D). The fraction of Rh (hydrodynamic radius) corresponding to the β-octamer is indicated in (A), while the presence of larger oligomers is indicated in (C). A plot of Rh as a function of sample temperature is shown in (D). Each data point is the average of at least 3 independent measurements, with scatter indicated by the error bars. At temperatures over 60°C, a second peak is observed, corresponding to larger oligomers, as indicated by the significantly higher Rh for this species. Irreversible aggregation was ruled out by rerunning low temperature curves after sample heating. The relative population (percent) of each species as a function of temperature is indicated by the numbers next to each data point.
Figure S3.2 Diffusion filtered $^{19}$F NMR spectra and physical diffusion profiles of β-state F198S ShaPrP(90-231). (A) $^{19}$F NMR diffusion-edited spectrum of F198S ShaPrP(90-231) acquired with a short gradient duration, such that signals are observed for monomer, octamer, and large oligomer. (C) T2-filtered diffusion-edited spectrum where a longer gradient period eliminates signal from fast-relaxing species (large oligomer). Whereas the peak representing the sum of octamer and large oligomer at -113 ppm, is fit to two peaks in (A), only a single line shape can be fit to this resonance in (C). In each spectrum, the experimental data is shown in black, with the deconvolved spectrum (blue), the reconstructed spectrum (green) and residual error (red). (B) Plot of the logarithm of the normalized intensity versus the parameter describing gradient strength for octamer and large oligomer corresponding to the spectrum shown in (A). (D) A similar plot for monomer and octamer diffusion, corresponding to the spectrum shown in (B). Parameters for the diffusion experiment described by A and B are as follows: Diffusion period = 600 ms, gradient duration = 100 µs. The parameters for C and D are: Diffusion period = 400 ms, gradient duration = 1.5 ms. For plots B and D, the natural logarithm of the normalized peak intensities ($I/I(0)$) can be plotted as a function of effective gradient strength according to the Stejskal-Tanner equation: $I(G) = I(0)*\exp[-\gamma^2G^2\delta^2(\Delta - \delta/3)*D_{app}]$, allowing determination of the apparent diffusion coefficient ($D_{app}$) from the slope. Note that exchange between states confounds the diffusion measurements, and the necessity of using low gradient duration in the octamer-large oligomer experiment introduces inaccuracies in the estimates of $D_{app}$. Thus, Dapp values obtained from plots C and D are not accurate estimates of the physical diffusion for any of the states, rather they allow comparison of relative diffusion rates – distinguishing large oligomers from octamers and monomers.
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Carlene Starck and Dr. Avi Chakrabartty for useful discussions. This work was funded by an operating grant to S.S. from PrioNet Canada. S.S. is a Canada Research Chairs Program chairholder (Tier II). R.S.P. acknowledges NSERC (Grant 261980) for a research discovery award.

3.8 REFERENCES


(60) Walsh, P.; Sharpe, S. In Neurodegenerative Diseases, Book 2; Chang, R. C.-C., Ed.; InTech Open Access Publisher : New York, 2011.
CHAPTER 4

Site-specific labeling of protein lysine residues and N-terminal amino groups with Indoles and Indole-derivatives

4.1 ABSTRACT

Indoles and indole-derivatives can be used to site-specifically label proteins on lysine and N-terminal amino groups under mild, non-denaturing reaction conditions. Hen egg white lysozyme (HEWL) was labeled with either fluoroindole or fluoroindole-2-carboxylate via electrophilic aromatic substitutions to imines formed in-situ in the presence of formaldehyde. The reaction is highly site-selective, is easily controlled by temperature, and does not eliminate the native charge of the protein, unlike many common lysine-specific labeling strategies. $^{19}$F NMR demonstrates that the fluoroindole tags are exquisitely sensitive to local environment and a computational study indicates that they are more sensitive than most traditional $^{19}$F NMR probes used for protein labeling. Given the site-specificity of these protein tags, their sensitivity to local chemical environment, the mildness of the reaction conditions (aqueous, buffered or unbuffered), and the low stoichiometry required for reaction, indole-derivatives should serve as a valuable addition to the bioconjugation toolkit. Whereas our focus was on applications to $^{19}$F NMR and to a lesser extent, mass spectrometry, it is evident that this labeling method can be adapted for a wide range of applications, given that it serves as a general bioconjugation strategy.

4.2 INTRODUCTION

Bioconjugation reactions play an important role in the study of macromolecules and are increasingly used in the synthesis of novel biomaterials and functionalized therapeutics (Harris and Chess 2003, Lundblad 2004, Pasut and Veronese 2006, Kalia and Raines 2010, Biju 2014). In addition, reactions which serve to site-specifically incorporate chemical tags into proteins are of particular importance in biochemistry as they provide ways of probing and/or chemically modifying specific residues that may be responsible for key structural and/or functional properties. Acetylation (acylations), methylation (alkylations), as well as succinimide- and maleimide based conjugations are all common techniques that target lysine or cysteine residues
for the purpose of either conferring new properties (and possibly functionality), or serve to introduce labels (isotopes, fluorophores, radicals, complexes) into the macromolecule of interest (Panchuk-Voloshina et al. 1999, Lundblad 2004, Hermansen 2008, Kahsai et al. 2011, Kuchar et al. 2012).

Lysine residues in particular serve as convenient labeling sites, due in part to their abundance of roughly 6-7% in proteins and their tendency to be solvent-exposed. (King and Jukes 1969, Hattori et al. 2013, Larda et al. 2012) As with arginine they tend to confer favorable solubility to the macromolecule given their net positive charge at physiological pHs. In addition, lysine residues are often involved in structurally and functionally important intra-, inter-domain, and inter-protein interactions such as cation-pi, hydrogen bonds, and salt-bridges, for which the net charge is essential (André et al. 2007) With regard to lysine side chain and N-terminal amine tagging, many existing approaches employ reagents such as anhydrides, acyl halides, thioacetates and active esters, isocyanates and isothiocyanates, and sulfonamides; all of which eliminate the net positive charge by converting the amine to a neutral conjugate (Means and Feeney 1971, Lundblad 2004, Hermanson 2008, Walker 1996, Syed et al. 1992). The elimination of the native charge of the protein can be perturbing to both structure and function and is therefore generally undesirable, especially at higher levels of labeling. Here we report a chemical tagging approach that employs indoles to site-specifically label lysine residues, under mild, non-denaturing, aqueous conditions. This labeling strategy preserves the native charge of the lysine side chain amino group and allows for the incorporation of functionalized/substituted indoles for a wide range of potential applications including: pegylation, protein immobilization, isotopic tagging for NMR, ICAT tagging for tandem mass spectrometry, incorporation of \( ^{18} \)F PET tracers, crosslinking, fluorophore labeling, UV-Vis, and intrinsic fluorescence.

4.3 MATERIALS AND METHODS

Formaldehyde (37% w/w) solution, egg white lysozyme, alpha-lactalbumin, 5-fluoroindole, and 5-Fluoroindole carboxylate were all purchased from Sigma Aldrich (Mississauga, ON, Canada).

Reactions with 5-fluoroindole or 5-fluoroindole carboxylate. Lyophilized lysozyme was dissolved in milli-Q H\(_2\)O (EMD Millipore) to a concentration of \(~200-250\) µM. An aliquot of
12C-formaldehyde solution (37% w/w) was added at a 3x stoichiometry relative to the number of free, reactive amine (number of lysines + number of N-terminal amines). Immediately following addition of formaldehyde, the solution was mixed and transferred to a vial or glass NMR tube (for time series data) containing solid 5-fluoroindole or 5-fluoroindole carboxylate at a stoichiometry of 1.25-2x relative to moles of aldehyde. Reactions conducted in vials were stirred gently with a micro stir bar. Note that 5-fluoroindole and 5-fluoroindole-2-carboxylate are only slightly soluble (with the latter the more soluble of the two). Experimentation with co-solvents (0.1-1%) to aid solubility were successful but were less desirable given the adverse effects on protein stability and less controlled reaction rates, and so were not continued. The reactions were carried out at room temp (23 °C) for 2-3 days for fluoroindole or 5-7 days for fluoroindole carboxylate prior to purification. Reactions done above 40 °C were found to achieve similar levels of conversion within 0.5-1.5 days whereas below 10 °C, conversion after 0.5 days was negligible by 19F NMR. Samples were either quenched with L-arginine and sterile filtered through 0.4 micron Millipore syringe filters, or simply rapidly diluted and buffer exchanged three to four times using Amicon-15 spin concentrators (EMD Millipore). Spectra acquired as a function of pH were buffer exchanged into 20 mM K+,Na+ phosphate at pH 6.0, 7.2, or 8.0. For the time-series experiments, spectra were acquired within 20 minutes of combining reagents in the NMR tube. NMR sample volumes were between 500-700 µL with 7-10% D2O.

NMR. NMR experiments were performed on either a 600 MHz Varian INOVA with a cryogenic probe tunable to fluorine or a 600 MHz VarianS outfitted with a HCN probe tunable to fluorine. Standard 19F 90˚ pulse widths were between 8.5-11 µs at a field strength of 14.2 kHz. A pulse-field gradient stimulated echo sequence with a 300 ms mix period was used for all diffusion experiments.

Electrospray MS. ESI+ mass spectra were acquired on a Waters Micromass ZQ mass spectrometer (Waters). Each protein spectrum was acquired over a 1 minute period with scans every 0.5 seconds. 1% formic acid was added to samples shortly before injection. Maximum entropy spectra were calculated using the MassLynx software suite (Waters).

Quantum-mechanical (QM) methods. All calculations were performed using Spartan’10 (Wavefunction Inc). Structures of methanethiol (-SMe) conjugates of BTFA, TFET, and
BTFMA were first energy minimized using Spartan’s built-in molecular mechanics minimization method. Methylamine conjugates of SETFA, 5-fluoroindole, and 5-fluoroindole carboxylate were also optimized in the same way. Equilibrium conformers were then predicted via ab initio calculations using the B3LYP hybrid functional with the 6-31G(d) basis set, which has been shown to yield reasonably accurate molecular geometries (Bauschlicher and Partridge 1995). DFT calculations using B3LYP and 6-31G(d,p) were used to obtain isotropic NMR chemical shifts for each of the structures studied, under six different solvent conditions (water, ethanol, methylene chloride, toluene, acetone, and dimethyl sulfoxide). For the trifluoromethyl groups, isotropic chemical shifts for each fluorine nucleus were averaged and normalized to the average isotropic shift determined using water as the solvent. Normalized values were plotted along a solvent coordinate (Py-scale) according to increasing polarity (Dong and Winnik 1982; Dong and Winnik 1985).

4.4 RESULTS AND DISCUSSION

Indole rings are surprisingly reactive toward electrophilic species. The C3 position of the indole ring is roughly \(10^{13}\) times more reactive than benzene toward aromatic substitutions. (Bandini 2013, Gribble 2000; Bandini and Eichholzer 2009) In the presence of aqueous formaldehyde, protein side chain and N-terminal amino groups tend to form electrophilic imine intermediates. This process is identical to the first steps associated with reductive aminations/alkylations of proteins (Means and Feeney 1968, Borch et al. 1971, Rayment 1997, Larda et al. 2012). These transient cationic intermediates are sufficiently reactive toward the C3 of indole rings such that conjugation readily proceeds under non-denaturing temperatures and pH. Reactions between protein amino groups and indoles (as well as imidazoles and/or phenols) in the presence of formaldehyde was noted by Fraenkel-Conrat et al. (1947), though earlier evidence for such reactions has existed since the advent of the tanning industry (Fraenkel-Conrat et al. 1947). Conjugations between amino groups and tyrosine residues in proteins, which occur similarly via mannich reactions (Betti reaction), have been reported but tend to require far greater stoichiometric ratios of reagent:protein and are prone to side reactions with tryptophan residues (Joshi et al. 2004, Romanini and Francis 2008, McFarland et al. 2008). Here, we demonstrate site-specific covalent attachment of fluoroindoles to the lysine side chain and N-terminal amino groups of hen egg white lysozyme (HEWL). This bioconjugation strategy is
achievable at low stoichiometric ratios of reagent to protein, thereby eliminating the common concerns associated with crosslinking at high protein concentrations in the presence of formaldehyde. We show that the reaction is highly specific for lysine side chain and N-terminal amino groups. The reaction can be performed at neutral, basic or acidic (pH 5-9) ranges in buffered or unbuffered solution and seems to be very selective for only lysine side chain and N-terminal amines. The only requirement for site-specific protein labeling is the absence of primary, or secondary amines, reducing agents, and any CH acidic compounds as buffer components during the reaction (as the presence of such buffer components can introduce cross-reactivity as well as reactions with tyrosine and imidazole moieties). Following reaction, the labeled protein can be exchanged into any buffer desired, however.

By employing 5-fluoroindole or 5-fluoroindole-2-carboxylate, fluorine probes can be conjugated to each reactive lysine on the protein as shown below (Figure 4.1A, 4.1B, respectively).
For both fluoroindole and fluoroindole-2-carboxylate labeled HEWL, there are an equivalent number of resolvable $^{19}$F resonances in the spectra as there are number of lysine residues + N-terminal amino groups in the protein, indicating a single label per residue (Figure 4.1C, 4.1D, Figure 4.2).

In order to definitively identify those resonances arising from fluoroindole and fluoroindole-2-carboxylate tags conjugated to protein, we performed diffusion-edited NMR to selectively eliminate all peaks arising from small molecule byproducts and reagent (Figure S4.1). Furthermore, to demonstrate that the fluoroindolation methodology is site-specific and that only lysine side chain and N-terminal amino groups are labeled, the reaction was tested against a fully-dimethylated lysozyme sample. A fully dimethylated protein has all reactive amine groups “capped” and cannot further react with formaldehyde to form the electrophilic imine intermediate that precedes indolation. As expected, the reaction does not proceed with the dimethylated protein indicating that the reaction is indeed specific for lysine sidechain and N-terminal amino groups.
terminal amino groups (Figure S4.2A, S4.2B). Arginine does not react under most conditions
given the elevated pKas of the guanidinium nitrogens and resonance. Additional NMR and mass
spectrometry-based control experiments, where small molecule amines were reacted with
formaldehyde and fluoroindoles, demonstrated that indoles indeed couple efficiently to primary
and secondary amines only (Figures S3 and S4). Chemical shift predictions and HSQC-TOCSY
experiments were also performed to confirm coupling between the $^{13}$CH$_2$ methylene “bridge”
(red in Figure 4.1) and the aromatic protons of the indole (not shown).

**Mass spectrometry**

ESI-MS mass spectra were acquired for 5-fluoroindole and 5-fluoroindole-2-carboxylate
labeled lysozyme samples equivalent to those shown in Figure 4.1 (Figure 4.3). In each case, the
$^{19}$F NMR spectra reveal 6-7 resonances, while the mass spectra demonstrate that there are on
average 2-3 indole tags incorporated per molecule. Therefore, whereas each molecule of protein
is labeled with a small number of indole tags, on average there is complete coverage/labeling of
all amino groups to some extent, with the more reactive lysines presumably exhibiting higher
degree of modification (Table 4.1).

![Figure 4.3 ESI+ Mass spectra for 5-Fluoroindole- and 5-Fluoroindolecarboxylate-labeled HEWL. A) (Left) ES+ maximum entropy calculated mass spectrum for 5-Fluoroindole-labeled egg white lysozyme. Inlay depicts original m/z spectrum used to product the max entropy spectrum. B) (Right) ES+ maximum entropy spectrum for 5-Fluoroindolecarboxylate-labeled HEWL. Each successive mass peak corresponds to the addition of a mass equivalent of either 5-fluoroindole + CH$_2$ or 5-fluoroindolecarboxylate+CH$_2$, where the CH$_2$ is from the formaldehyde carbon during imine formation. Inlays are original mass spectra used to generate the max entropy spectra. Figure S67 shows the maximum entropy mass spectrum for unlabeled HEWL with MW=14306 Da.]
Table 4.1 Percent Modification of Proteins as quantified by ESI-MS.

<table>
<thead>
<tr>
<th></th>
<th>0 mod</th>
<th>1 mod</th>
<th>2 mod</th>
<th>3 mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-fluoroindole</td>
<td>56.5</td>
<td>34</td>
<td>8.5</td>
<td>~1</td>
</tr>
<tr>
<td>HEWL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-fluoroindole-2-</td>
<td>48</td>
<td>36</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>carboxylate HEWL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*note that higher yields can be achieved but under the conditions employed (high concentration) for NMR, precipitation occurs at higher levels of protein modification. In addition, 1-2 modifications per molecule are most appropriate under the adoption of fractional labeling practices.

While fractional (partial) labeling of proteins is not an uncommon strategy, it introduces concerns about sample homogeneity and “mixtures of states”. Yet many site-specific chemical tags used for protein labeling take advantage of this approach (Joshi et al. 2004, Romanini and Francis 2008). Prior work has demonstrated the advantages of fractional labeling proteins to prevent perturbation of protein structure (Kitevski-Leblanc et al. 2013). Given the modest number of probes conjugated to each protein (Figure 4.3), prolonged sample stability at elevated temperatures, and the lack of inhomogeneity in the NMR spectra of fluoroindole-2-carboxylate labeled HEWL (Figures S4.5, S4.6), it seems that the partial labeling approach does not induce multiple protein conformational states for this tag. Unlike fluoroindole-2-carboxylate spectra, those for fluoroindole do exhibit inhomogeneity, but only at elevated temperature. At low temperature, the inhomogeneity is ameliorated, suggesting that rather than a multitude of protein states, it is likely the tag itself that exhibits conformational sampling. This is perhaps not surprising given the greater hydrophobicity of fluoroindole over fluoroindole-2-carboxylate, where favourable interactions with solvent would be expected to reduce non-specific intra-protein interactions.

**Probing protein topology by monitoring differential lysine reactivity.**

The pKas for lysine side chain amino groups are typically between 9.3-10.3 and are dependent on factors such as solvent exposure, intra- and intermolecular interactions (H-bonds, salt bridges), and steric factors (Gao et al. 2006, Gerken et al. 1982, Imoto et al. 1972, Jenoft et al. 1981, Brown et al. 1976). It is therefore reasonable to assume that it is possible to distinguish
individual lysine residues simply by their differences in reactivity toward a chemical reagent. Furthermore, any changes in their reactivity profile from the “normal” condition may indicate conformational changes in a protein and/or ligand binding events. This principle was clearly demonstrated in a 2011 study on ligand-induced changes in the β2-adrenergic receptor using an acetylating agent as the probe for lysine reactivity (Kahsai et al. 2011). From Figure 4.1C and 4.1D, the difference in peak integrals for the $^{19}$F resonances of fluoroindole and fluoroindole-2-carboxylate labeled lysines suggests variances in their relative reactivities. To further demonstrate such differences, $^{19}$F NMR spectra were acquired as a function of time for the reaction between 5-fluoroindole, formaldehyde, and lysozyme (Figure 4.4). The same was done for the reaction with 5-fluoroindole carboxylate (Figure 4.5). It is apparent that many of the lysine residues can readily be distinguished by their distinct reaction profiles. Relative differences in reactivity between the two reagents are also very pronounced, with fluoroindole carboxylate exhibiting a much slower reaction profile than fluoroindole. This is most likely due to the deactivation of the C3 position by the carboxylate group at C2 for fluoroindole-2-carboxylate.

Figure 4.4 5-fluoroindole labeling of HEWL versus time. (A) Plot of absolute peak integral versus time for Fluoroindole labeling of 500 µM HEWL in 20 mM phosphate pH 7.2 at 25 ºC with 3X stoichiometric excess of formaldehyde relative to the number of reactive primary amine and 1.25 stoichiometric excess of solid 5-fluoroindole over moles of formaldehyde. (B) Overlay of spectra as reaction time progresses. All peaks shown correspond to labeled protein and are numbered according to the profiles in plot A. The inlay is a deconvolution of the $^{19}$F NMR spectra acquired after 1030 minutes of reaction time. A larger depiction of a similar deconvolution (though of fluoroindole carboxylate) is shown in the subsequent figure. Presumably peak 5 is the N-terminal amine as determined by guanidine unfolding of the labeled protein showing increased intensity at peak 5 as well as mass spectrometry data of chymotrypsin digests that suggest this site is more readily labeled than others. Also, NMR of chymotrypsin digests of the labeled protein show increasing signal in the vicinity of peak 5.
Most lysine residues in proteins are surface-exposed and might therefore be expected to have very similar chemical shifts for side chain protons. Table S4.1 shows that for HEWL, all lysines are more than 40% exposed over the total residue accessible surface areas (ASA). Indeed, $^1$H chemical shift dispersions for lysine side chain protons ($\text{C}_\varepsilon$ methylene, $\text{N}_\zeta$ amino, $\text{CH}_3$ for methyllysines) are generally only between 0.5-0.6 ppm (Gao et al. 2006, Larda et al. 2012, Zandarashvili et al. 2013). Therefore most studies of protein lysine resonances by NMR employ detection of $^{13}$C or $^{15}$N (either directly or indirectly) to afford better resolution of otherwise overlapping resonances. With regard to fluoroindoles, fluorine nuclei are known to be highly sensitive to changes in chemical environment and the inductive effects from adjacent conjugated

* Note that resonance 8 corresponds to a small molecule in solution. Its intensity was static and it was present since the start of the reaction.

**Figure 4.5** 5-fluoroindole-2-carboxylate labeling of HEWL versus time. (A) Plot of peak integral obtained from deconvoluted spectra vs time for Fluoroindolecarboxylate labeling of 500 µM HEWL in 20 mM phosphate pH 6.0 using the same stoichiometry in the preceding figure. (B) Representative deconvolution of the $^{19}$F NMR spectra for Fluoroindolecarboxylate-labeled HEWL with peak numbers corresponding to each of the profiles in plot A. The inlay corresponds to peak integrals for non-deconvoluted spectra showing linear reaction progression across the entire interval monitored. The deconvolutions used to obtain plot A could only be performed with sufficient peak intensity, hence why the intensity starts above 200. Note that the reaction had not yet plateaued even after 1030 minutes indicating that reactions with fluoroindole carboxylate are far slower than for fluoroindole at the same temperature. The deconvolution was necessitated by the fact that peaks 4 and 5 could not be independently integrated which would otherwise prevent reliable estimation of the rate of reaction at either site.

**Fluoroindole Probe Sensitivity for NMR-based applications**

Most lysine residues in proteins are surface-exposed and might therefore be expected to have very similar chemical shifts for side chain protons. Table S4.1 shows that for HEWL, all lysines are more than 40% exposed over the total residue accessible surface areas (ASA). Indeed, $^1$H chemical shift dispersions for lysine side chain protons ($\text{C}_\varepsilon$ methylene, $\text{N}_\zeta$ amino, $\text{CH}_3$ for methyllysines) are generally only between 0.5-0.6 ppm (Gao et al. 2006, Larda et al. 2012, Zandarashvili et al. 2013). Therefore most studies of protein lysine resonances by NMR employ detection of $^{13}$C or $^{15}$N (either directly or indirectly) to afford better resolution of otherwise overlapping resonances. With regard to fluoroindoles, fluorine nuclei are known to be highly sensitive to changes in chemical environment and the inductive effects from adjacent conjugated
bonds makes fluoroaromatics particularly sensitive (Gerig 1997, Gerig 2001, Vulpetti and Dalvit 2012, Dalvit et al. 2014, Kitevski-LeBlanc and Prosser 2012). $^{19}$F NMR spectra of fluoroindole and fluoroindole-2-carboxylate labeled HEWL exhibit chemical shift dispersions four times greater that afforded by $^1$H. Figures S4.5 and S4.6 show pH and temperature-dependence of $^{19}$F chemical shifts, which further demonstrate sensitivity to relatively small changes in local environment. These results are consistent with the fact that fluoroindole probes are able to sensitively register subtle differences in local electrostatic, steric, and dipolar environments.

Many existing chemical reagents for the site-specific incorporation of NMR-active labels into proteins target cysteine residues. Few $^{19}$F labels exist for lysine labeling, and the majority of these involve acylations that eliminate the charge on the amine. In the design of new $^{19}$F labels, it would be useful to be able to predict how well one tag performs over another in terms of sensitivity to local environmental changes. Here, we employed DFT calculations to calculate isotropic chemical shifts for fluorine nuclei in thio- or amino-conjugates of several known $^{19}$F-NMR tagging reagents (Figure 4.6) across varying solvent conditions (Figure 4.7).

![Chemical Structures](image)

**Figure 4.6** Structures of six fluorine tags used for $^{19}$F-labeling of cysteine (BTFA, TFET, BTFMA) and lysine (SETFA, FINDOLE, FINDCBXY) residues. Note that DFT calculations were performed on methanethiol (CH$_3$S$^-$) conjugates of the cysteine tags (where the thiolate replaces the bromine), whereas methylamine (CH$_3$NH$_2$) was conjugated to SETFA and trimethylmethanammonium ([(CH$_3$)$_3$NH]$_2$) was conjugated to both 5-fluoroindole and 5-fluoroindole-2-carboxylate at the C3 position. These conjugates are small molecule analogs of structures that would otherwise form following labeling of the appropriate amino acid residue.
3-bromo-1,1,1-trifluoropropan-2-one (BTFA), 2,2,2-trifluoroethanethiol (TFET), and 2-bromo-N-(4-(trifluoromethyl)phenyl)acetamide (BTFMA) are all $^{19}$F cysteine tags whereas S-Ethyl trifluoroethanethiolate (SETFA) is a known $^{19}$F tag used for lysine labeling (Kitevski-LeBlanc and Prosser 2012, Adriaensens et al. 1988). From prior work (unpublished), A methane-thiol conjugate of BTFMA was shown to exhibit the highest magnitude of changes in chemical shift across all solvent conditions among the three cysteine probes (Ye, Larda, Prosser). Conjugates of BTFA and SETFA exhibit comparable $^{19}$F chemical shift sensitivities but are the least sensitive of the tags investigated. The same calculations performed for 5-fluoroindole and 5-fluoroindole-2-carboxylate conjugates demonstrates that these probes are the most sensitive of all the $^{19}$F tagging reagents (Figure 4.7). This is perhaps not surprising given that fluoroaromatics are more sensitive to inductive effects due to the conjugated, delocalized pi electrons and considering the position of the fluorine nucleus on the indole ring. The fluorine atom, being directly bonded to the ring, is likely more susceptible to ring-current and inductive effects than for a trifluoromethyl group as in the case of BTFMA. This further lends support for the utility of fluoroindolation of protein lysine residues toward monitoring structural changes in proteins and functional binding events.

![Figure 4.7 DFT-based normalized $^{19}$F chemical shifts for each fluorine tag across varying solvent polarities. Shifts are normalized to the chemical shift of the tag in H$_2$O. Isotropic NMR shifts for each of the tags were calculated using the B3LYP hybrid functional and the 6-31G(d,p) basis set, at six different solvent conditions [left to right: toluene (Py = 1.04), ethanol (Py = 1.18), methylene chloride (Py = 1.35), acetone (Py = 1.64), water (Py = 1.87), and dimethyl sulfoxide (Py = 1.95)]. (Dong and Winnik 1982; Dong and Winnik 1984)](image)
4.5 CONCLUSIONS

We describe a chemical tagging method for the site-specific labeling of lysine residues. The reaction takes advantage of the highly nucleophilic C3 position of indole rings and can therefore be used to conjugate various indole analogs onto proteins site-specifically under mild, non-denaturing, aqueous conditions. We have shown that the fluoroindole is a useful probe for fluorine NMR and mass spectrometry whereas we also foresee applications toward protein pegylation, crosslinking, immobilization, ICAT tagging (using indole-H$_7$ and indole-d$_7$), $^{18}$F-PET imaging, as well as UV-Vis and intrinsic fluorescence. With regard to fluoroindole probes; they are highly sensitive to local chemical environment, more so than most commonly used fluorine tags for NMR as shown by DFT and offer the advantage that they allow retention of the native protein charge. The indolation reaction is rapid and may be conducted under very mild conditions and low stoichiometry and is therefore a useful addition to the toolkit of bioconjugation reactions.

4.6 SUPPORTING INFO

Table S4.1 Relative % solvent exposure for lysine residues in hen egg white lysozyme. Calculated from structure 193L.

<table>
<thead>
<tr>
<th>Lysine</th>
<th>molmol$^a$</th>
<th>getarea$^b$</th>
<th>surfaceracer$^c$</th>
<th>DSSP$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79</td>
<td>60</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>75</td>
<td>74</td>
<td>76</td>
<td>73</td>
</tr>
<tr>
<td>33</td>
<td>56</td>
<td>57</td>
<td>65</td>
<td>59</td>
</tr>
<tr>
<td>96</td>
<td>41</td>
<td>42</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>97</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>116</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>


Figure S4.1 Diffusion-edited $^{19}$F NMR for 5-fluoroindole and 5-fluoroindole-2-carboxylate labeled HEWL at 35 °C. Sample was prepared as described previously and buffer exchanged 5x to milliQ H$_2$O (concentration ~ 0.5-1 mM). (Left) $^{19}$F NMR spectrum of the sample before (cyan) and after (red) application of diffusion-filter to eliminate fast diffusing species. Spectra acquired at 35 °C. (Right) The equivalent experiment for 5-fluoroindole labeled HEWL at 5 °C.

Figure S4.2 (A) $^{19}$F NMR spectra for reactions between 5-fluoroindole, formaldehyde and unmethylated HEWL (cyan) and 5-fluoroindole, formaldehyde, and fully $^{13}$C-Dimethylated HEWL (red). Both spectra are normalized to the intensity of the reagent (fluoroindole) peak. (B) $^1$H,$^{13}$C hsqc spectrum for fully $^{13}$C-dimethylated HEWL. Note that no intensity was observed in the monomethyl region (only dimethyllysine region shown) indicating full dimethylation of the protein. Complete dimethylation ensures that all lysine side chain and N-terminal amino groups are “capped” thereby preventing reaction with indoles in the presence of formaldehyde. Only small molecule byproduct is formed (S2.A) as indicated by the asterisks and no peaks occur in the region of labeled amine (greyed box region). Note that without normalizing to reagent peak height, the intensities of the small molecule species (asterisk) are far greater than those of the conjugated indole resonances. The resonances corresponding to conjugated indole persist following dialysis and molecular weight filtration, whereas the small molecule peaks are lost following extensive dialysis.
Figure S4.3 (A and B) $^{19}$F NMR comparisons between the following reactions: (Green) 3 mg Tris(hydroxymethyl)aminomethane in ~90% H$_2$O:D$_2$O with 3 mg 5-fluoroindole and 0.5 uL formaldehyde. (Blue) 3 mg BisTris in ~90% H$_2$O:D$_2$O with 3 mg 5-fluoroindole and 0.5 uL formaldehyde. (Red) 3 mg 5-fluoroindole with 0.5 uL formaldehyde in ~90% H$_2$O:D$_2$O. Note that the reaction between formaldehyde and fluoroindole alone exhibits two distinct byproducts not formed in the reactions with bis-tris or tris. Most importantly, note that only the reaction with tris exhibits a peak at -123.9 ppm indicating reaction with free amine. Mass spectra confirm the results exhibited here, that fluoroindole only couples to tris (primary amine) and not to bis-tris having a tertiary amine. Furthermore, a single peak for labeling of tris suggests that only single fluoroindole couples to a primary amino group (no double-labeling). The fluoroindole reagent peak occurs at -125.55 ppm.

Figure S4.4 Lysine hydrochloride + 5-fluoroindole + $^{12}$C formaldehyde ESI$^+$ mass spectrum. Molecular weights and structures of relevant compounds which dominate the m/z peak in the spectrum are shown.
Figure S4.5 $^{19}$F NMR spectra of fluoroindole and fluoroindole-carboxylate labeled HEWL versus Temperature. (Left column) 5-fluoroindole-labeled, and (right column) 5-Fluoroindole-carboxylate labeled egg white lysozyme in unbuffered milliQ H$_2$O at 5 °C (top), 25°C (center) and 45 °C (bottom). Concentration of fluoroindole and fluoroindole-2-carboxylate labeled HEWL was between 0.5-1 mM. Both samples were buffer exchanged into pure milliQ H$_2$O. Note that for the Fluoroindole sample, resonances are exchange broadened at high temperature whereas fluoroindole-2-carboxylate labeled HEWL spectra seem to be well resolved at all temperatures investigated (5-55 °C). Note that the resolution of resonances seems better in unbuffered solution than in buffered, salty, solution. Peak deconvolution is shown in blue with the residual error in red.
Figure S4.6 $^{19}$F NMR spectra of fluoroindole and fluoroindole-carboxylate labeled HEWL versus pH. 5-fluoroindole (left column) and 5-fluoroindole-2-carboxylate labeled (right column) HEWL $^{19}$F NMR spectra acquired at pH 6.0 (top), 7.2 (middle), and 8.0 (bottom) at 35 °C. Samples were prepared as described in Figure 1 but each sample was dialyzed into 20 mM K+/Na+ phosphate at the respective pH indicated. Note the improved resolution of resonances for fluoroindole-2-carboxylate labeled HEWL at pH 6.0 (bottom right). Note also that at all pHs investigated, the samples showed no signs of degradation over extended periods of storage at room temperature.
4.7 REFERENCES


Bauschlicher CW Jr. Partridge H (1995) The Sensitivity of B3LYP Atomization Energies to the Basis Set and a Comparison of the Basis Set Requirements for CCSD(T) and B3LYP. NASA Ames Research Center, Moffett Field, CA 94035, USA DOI: 10.1016/0009-2614(95)91855-R


CHAPTER 5
A Comparison of Chemical Shift Sensitivity of Trifluoromethyl Tags: Optimizing $^{19}\text{F}$ NMR Studies of Proteins

5.1 ABSTRACT

The elucidation of distinct protein conformers or states by fluorine ($^{19}\text{F}$) NMR requires fluorinated moieties whose chemical shifts are most sensitive to subtle changes in the local dielectric and magnetic shielding environment. In this study we evaluate the effective chemical shift dispersion of a number of thiol-specific trifluoromethyl probes [i.e. 2-bromo-$N$-(4-(trifluoromethyl)phenyl)acetamide (BTFMA), $N$-(4-bromo-3-(trifluoromethyl)phenyl)acetamide (3-BTFMA), 3-bromo-1,1,1-trifluoropropan-2-ol (BTFP), 1-bromo-3,3,4,4,4-pentafluorobutan-2-one (BPFB), 3-bromo-1,1,1-trifluoropropan-2-one (BTFA), and (2,2,2-trifluoroethyl)-1-thiol (TFET)] under conditions of varying polarity. In considering the sensitivity of the $^{19}\text{F}$ NMR chemical shift to the local environment, a series of methanol/water mixtures were prepared, ranging from relatively non-polar (MeOH:H$_2$O = 4) to polar (MeOH:H$_2$O = 0.25). $^{19}\text{F}$ NMR spectra of the tripeptide, glutathione ((2S)-2-amino-4-{{[(1R)-1-[(carboxymethyl)carbamoyl]-2-sulfanylethyl]carbamoyl}butanoic acid), conjugated to each of the above trifluoromethyl probes, revealed that the BTFMA tag exhibited a significantly greater range of chemical shift as a function of solvent polarity than did either BTFA or TFET. DFT calculations using the B3LYP hybrid functional and the 6-31G(d,p) basis set, confirmed the observed trend in chemical shift dispersion with solvent polarity.

The work presented in this chapter was recently submitted to the Journal of Biomolecular NMR as:
Libin Ye, Sacha Thierry Larda, Yi Feng Frank Li, Aashish Manglik, R. Scott Prosser. A Comparison of Chemical Shift Sensitivity of Trifluoromethyl Tags: Optimizing $^{19}\text{F}$ NMR Studies of Proteins.

My contributions to this work include all QM-based experiments/calculations and associated analyses. Libin Ye conducted the solution NMR experiments.
5.2 INTRODUCTION

While more than one hundred elements are NMR active, the fluorine nucleus has the distinction of possessing one of the highest gyromagnetic ratios, and thus, greatest sensitivity for NMR, next to tritium and $^1$H nuclei. $^{19}$F NMR also exhibits a remarkable range of chemical shift; common fluorinated organics span a chemical shift range of ~1100 ppm. For this reason fluorine NMR has had a significant impact in fragment based drug discovery and protein structure function studies (2,3) (4-8). In $^{19}$F NMR studies of proteins, fluorinated amino acid derivatives (i.e. tryptophan, phenylalanine, tyrosine, and methionine) are commonly incorporated by biosynthetic means, in which case, the chemical shift dispersion arises not from variations of chemical structure but rather from differences in local secondary and tertiary structure, solvent exposure, and conformational dynamics (2). There are examples of protein studies where $^{19}$F NMR resonances arising from fluorophenylalanine (9) and fluorotryptophan (10) span as much as 10-20 ppm, reflecting the sensitivity of the chemical shift to environment. It is this exquisite sensitivity to environment that we generally wish to exploit in many protein NMR studies, for purposes of delineating distinct conformers or states, and, through relaxation experiments, their interconversion rates and lifetimes.

In this study we compare the chemical shift sensitivity of a number of common thiol-reactive trifluoromethyl (-CF$_3$) probes whose purpose is to monitor conformers and states of proteins of interest. In each case, these probes are incorporated into the protein through 1- or 2-step reactions under non-denaturing aqueous conditions, as described elsewhere and in the Materials and Methods section. Trifluoromethyl probes are a preferred form of NMR tag for $^{19}$F NMR studies of proteins, given their modest amplification in the signal to noise ratio resulting from three equivalent nuclei and the fact that fast methyl rotation further averages the chemical shift anisotropy (CSA). Since $^{19}$F NMR relaxation is often dominated by CSA at high magnetic fields, the use of probes with relatively small CSA is key to the study of protein states, particularly in cases where the protein of interest is on the order of 100 residues or more. It is also possible, for purposes of further enhancing the signal to noise ratio, to resort to sulfhydryl specific probes, consisting of higher numbers of equivalent fluorine nuclei, such as perfluoro tert-butyl moieties (11) and even larger structures (12). In this case, the challenge becomes
avoiding perturbation to the states or conformations of interest in the protein. In many cases, this can be assessed either through functional assays or by $^{15}$N,$^1$H spectroscopy or circular dichroism, which provides a global perspective of the state of the protein before and after labeling.

In situations employing a cysteine-specific trifluoromethyl $^{19}$F NMR probe, we address the question “Are some CF$_3$ probes better than others?” In the context of defining states, this question amounts to an issue of chemical shift sensitivity, which may loosely be defined as the maximum possible range of chemical shifts accessible to the tag, as a function of environment, divided by the line width. Thus, a small CSA term, a long transverse relaxation time ($T_2$), and a sensitivity of chemical shift to environment, are required.

Fluorine chemical shifts are known to be strongly influenced by long-range electrostatic effects (13) where solvent accessibility, van der Waals interactions, and local magnetic effects from aromatics (ring currents) and electronically anisotropic species all contribute to difference in shielding at the fluorine nucleus. In particular, electrostatic effects and van der Waals interactions couple with the fluorine lone pair electrons resulting in significant modulation of the paramagnetic shielding term (14). For example, an adjacent carbonyl moiety typically exerts strong shielding effects on the fluorine nuclei of the trifluoromethyl group (15). Other moieties may contribute to the effective partial charge of the CF$_3$ group, thereby allowing greater solvent clustering and thus, deshielding effects. One might further speculate that the polarizability and molecular geometry of species adjacent to the CF$_3$ group could in turn influence the range of shielding effects and thus the sensitivity of the CF$_3$ moiety to environment. A second possibility is that the nearby atoms, constituting the trifluoromethyl probe physically shield the CF$_3$ group from solvents or the local environment, thereby reducing the overall sensitivity of the probe to environment.

In the current study, we compare the sensitivity of the observed $^{19}$F NMR chemical shift to polarity for a variety of CF$_3$ probes. Polarity is controlled by varying the ratio of methanol to water in mixtures. Significant differences in the chemical shift dispersion are observed. To ascertain the possible origin of these differences we employed Density Functional Theory (DFT) calculations to corroborate the dependence of the chemical shift on polarity while estimating the partial charge on the fluorine atoms and the carbon atom of the CF$_3$ tag, in the hopes of identifying a correlation.
5.3 MATERIALS AND METHODS

Reduced glutathione (GSH) was conjugated with the following tags as described below:

1) BTFA. GSH was dissolved in 5 mL phosphate buffered saline (PBS) buffer, pH 7.4, to a final concentration of 10 mM, followed by the addition of BTFA to a final concentration of 1 mM. The mixture was heated to 65°C for 5 min and the reaction was allowed to continue overnight at room temperature.

2) BPFB, BTFP, 3-BTFMA and BTFMA. GSH was dissolved in 5 mL PBS, pH 7.4, to a final concentration of 10 mM followed by the addition of the tag to a final concentration of 1 mM. The reaction mixture was then incubated at room temperature overnight without heating.

3) TFET. The reaction was completed in two steps. 2,2'-dipyridyl disulfide (DPS) was added to 10 mM GSH (pH 7.4 in PBS buffer) in a final concentration of 5 mM, and the mixture was left at room temperature for 6 hours in the dark. Subsequently, the TFET was added into solution with a final concentration of 1 mM, and the reaction was allowed to continue overnight at room temperature, thereby replacing the DPS moiety by TFET.

NMR

A series of MeOH/water mixtures were prepared, ranging from MeOH/water (v/v) 4:1 through 1:4. In all cases the labeled peptide was observed to be soluble. NMR experiments were performed on a 600 MHz ($^1$H Larmor frequency) Varian Inova spectrometer equipped with a cryogenic probe capable of $^{19}$F NMR spectroscopy, and the chemical shift sensitivity measurements were performed at 35°C. T$_2$ relaxation times were measured using a standard Hahn echo experiment, using a repetition time of 2 s, and 12-15 echo delay times. Spectra were processed with MestReNova 9.0.1 software, and $^{19}$F NMR peaks were referenced to NaF (119.25 ppm).

LC-MS

LC-MS was used to monitor the extent of conjugation. The sample was diluted into 50% methanol solution with final concentration of 100 µM supplemented with 0.1% TFA, and centrifuged at 5000g for 10 min to remove possible precipitated solute/contaminate prior to LC-
MS. The flow rate was set to 40 µL/min, and the spectra were acquired with 30 scans with a range of M/Z between 100 and 900 to confirm the conjugation between $^{19}$F tag and GSH.

**Quantum-mechanical (QM) methods**

All calculations were performed using Spartan’10 (21,22). Structures of methanethiol (-SMe) conjugates of BTFA, TFET, and BTFMA were first energy minimized using Spartan’s built-in molecular mechanics minimization method (23). Equilibrium conformers were then predicted via *ab initio* calculations using the B3LYP hybrid functional with the 6-31G(d) basis set, which has been shown to yield reasonably accurate molecular geometries (23). DFT calculations using B3LYP and 6-31G(d,p) were used to obtain isotropic NMR chemical shifts and atomic partial charges (electrostatic) for each of the structures studied, under six different solvent conditions (water, ethanol, methylene chloride, toluene, acetone, and dimethyl sulfoxide). For the trifluoromethyl groups, isotropic chemical shifts for each fluorine nucleus were averaged and normalized to the average isotropic shift determined using water as the solvent. The same was done for $^{19}$F electrostatic partial charges. Normalized values were plotted along a solvent coordinate (Py-scale) according to increasing polarity (16,17).

5.4 RESULTS AND DISCUSSION

To systematically evaluate spectroscopic features, T$_2$ relaxation, and $^{19}$F NMR chemical shift sensitivity to environment, we consider herein a variety of sulfhydryl reactive trifluoromethyl probes, shown in Figure 5.1. BTFA and TFET represent two commonly used $^{19}$F NMR probes that are expected to minimally perturb the target protein, due to their relatively small size. With the exception of TFET, all of the probes may be conjugated to free thiols in a single reaction step. Conjugation of TFET to proteins, generally involves an additional step where the free thiol(s) are first reacted with dithiopyridine, a relatively bulky moiety, which would be expected to preferentially react only with the most exposed cysteine residue(s), possibly avoiding attack of largely-buried cysteines or disulfide linkages. One disadvantage of TFET is that the resulting disulfide linkage is somewhat vulnerable to reduction in the presence of reducing agents. BTFA is an equally popular trifluoromethyl tag and may have the advantage that the S-C bond established following reaction with thiols and thiolates is stable. Note that the trifluoromethyl moiety in BTFA possesses no scalar coupled protons, thus reducing possible line
broadening through dipolar relaxation and obviating the need for $^1$H decoupling.

A variety of other commercially available trifluoromethyl tags, depicted in Figure 5.1, are expected to provide a range of effects on the fluorine valence electron density, and thus sensitivity to chemical shift. In particular, note that in the case of BTFMA and 3-BTFMA, the trifluoromethyl substituent is directly conjugated to the phenyl group. The delocalized electron density of the aromatic system is highly polarizable and thus, sensitive to environment. The polarizability of the phenyl moiety adjacent to the CF$_3$ group would therefore be expected to contribute to a greater range of shielding effects on the fluorine nuclei, depending on environment. Furthermore, ring current effects are expected to contribute additional shielding/deshielding effects to the $^{19}$F nuclei of the trifluoromethyl group. BTFMA and 3-BTFMA are comparable in size to current nitroxide spin-labels used in ESR applications. This might be advantageous in a study involving both NMR and ESR since possible steric perturbations would at least be similar in both cases.

![Chemical structures](image)

**Figure 5.1** Structures of thiol-reactive trifluoromethyl probes used in this study.

To examine the potential chemical shift sensitivity of the probes, shown in Figure 5.1, all were first conjugated to a small test peptide, glutathione, or ((2S)-2-amino-4-[[((1R)-1-[(carboxymethyl)carbamoyl]-2-sulfanylethyl]carbamoyl]butanoic acid) via the cysteine sulfhydryl. $^{19}$F NMR spectra of the $^{19}$F-labeled peptides were then separately recorded in
solvents of varying polarity. Since our interest is in studying proteins by $^{19}$F NMR under aqueous conditions, we elected to make use of water combined with methanol over a variety of concentrations such that the water to MeOH ratio ranged from 0.25 to 4.0. MeOH has a dielectric constant of 32.7 and that of water is 80.1, while both solvents are completely miscible. Thus, mixtures of the two solvents provide a range of polar environments while still solubilizing the peptide. While there are many definitions of polarity, we make use of the well-known Py scale for polarity, which is based on the vibronic fine structure of pyrene fluorescence (16,17). Pure water exhibits a Py polarity index of 1.87, while the corresponding value for MeOH is 1.35. Moreover, the Py polarity values for MeOH/water mixtures have been previously determined and were therefore used in the current study (18).

Figure 5.2 reveals the change in the observed $^{19}$F NMR chemical shifts as a function of polarity. Note that the chemical shift changes are defined with respect to the observed shifts for the least polar solvent mixture (i.e. MeOH/H$_2$O = 4, Py = 1.454). While all of the trifluoromethyl probes reveal a sensitivity of chemical shift to polarity, both probes where there is a CF$_3$ group directly conjugated to an aromatic, exhibit the greatest chemical shift range by a significant margin over BTFA. Observed changes in chemical shift were fit to the formula $\Delta \delta = A/(1+\exp(-(Py-x_0)/w))$, where $A$, $x_0$, and $w$ represent fitting parameters, and are presented in Supplemental Table 1, for each probe. Note in particular, that $A$ represents the magnitude of the chemical shift change, while $1/w$ is a measure of steepness of the observed profile. BTFMA clearly exhibits the greatest extent of change in chemical shift ($A$) over the range of polarities investigated.
Figure 5.2 Changes in $^{19}$F chemical shift ($\Delta \delta$) of various CF$_3$ tags as a function of solvent polarity. Note that all changes are referenced to the most hydrophobic solution investigated (i.e. MeOH/H$_2$O = 4.0 or Py=1.454). Observed changes in chemical shift were fit to the formula $\Delta \delta = A/(1+\exp(-(|Py-x_0|)/w))$, where $A$, $x_0$, and $w$ represent fitting parameters, for each probe.

To understand the origin of our experimental results, Density Functional Theory (DFT) calculations of the average $^{19}$F NMR chemical shift were performed on three trifluoromethyl tags, anticipated to be most useful for labeling cysteine residues in NMR studies - namely, BTFA, TFET and BTFMA. To facilitate the DFT calculations, the above probes were each assumed to be conjugated to methanethiol, rather than glutathione, as described in the Materials and Methods section. Using the B3LYP hybrid functional and the 6-31G(d,p) basis set, isotropic $^{19}$F NMR chemical shifts were estimated for each of the three tags, both in the vapor phase (Py = 0.4) and in solvents of varying polarity. These chemical shifts were then normalized to the predicted shifts in water, as shown in Figure 5.3A. There are two clear trends that are apparent
from the simulations in Figure 5.3 and are most evident in the case of BTFMA. Firstly, the normalized shifts appear to increase with polarity, indicating larger chemical shift differences between the three tags at lower solvent polarity. Secondly, protic solvents (namely ethanol, for which Py = 1.18, and water, for which Py = 1.87) exhibit very pronounced maxima. As is clear from Figure 3B, there also appears to be a direct correspondence between the partial charge on the $^{19}$F atom and the observed shift. This correlation is further illustrated in Supplementary Figure S2. Partial charge is expected to be a consequence of solvent polarity and the chemical structure of the probe itself. As evidenced by the DFT calculations, the $^{19}$F atoms on the CF$_3$ moiety associated with BTFMA, exhibit the greatest range of partial charges as a function of polarity (solvent) and correspondingly, the greatest range of $^{19}$F NMR chemical shifts.

We emphasize that there are many choices of polarity and the Py scale, used in this paper is one of many. Supplementary Figure S5.1 shows normalized shifts for the same solvents introduced in Figure 3, though using a different polarity scale. In this case, the trend in shift with polarity appears more monotonic. Our goal in this paper was simply to determine differences or sensitivities in shift to changes in polarity or solvent and the exact shape of the curve is outside the scope of this paper. It is also important to recognize that the calculated chemical shift and partial charge profiles for BTFMA, BTFA, and TFET, shown in Figure 5.3, explore a wide range of solvents and a correspondingly wide range of polarities. In contrast, the experimental results shown in Figure 5.2, which depict chemical shift sensitivity to polarity for a series of probes, were determined only in methanol/water mixtures over a more limited range of polarities. Moreover, between Py values of 1.454 and 1.78, the experimental and DFT-based results are consistent. The main point is that the computational studies corroborate the observation that BTFMA exhibits a greater sensitivity of $^{19}$F NMR chemical shift to polarity, suggesting it may serve as an effective NMR probe in protein NMR studies.
The above studies were in part motivated by observations from prior and current $^{19}$F NMR studies of detergent-stabilized membrane receptor, $\beta_2$AR, in which a specific cysteine moiety (Cys 265) was tagged by BTFA (19), TFET (20), or BTFMA. In our hands we observed dramatically improved chemical shift dispersion upon resorting to BTFMA. While this is at first surprising that the CF$_3$ group, conjugated to the aromatic (i.e. BTFMA) should exhibit dramatically improved chemical shift sensitivity over more conventional probes, the model

**Figure 5.3** (A) DFT-based normalized chemical shifts for three CF$_3$ tags conjugated to methanethiol (shown in C). Isotropic NMR shifts for each of the tags were calculated using the B3LYP hybrid functional and the 6-31G(d,p) basis sets, under six different solvent conditions [left to right: toluene (Py = 1.04), ethanol (Py = 1.18), methylene chloride (Py = 1.35), acetone (Py = 1.64), water (Py = 1.87), and dimethyl sulfoxide (Py = 1.95)]. (B) DFT-based normalized average partial charges for $^{19}$F atoms for each of the above three CF$_3$ tags, as a function of solvent polarity. The partial charges for each of the $^{19}$F atoms in the CF$_3$ group were averaged assuming isotropic conditions and fast methyl rotations and then normalized to the average $^{19}$F partial charge for the conjugate in H$_2$O. (C) Structure of the three methanethiol conjugates for which DFT calculations were performed.

### 5.5 CONCLUSIONS AND FINAL REMARKS

The above studies were in part motivated by observations from prior and current $^{19}$F NMR studies of detergent-stabilized membrane receptor, $\beta_2$AR, in which a specific cysteine moiety (Cys 265) was tagged by BTFA (19), TFET (20), or BTFMA. In our hands we observed dramatically improved chemical shift dispersion upon resorting to BTFMA. While this is at first surprising that the CF$_3$ group, conjugated to the aromatic (i.e. BTFMA) should exhibit dramatically improved chemical shift sensitivity over more conventional probes, the model
studies and DFT studies in this paper confirm these observations. Moreover, the correspondence between observed chemical shift measurements in MEOH/water and the DFT simulations suggest that much insight into the potential of next generation fluorinated tags could be obtained by computational and experimental studies such as those outlined above.

Acknowledgments
We would like to thank the Natural Sciences and Engineering Research Council of Canada for support of this research.

5.6 SUPPLEMENTARY INFORMATION

Table 5.1 Fitting parameters (A, x_0 and w) used to interpret the dependence of chemical shift change Δδ of CF3 tags to solvent polarity, using the formula Δδ = A/(1+exp(-(Py-x_0)/w))

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>x_0</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTFA</td>
<td>0.373524</td>
<td>1.58831</td>
<td>0.043651</td>
</tr>
<tr>
<td>BPFB</td>
<td>0.520183</td>
<td>1.5827</td>
<td>0.060177</td>
</tr>
<tr>
<td>BTFP</td>
<td>0.588046</td>
<td>1.60626</td>
<td>0.048339</td>
</tr>
<tr>
<td>BTFMA*</td>
<td>0.679928</td>
<td>1.63681</td>
<td>0.054611</td>
</tr>
<tr>
<td>TFET</td>
<td>0.538624</td>
<td>1.60405</td>
<td>0.045583</td>
</tr>
<tr>
<td>BTFMA</td>
<td>0.705462</td>
<td>1.61827</td>
<td>0.052505</td>
</tr>
</tbody>
</table>
**Supplementary Figures**

**Figure S5.1** $^{19}$F NMR chemical shifts (normalized to $^{19}$F shifts in H$_2$O) as a function of solvent polarity. This plot uses a relative solvent polarity scale where H$_2$O is arbitrarily assigned a polarity = 1. This relative polarity scale is defined by chromophore absorbance in various solvents (24).

**Figure S5.2** Correlation between normalized $^{19}$F chemical shifts and normalized $^{19}$F electrostatic partial charges. The slope of each plot indicates the extent of $^{19}$F shift changes for a given change in partial charge on the fluorine. BTFMA exhibits the greatest chemical shift change for a given change in partial charge. Each point corresponds to a chemical shift and partial electrostatic charge for a given solvent (or solvent polarity).
REFERENCES


CHAPTER 6

Saturation Transfer NMR: Aliphatic Cross Saturation Spin Diffusion

This section introduces the NMR cross-saturation effect and the aliphatic-saturation spin-diffusion experiment for the detection of subtle conformational changes (or changes in stability) in proteins that are otherwise undetectable by standard $^{15}\text{N}, ^1\text{H}$ NMR.

6.1 Cross-Saturation (CS) and Spin Diffusion:

In the study of proteins and other biopolymers by NMR, it was readily discovered that the spin-lattice relaxation times (T1s) among proton spins were more homogeneous than would otherwise be expected (Kalk and Berendsen 1976, Akasaka 1981, Russu and Ho 1982). This phenomenon was attributed to an NOE (cross-relaxation) effect known as “spin diffusion” whereby nuclear spins exchange magnetization via through-space dipole-dipole interactions. Spin diffusion is a prominent contributor to spin-lattice relaxation and given that it propagates as a cross-relaxation effect, it tends to lead to an equalization of the T1s of the proton pool. It is therefore often a problem for the determination of intramolecular distances via NOESY NMR at long mixing times because as the mixing time increases, the probability of indirect vs direct cross-relaxation transfers between nearby spins also increases. Therefore, rather than an NOE occurring between two spins (A and C) directly, the NOE may be transferred from (A) to an intermediate spin (B), then to (C). This confuses distance measurements and thus mixing times for NOESY experiments are usually reduced in order to avoid such spin diffusion effects. Methods have also been devised to reduce losses in signal intensity due to spin-diffusion related T1 rate enhancements. Protein deuteration is one such technique, which reduces the total number of proton spins (proton pool) and therefore the extent of cross-relaxation mediated magnetization transfers to the lattice (Bennett and Venters 1996).

Whereas spin-diffusion cross-relaxation effects can complicate distance measurements, they can be useful in the study of both inter- and intramolecular processes. For example, cross-relaxation phenomena are commonly employed in NMR to elucidate networks of spins that are
“spin-diffusion-coupled”. One typical method that takes advantage of cross-relaxation between two coupled spins (A and B) is to selectively saturate one of the spins (A) for a given length of time whilst monitoring site B. If A and B are spin-diffusion coupled, then it can be expected that saturation of A will result in partial or full saturation of B. This technique is commonly referred to as saturation transfer (though more appropriately cross-saturation as it can otherwise be confused with CEST – chemical exchange saturation transfer) and sees regular use in saturation transfer difference experiments (STD). In STD NMR, protein resonances at one frequency (1) are selectively irradiated with rf in one experiment, whereas a region of the spectrum devoid of resonances (2) is saturated in a second experiment. Taking the difference between the two spectra (2-1) reveals only those resonances connected by a spin-diffusion network to the nuclear spins that were saturated in (1). This method is frequently used to screen ligands for binding to a protein of interest (Mayer and Meyer 1999, Mayer and Meyer 2001, Streiff et al. 2004, Dalvit et al. 2000). In this case, the protein is once again selectively irradiated with rf, but this time it is in solution with one or more potential ligands. The difference experiment, in this case, shows only those resonances corresponding to ligands connected via a spin diffusion network to the saturated protein spins, thereby identifying ligands that bind to the protein (Figure 6.1).

**Figure 6.1 Saturation Spin Diffusion Experiment.** (A) Scheme at left illustrates saturation of a receptor-ligand complex. Saturation of the receptor is transferred to the ligand protons that are in contact with the receptor via spin diffusion. (B) Saturation transfer difference (STD) experiment with a T1rho filter. The T1rho filter eliminates all broad resonances arising from protein whereas the STD experiment shows only those resonances arising from ligands bound to the receptor which was originally saturated (panel F). Reprinted with permission from J. Am. Chem. Soc., (2001) 123 (25), pp 6108–6117. Copyright © American Chemical Society
Another application of the cross-saturation effect is to monitor spin-diffusion networks within a protein. Selective saturation of aliphatic or aromatic resonances in a protein can be used to probe spin-diffusion networks within the hydrophobic core. Such networks can give insight into changes in protein structure if monitored alongside a chemical (ligand or denaturant), thermal, or pressure titration (Larda et al. 2013). These techniques have also been applied toward monitoring bound-waters of hydration in proteins (Kutyshenko and Cortijo 2000).

6.2 $^{15}$N, $^1$H Cross-Saturation Spin Diffusion of Proteins as a general technique for monitoring core packing and subtle conformational changes: application to Ca$^{2+}$-Calmodulin

6.2.1 ABSTRACT

We demonstrate the application of $^{15}$N, $^1$H-detected aliphatic saturation cross-relaxation experiments as a technique for monitoring subtle changes in protein core stability and/or dynamics. Saturation of $^1$H aliphatic resonances with a low power rf pulse of ~1 sec duration causes enhanced relaxation of adjacent spins coupled via spin-diffusion. The efficiency of spin-diffusion across the network of NOE-coupled protons in a protein core can be a sensitive measure of dynamics and intramolecular distance. The method was used to confirm the presence of a recently discovered stable desolvated intermediate state of *Xenopus* Ca$^{2+}$-calmodulin which exhibits native secondary structure and is otherwise undetectable by standard $^{15}$N, $^1$H NMR techniques. The results are entirely consistent with near-UV CD, $^{19}$F solvent isotope chemical shifts, and $^1$H, $^1$H NOESY experiments. It is possible that this newly discovered intermediate state of calmodulin is relevant to presampling of ligand-binding competent states of the protein, thereby linking subtle structural changes in Ca$^{2+}$-CaM to its physiological function.
6.2.2 INTRODUCTION

Protein domains are well-ordered tertiary structures comprised of interacting secondary structure elements. Protein denaturation (whether via temperature, chemical denaturant, or pressure) disrupts intra- and inter-domain interactions that otherwise maintain the stability and fold of the protein. Electrostatic (H-bonds, salt bridges) and dipolar forces (Van der Waals) between adjacent residues and functional groups in proteins, generally serve to reduce conformational sampling and in doing so, define the native state structure. The entropic penalty for such reduction in conformational sampling is typically satisfied by a combination of positive enthalpies of formation and favourable entropy in the desolvation of hydrophobic residues. The result is that proteins generally become less dynamic in going from their denatured states to native, folded conformations. Reduced dynamics favour propagation of spin diffusion effects and in general, side chain protons near the surface of the protein or in sparsely populated regions of the protein will typically exhibit less efficient spin diffusion propagation than protons in the hydrophobic core (Kalk and Berendsen 1976). For relatively large, slow tumbling proteins in solution (>20kDa at 100 MHz, smaller at larger field strengths), the efficiency of spin diffusion can be used as a measure of the extent of core packing interactions. Spin diffusion can therefore be used to obtain useful information about the stability of proteins as a function of temperature, pressure, pH, and chemical denaturant.

6.2.3 MATERIALS AND METHODS

A plasmid (pET21b) encoding Xenopus laevis calmodulin (residues 1-148) was transformed into BL21(DE3) competent cells. A 25 mL overnight culture of transformed E. coli BL21 (DE3) cells in LB was used to inoculate 1L of M9 media (0.3% D-glucose, 0.1% $^{15}$NHCl$_4$, 100 mg/L ampicillin, 10 mg/L thiamine, 10 mg/L biotin, 1 mM MgSO$_4$, 0.1 mM CaCl$_2$). The cell cultures were left to grow at 37°C with shaking until an OD600 = 0.8 was reached. At this point, each 1L culture was supplemented with 21 mg of DL-3-fluorophenylalanine and 14 mg DL-phenylalanine (previously dissolved in a minimal amount of water) Expression was then immediately induced with 238 mg/L isopropyl β-D-1-thiogalactopyranoside (IPTG). Cell cultures were harvested after 3 hours by centrifugation at 7000 rpm for 30 minutes at 4°C. Harvested cells were then suspended in 25 mL of 50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM
imidazole, and 1 mM phenylmethylsulfonyl fluoride (PMSF) buffer at pH 8. The cells were then incubated at 4°C for 20 minutes with the addition of 1 mg/mL of lysozyme followed by sonication. The suspension was then centrifuged at 11000 rpm for 35 minutes at 4°C to pellet cell debris. The cleared lysate was then loaded onto a Ni-NTA agarose resin (Qiagen, Mississauga, Ontario, Canada) and shaken for 1 hour at 4°C. After this, the flow through was discarded and the resin washed with two column volumes of wash buffer (50 mM sodium phosphate 300 mM sodium chloride, 12 mM imidazole, and 1 mM PMSF, pH 8.0). Protein was eluted with 250 mM imidazole, 50 mM sodium phosphate 300 mM sodium chloride, and 1mM PMSF, pH 8.0. The eluted protein was precipitated by addition of cold 50% (w/v) trichloroacetic acid (TCA) to a final concentration of 6% (w/v). The protein precipitate was then centrifuged at 6500 rpm for 20 minutes at 4°C. The solution was then decanted and the protein pellet was resuspended in 1 M TRIS base, 0.5 M TRIS HCl, 1 mM MgCl$_2$, 1 mM CaCl$_2$ at pH 7.5 and left to resuspend for 12 hours. The labeled protein was then further purified using a phenyl sepharose column as described previously (Biochem. Pharmacol. 40, 153-160). Purified CaM samples were stabilized into either a) 15 mM Bis–Tris, 100 mM KCl, pH 6.72, 5 mM CaCl$_2$, or b) 0.1 M KCl, 20 mM Tris HCl, and 9 mM CaCl$_2$ buffer adjusted to pH 8. NMR samples were prepared with 10% D$_2$O for lock.

NMR Experiments. Spin diffusion experiments were performed by selectively saturating aliphatic resonances of the protein using a CW saturation pulse with a B1 field of 5200 Hz, for a period of 1 sec in a standard $^{15}$N,$^1$H HSQC. At each temperature (between 16-70 °C), two spectra were acquired: one with the saturation applied on resonance at ~0.7 ppm while the second spectrum was acquired with an off-resonance saturation pulse applied at ~15 ppm. Processing of raw FIDs was done with NMRPipe (Delaglio et al. 1995) and spectral processing was done with NMRViewJ (One Moon Scientific, Inc.).

6.2.4 RESULTS AND DISCUSSION

Preliminary experiments demonstrating utility of saturation transfer to effect cross-relaxation within the spin-diffusion network of Ca$^{2+}$-calmodulin were conducted at low (16 °C) and high (44 °C). Two sets of spectra were collected at these temperatures, one with the application of an
on-resonance saturation pulse at ~0.7 ppm, and the other with an off-resonance saturation pulse at ~15 ppm. On- and off-resonance spectra at each temperature were processed with NMRPipe and differenced (Figure 6.2). The figure clearly demonstrates that the efficiency of spin-diffusion across the $^1$H network within calmodulin is greater at low temperature (16 °C, blue) than at higher temperatures (44 °C, red).

Presumably, as temperature increases, so do protein dynamics (more rapid/frequent side chain motions resulting in less defined tertiary contacts) and there is a loss of the cohesiveness in the spin-diffusion network.

Subsequently, on and off-resonance $^{15}$N,$^1$H spectra were acquired across a 54 °C temperature range (16 °C to 70 °C). Peak assignments were made by comparing $^{15}$N,$^1$H resonances to those published previously for calcium-bound calmodulin (Mal and Ikura 2006). Peak intensities were measured for fifteen residues in both on- and off-resonance saturation experiments and the ratio of these intensities were plotted as a function of temperature (Figure 6.3). Note that most of the residues selected for analysis correspond to aliphatics and aromatics in the hydrophobic regions of calmodulin.
Interestingly, the plots for individual residues across the temperature range exhibit sigmoidal trends, with a plateau above 66 °C. At this temperature, the protein is still stable and folded as the melting temperature for calcium-loaded Xenopus calmodulin is ~110 °C (Figure 6.4).

Figure 6.4 15N,1H HSQC of fluoro phenylalanine-labeled Ca2+-Calmodulin at 70 °C. The protein is well folded at this temperature and exhibits native like secondary structure (confirmed by far-UV CD). However, near-UV CD and spin-diffusion experiments demonstrate a tertiary structure transition above 66 °C. This is consistent with the formation of a stable intermediate state.
This coincides well with near-UV circular dichroism (CD) experiments which monitor changes in tertiary structure (as opposed to far-UV which report on secondary structure). Far-UV CD and standard $^{15}\text{N}_{-1}^{1}\text{H}$ chemical shifts versus temperature suggest retention of native secondary structure (Kitevski-Leblanc et al. 2013).

The ratio between near- and far-UV CD ellipticity demonstrates a transition in tertiary structure between 60-80 °C, attributed to the onset of a stable intermediate state of the protein (Figure 6.5).

It is therefore apparent that whereas standard $^{15}\text{N}_{-1}^{1}\text{H}$ NMR spectroscopy and far-UV CD are somewhat insensitive to the detection of the intermediate state populated between 66-80 °C, near-UV CD and spin-diffusion measurements are perfectly suited for this purpose. The intermediate state of Ca$^{2+}$-Calmodulin at these temperatures was further characterized by $^{19}\text{F}$-Phe NMR, via solvent isotope and paramagnetic shifts which provided information about local solvent (H$_2$O/D$_2$O) and oxygen accessibility. Given that the $^{19}\text{F}$ reporters are fluorophenylalanines, they provide a measure of changes occurring in the hydrophobic core of the protein. Water accessibility was shown to decline with increasing temperature until a plateau was reached between ~60-70 °C, after which solvent accessibility increased. The trend for oxygen

![Figure 6.5 Changes in the degree of the tertiary structure of 3F-Phe CaM as a function of temperature, as monitored by the near-UV ellipticity signature, $\langle\theta\rangle_{253-275}$. As highlighted, a plateau in the degree of the tertiary structure seen from 65 to 80 °C signifies the onset of an intermediate state. At higher temperatures, beyond 80 °C, the protein gradually acquires a more unfolded state. Reprinted with permission from Biochemistry (2013) 52: 5780–5789. Copyright © American Chemical Society](image-url)
accessibility (and therefore hydrophobicity) was opposite to water accessibility (Figure 6.6). Both results are consistent with a gradual desolvation of the hydrophobic core (Kitevski-Leblanc et al. 2013).

Presumably, the combination of desolvation of the hydrophobic core and the increase in temperature would be accompanied by the formation of a liquid-alkane-like protein interior comprised of highly dynamic aliphatic and aromatic side chains. Indeed $^1$H,$^1$H-NOESY spectra showing aliphatic vs aromatic contacts demonstrate reduced NOE efficiency as temperature increases, consistent with this hypothesis (Figure 6.7).

![Figure 6.6](image1.png)

**Figure 6.6** (A) Temperature dependence of the $^{19}$F solvent isotope shifts, $\Delta\delta^*(H_2O)$. The solvent isotope shifts are normalized by an internal standard, 4F-Phe, and provide a quantitative measure of solvent accessibility. (B) Temperature-dependent changes in oxygen accessibility to the protein interior as measured by the temperature dependence of $^{19}$F paramagnetic shifts, $\Delta\delta^*(O_2)$, resulting from the dissolution of oxygen under a partial pressure of 20 bar. Paramagnetic shifts are normalized by the paramagnetic shifts observed for an internal standard, 4F-Phe. Reprinted with permission from Biochemistry (2013) 52: 5780–5789. Copyright © American Chemical Society

![Figure 6.7](image2.png)

**Figure 6.7** The aliphatic aromatic contact region of $^1$H,$^1$H NOESY spectra of 3-F-Phe enriched Ca$^{2+}$-CaM at 30°C (black), 50°C (red), and 70°C (blue). A mixing time of 140 ms and a $^1$H Larmor frequency of 500 MHz. Contour levels were normalized to buffer peak intensity. Reprinted with permission from Biochemistry (2013) 52: 5780–5789. Copyright © American Chemical Society
The $^1\text{H},^1\text{H}$ NOESY data indicates a loss of NOE connectivity between aliphatic and aromatic proteins due to either increased inter-residue distances or increased dynamics in the protein hydrophobic core. Enthalpic and entropic contributions to oxygen partitioning in the hydrophobic core seemed to suggest that the intermediate state is indeed stabilized by greater conformational sampling events in the vicinity of the fluoro-aromatic probes (Kitevski-Leblanc et al. 2013). In either case, the spin-diffusion network is less defined at elevated temperatures as was suggested by the $^{15}\text{N},^1\text{H}$ aliphatic cross-saturation experiments shown earlier.

6.2.5 CONCLUSIONS

The $^{15}\text{N},^1\text{H}$ aliphatic cross-saturation/spin diffusion experiment is a useful tool for monitoring subtle changes in tertiary structure in proteins, specifically in the region of the hydrophobic core and effectively complements $^1\text{H},^1\text{H}$ NOESY and near-UV CD experiments. Whereas standard $^{15}\text{N},^1\text{H}$ NMR and far-UV CD did not elude to the existence of a stable intermediate state of Ca$^{2+}$-Calmodulin, $^{15}\text{N},^1\text{H}$ cross-saturation experiments, near-UV CD, $^1\text{H},^1\text{H}$ NOESY, and both solvent isotope and paramagnetic-induced $^{19}\text{F}$ shifts provided unequivocal evidence for a stable intermediate state of the protein between ~60-76 °C. This intermediate state is characterized by a desolvated hydrophobic core and stabilized by a gain in conformational entropy (side chain dynamics) and an increase in entropy associated with the release of ordered water. It is likely that sampling of this intermediate desolvated state occurs under physiological conditions (but is populated to a much lesser extent than at 60-78 °C) and is perhaps responsible for enhancing calmodulin’s competency toward ligand binding. The release of water and increased conformational entropy would allow the protein to sample states which are more competent toward binding, thereby enhancing both selected- and induced-fit modes of association with its partner.

6.3 Application of $^{15}\text{N},^1\text{H}$ cross-saturation spin-diffusion experiments toward monitoring subtle differences in protein stability associated with the presence of chemical stabilizers or unnatural amino acid incorporation.
6.3.1 INTRODUCTION

Earlier, $^{15}$N,$^1$H cross-saturation spin-diffusion was shown to be useful toward monitoring subtle changes in protein tertiary structure and in particular, changes in the spin-diffusion coupled network of the hydrophobic core. These experiments were based on residue-specific changes in spin-diffusion connectivity between backbone amide protons and adjacent aliphatic and aromatic side chain protons of calcium-bound Xenopus calmodulin. To test the robustness of this technique toward monitoring of minute changes in protein stability, $^{15}$N,$^1$H aliphatic cross-saturation experiments were conducted for: a) standard Ca$^{2+}$-Calmodulin, b) fluoro-Phe-labeled Ca$^{2+}$-Calmodulin, c) Ca$^{2+}$-Calmodulin in the presence of 5% trifluoroethanol (a reagent known to enhance stability of alpha-helical regions of proteins) (Sönnichsen et al. 1992), and d) human Small Ubiquitin-like Modifier protein (SUMO). The cross-saturation experiments were performed at temperatures ranging from 16 °C to 70 °C.

6.3.2 METHODS

$^{15}$N,$^1$H,$^{19}$F-Phe labeled Xenopus calmodulin was expressed as described in the previous section. Non-fluorinated $^{15}$N,$^1$H-CaM was also expressed using the same protocol, with the exception that no glyphosate, DL-3-fluorophenylalanine, or DL-phenylalanine were added to the expression media. $^{15}$N,$^1$H-labeled 6X-His-SUMO was donated by the laboratory of Dr. Voula Kanelis at the University of Toronto, and was prepared as described previously (de Araujo et al. 2011). Calmodulin samples were exchanged into 15 mM Bis–Tris, 100 mM KCl, pH 6.72, 5 mM CaCl$_2$. NMR samples were between 250-500 µM protein with 10% D$_2$O for lock.

NMR experiments. $^{15}$N,$^1$H aliphatic saturation, spin diffusion NMR experiments were performed as described in the previous section on a 500 MHz Varian NMR spectrometer equipped with a standard room temperature HCN probe. Both on- and off-resonance saturation spectra were acquired over a range of temperatures between 16 °C and 64 °C. Standard pulse widths were 7-10 µsec for $^1$H and 42-46 µsec for $^{15}$N. Temperature-dependent changes in pulse widths and resonance-offsets were recalibrated at three equally spaced temperature points. This is important as small variances can lead to differences in initial intensities. Ideally, temperature-dependent effects should be calibrated at every temperature immediately prior to conducting the cross-saturation $^{15}$N,$^1$H experiments.
6.3.3 RESULTS AND DISCUSSION

Application of a low power cw saturation pulse on proton resonances at ~0.7 ppm was shown earlier to yield favorable spin-diffusion across residues in the hydrophobic core of calmodulin. The spin diffusion effect clearly extends from the aliphatic protons to the backbone amide protons being monitored in the $^{15}$N,$^1$H HSQC. Between nine to fifteen residues which exhibited sufficiently large changes in intensity between on- and off-resonant saturation spectra were assigned by comparing chemical shifts to those published previously (Mal and Ikura 2006). A comparison between datasets collected for $^{15}$N,$^1$H-labeled calmodulin in the presence and absence of 5% trifluoroethanol (TFEtOH) clearly demonstrate differences in hydrophobic core dynamics (or interatomic distances) between the two preparations (Figure). Trifluoroethanol is known to act as a cosolvent that stabilizes intramolecular hydrogen bonding (in particular for alpha helical segments of proteins) (Sönnichsen et al. 1992). Previously, it was shown that calmodulin undergoes a transition to a stable intermediate state with a desolvated hydrophobic core between 60-70 °C (Kitevski-Leblanc et al. 2013). It would seem that in the presence of trifluoroethanol, the onset of this transition is pushed toward higher temperatures, consistent with the stabilizing effects attributed to the cosolvent (Figure 6.8).

Figure 6.8 Aliphatic saturation spin diffusion Plots comparing standard $^{15}$N,$^1$H-labeled Ca$^{2+}$-CaM in the presence and absence of 5% TFEtOH. Average of the normalized integral differences between on-resonance and off-resonance saturation spectra. (A) red/blue: Plots employing only data for identical residues across both datasets (red, blue - 9 residues: I100, D64, I27, A57, N137, V136, A102, L32, and L105), green/purple: Plots incorporating additional assignable residues to the red/blue data sets. Note that addition of TFEtOH changes peak positions and so only those residues that could be definitively assigned were considered. The normalized data extends beyond unity in either dataset because of poorer S/N ratio at the lowest temperature measured. Regardless 16C was selected as the reference as spin-diffusion effects are expected to be more pronounced at lower temperatures as described earlier. (B) Structure of Calcium-bound calmodulin (PDB:3CLN) with the nine aliphatic residues common to both the No TFEtOH and 5% TFEtOH datasets shown in purple. Orange spheres show positions of the four calcium ions.
It should be noted that trifluoroethanol is thought to enhance the likelihood of formation of helical structures inherent to the primary sequence of a protein rather than directly inducing helices regardless of sequence identity (Sönnichsen et al. 1992). It does so by stabilizing intramolecular hydrogen bonds while seemingly to disfavoring hydrophobic interactions. It is perhaps surprising then, that in the presence of 5% TFEtOH spin-diffusion effects seem to remain stronger (ratio stays closer to unity in plots) until higher temperatures than in the absence of the cosolvent. If aliphatic saturation cross-saturation effects are dependent on proton-proton distances within the hydrophobic core, would the effects not be stronger in the presence of greater number of Van der Waals interactions? This may be explained by the fact that protein interiors often contain bound water, with calmodulin in particular exhibiting a more solvated core upon calcium binding, and that spin-diffusion can propagate across such waters of hydration (Stoclet et al. 1987, Lee et al. 2000). In fact, this is entirely consistent with the delayed onset (higher temperature) transition to the desolvated intermediate in the presence of TFEtOH. The native-state, which has a relatively solvated core is stabilized by trifluoroethanol and only at higher temperatures does the protein lose bound water to adopt the desolvated intermediate.

The presence of trifluoroethanol induces relatively subtle changes in protein conformation for regions that have no propensity toward adopting helical structures (Sönnichsen et al. 1992). Unlike thermo-stabilizing mutations to proteins, the overall changes caused by the addition of trifluoroethanol could be said to be minimal. What then for the introduction of unnatural amino acids into proteins?

Earlier, the introduction of fluoro-phenylalanine into Xenopus calmodulin was shown to allow for direct monitoring of solvent and oxygen accessibility to the hydrophobic interior of Ca$^{2+}$-CaM (Kitevski-Leblanc et al. 2013). The replacement of a proton with a fluorine nucleus is generally thought to be minimally perturbing to proteins given that the radius of the $^{19}$F atom is comparable to the size of a water molecule (Lau and Gerig 1997). That said, it has been observed experimentally, that the incorporation of F-Phe into Calmodulin is only well tolerated up to ~70% (70% F-Phe, 30% Phe) and at higher levels of labeling, the protein is destabilized (unpublished). Similarly changes in protein stability upon incorporation of other mono-
fluoroaromatics been noted by others (Schuler et al. 2002). Consequently, preparations of F-Phe-labeled calmodulin involve partial labeling of the protein to ensure stability. Such samples exhibit one fluorine resonance per phenylalanine residue, consistent with the presence of a homogeneous sample rather than a superposition of structures which would otherwise be expected for unstable proteins (Figure 6.9).

Figure 6.9 X-ray crystal structure and $^{19}$F NMR spectrum of CaM. (a) X-ray crystal structure of the calcium-loaded state of CaM, highlighting the location of each phenylalanine residue (PDB file 3CLN). (b) $^{19}$F NMR spectrum, collected at 50 °C, of 70% 3F-Phe fractionally labeled CaM. A spectral deconvolution is shown to distinguish all eight 3F-Phe resonances. Reprinted with permission from Biochemistry (2013) 52: 5780–5789. Copyright © American Chemical Society

Though $^{15}$N,$^{1}$H NMR spectra of F-Phe-labeled Ca$^{2+}$-CaM are superimposable with those of non-fluorinated calmodulin, it is likely that the introduction of fluorines induces minute structural perturbations in the protein that are otherwise undetectable by standard two-dimensional NMR experiments. In order to determine whether this was indeed the case for F-Phe calmodulin, the $^{15}$N,$^{1}$H aliphatic saturation, cross-relaxation experiments were performed on standard Ca$^{2+}$-calmodulin (CaM), fluoro-phenylalanine-labeled CaM, CaM + 5% trifluoroethanol, and on human 6X-his-SUMO (a control exhibiting simple two-state unfolding without a stable intermediate). Unlike the earlier cross-saturation experiments described above, the data analyses were different for these experiments. Rather than measuring peak integrals for select assigned $^{15}$N,$^{1}$H resonances, the entire NH regions (5-11.6 ppm) were integrated for both
on- and off-resonance saturation spectra and differenced. As before, however, the differences obtained at a given temperature were normalized to that obtained at the lowest temperature (Figure 6.10). The advantage of using the entire NH region is that it allows for general comparisons across proteins of different structure and primary sequence.

Yet, there are disadvantages for the simple comparison of entire NH regions in the aliphatic saturation spin-diffusion experiment. From the figure above, it is difficult to distinguish the plot for Ca\(^{2+}\)-calmodulin from the plot of calmodulin in the presence of 5% trifluoroethanol. While little to no difference is apparent between these two curves, earlier it was shown that clear differences exist between these two preparations when comparing only those side chains within the hydrophobic interior. Therefore, considerable sensitivity to minute differences in the spin-diffusion network is lost by simply comparing entire NH regions. In general, however, comparing the efficiency of spin-diffusion across temperature (using entire NH integrals) will give insight into transitions to intermediate or unfolded states. The 6X-his-SUMO plot exhibits an almost linear change in spin-diffusion efficiency across temperature suggesting a gradual unfolding process (two state folding-unfolding transition) whereas each of the calmodulin plots exhibit trends that can be reasonably fit to sigmoidal curves indicating more than two-state transitions, given that global unfolding of CaM occurs between 110-115 °C (Kitevski-Leblanc et al. 2013). It is evident that the F-Phe labeled calmodulin exhibits a shifted transition temperature.
relative to non-fluorinated CaM. This is highly suggestive that biosynthetic labeling with fluorine introduces minute changes to the protein stability. Alternatively, if one assumes that the transition is to the desolvated intermediate, the results may simply suggest that the F-Phe CaM more readily desolvates in order to accommodate the increased hydrophobicity contributed by replacement of protons with fluorine nuclei.

6.3.4 CONCLUSIONS

$^{15}$N,$^1$H aliphatic saturation, spin diffusion (cross-relaxation) experiments are an ideal complement to other spectroscopic techniques for probing minute differences in protein stability. Whereas standard $^{15}$N,$^1$H HSQC spectra alone (which are commonly used to profile protein structure by NMR) may, in some cases, provide little evidence for subtle differences in protein core dynamics and stability, the spin diffusion experiments described above can readily elucidate such changes. Calmodulin biosynthetically labeled with fluoro-phenylalanine exhibits $^{15}$N,$^1$H NMR spectra that are superimposable with spectra of non-fluorinated CaM and without further inspection, it would seem that fluorination has little effect on the protein. A stable intermediate state of calmodulin, which has eluded detection until recently because of its identical secondary and nearly-identical tertiary structure to the native state of the protein, can be probed by experiments sensitive to small differences in hydration, hydrophobicity, and core dynamics. Near-UV CD, $^1$H,$^1$H NOESY, solvent-isotope induced fluorine chemical shifts, and the $^{15}$N,$^1$H aliphatic saturation spin diffusion experiments are all techniques that have provided evidence for the desolvated intermediate. Several major advantages of the spin-diffusion experiment are: i) it is simple to implement, ii) it requires no unnatural amino acid incorporation, and iii) it can be informative even in the absence of resonance assignments via integration of the entire NH region of both on- and off-resonance saturation HSQC spectra. Though for greater sensitivity to minute structural differences in core stability and dynamics, analysis of NH integrals for individual core hydrophobic residues is ideal. This has the added benefit of providing insight into those residues that are most strongly spin-diffusion coupled to the saturated resonances, thereby providing more localized structural information.
6.3.5 REFERENCES


Bauschlicher CW Jr, Partridge H (1995) The Sensitivity of B3LYP Atomization Energies to the Basis Set and a Comparison of the Basis Set Requirements for CCSD(T) and B3LYP. NASA Ames Research Center, Moffett Field, CA 94035, USA DOI: 10.1016/0009-2614(95)91855-R


CHAPTER 7

Conclusions and Future Directions

7.1 CONCLUSIONS

The work described in this thesis has an underlying theme focused around site-specific labeling of proteins. Reductive methylation has proven to be a useful approach for incorporating NMR-active $^{13}$C-methyl tags into proteins under non-denaturing conditions for structural/functional studies. Whereas we have introduced methodological improvements to the reductive methylation of proteins (stoichiometry, monoalkylation, CH$_2$D-labeling – Chapter 2), $^1$H and $^{13}$C spectra for methylated proteins have always exhibited complex patterns that have, until now, eluded complete characterization. The kinetics and thermodynamics of dimethyllysine methyl exchange provide detailed insight into structurally and functionally relevant interactions and dynamic processes within proteins. Understanding these processes for model compounds and comparing them to those occurring within biologically-relevant systems helps to elucidate important interactions. As an example, leucine side chain methyl relaxation cpmg dispersions have been used to monitor excited states of proteins by NMR (Hansen et al. 2010). Given the similarities between dimethylamino groups and leucine side chains shown earlier, one obvious direction for future studies would be to monitor dimethyllysine methyl cpmg dispersions. An in-depth understanding of the proton-exchange and nitrogen inversion kinetics described earlier would therefore be key to discerning functionally-relevant changes in relaxation rates from uncoupled kinetic processes. Monitoring methyllysine dynamics via cpmg dispersions is a potential way of characterizing protein conformational “switching” for those proteins exhibiting salt-link dependent activation kinetics such as the β2-adrenergic receptor (Bokoch et al. 2010). It would therefore be possible to monitor drug-dependent changes in activation-deactivation kinetics simply by monitoring salt-link switching via dimethyllysines.

One of the more recent advances in $^{19}$F-labeling of proteins was the introduction of glyphosate to inhibit aromatic amino acid synthesis and the addition of fluoro-phenylalanine, fluoro-tryptophan, and or fluoro-tyrosine to the expression media (Kim et al. 1990). This biosynthetic labeling strategy has proven useful on many occasions for structural biological studies (Kim et
The exquisite sensitivity of the fluorine nucleus to changes in local electrostatic and van der Waals interactions makes it an ideal probe for studies of protein structure and function. \(^{19}\text{F}\)-labeling of hamster prion with fluoro-phenylalanine allowed for the elucidation of soluble large-oligomers, which have never before been observed spectroscopically, and characterization of the exchange between monomeric, octameric and large oligomeric states (Larda et al. 2013). Thermodynamic analyses demonstrated differences between the aggregation propensity of wt hamster prion and a disease-relevant mutant. Whereas these prior experiments employed prion protein constructs exhibiting stable alpha-to-beta conversion, we have recently identified a mutant exhibiting improved exchange kinetics between the alpha and beta states under conditions of mild denaturant. This opens the possibility toward real-time monitoring of alpha to beta templating and conversion, which are thought to be critical events in priogenesis. Studies of this nature are critical for understanding the prion misfolding pathway and ultimately how prion diseases work on a molecular level.

Whereas cross-saturation spin-diffusion experiments are regularly employed for monitoring and mapping ligand binding and also for characterizing protein-protein binding, the technique is less often employed for monitoring intramolecular conformational changes in proteins. The aliphatic saturation spin-diffusion methodology described earlier demonstrates superb sensitivity for monitoring minute changes in protein conformation and stability and can be used to identify dynamic intermediate states that are otherwise invisible to more traditional NMR methods. Whereas the use of first-increment amide spectra in the absence of full \(^{15}\text{N},^{1}\text{H}\) assignment allows for detection of conformational transitions, it was shown that residue-specific information (especially for those residues tightly coupled to the aliphatic spin diffusion network) provides much more sensitive information about core-packing and intramolecular dynamics. We have recently compared the spin-diffusion efficiency for either only N- or only C-terminal residues for calmodulin and found that the N-terminal residues are much more affected by the aliphatic saturation than are C-terminal residues (Figure 7.1).
This is entirely consistent with recent $^{19}$F cpmg measurements of fluoro-phe labeled calmodulin which demonstrate that the N-terminal domain is intrinsically more flexible/dynamic than the C-terminal domain (Hoang and Prosser 2014). It is therefore apparent that the aliphatic spin-diffusion experiments can delineate differences in spin-diffusion networks within a protein across different sets of coupled residues. It might therefore be useful for identifying allosteric networks/pathways in proteins with complex functions including GPCRs (Bhattacharya and Vaidehi 2014, Leioatts et al. 2014).

![Graph showing spin diffusion efficiency versus temperature for residues in either the C-terminal domain or N-terminal domain of Ca$^{2+}$-calmodulin.](image)

**Figure 7.1** Spin diffusion efficiency versus temperature for residues in either the C-terminal domain or N-terminal domain of Ca$^{2+}$-calmodulin. N-terminal domain exhibits greater magnitude of changes in spin diffusion efficiency across temperature, consistent with its greater flexibility/dynamics as shown via cpmg in recent studies (Hoang and Prosser 2014).

Finally, whereas the indole-based tagging methodology described in this thesis was developed as a new way for site-specifically incorporating $^{19}$F into proteins, it is far better suited as a general bioconjugation method to link amines. Therefore, although fluoroindoles were shown to be more sensitive to solvation effects (and therefore presumably local chemical environment) than many common $^{19}$F tags for protein labeling, the indole-conjugation chemistry should not be sold short on use for $^{19}$F-labeling of proteins only. Currently, we are working on employing this conjugation strategy for site-specific functionalization of proteins with moieties which confer favourable solution properties and stability to proteins under physiologically relevant conditions.
Other applications include crosslinking, immobilization, $^{18}$F-PET imaging, and ICAT tagging for mass spectrometry.

7.2 REFERENCES


