Investigating the Folding Network of Calmodulin Using Fluorine NMR

by

Joshua Nam Hoang

A thesis submitted in conformity with the requirements for the degree of Master of Science
Department of Chemistry
University of Toronto

© Copyright by Joshua Hoang 2013
Investigating the Folding Network of Calmodulin Using Fluorine NMR

Joshua Nam Hoang

Master of Science

Department of Chemistry
University of Toronto

2013

Abstract

Protein folding pathways can be extraordinarily complex. In this study, circular dichroism (CD) and $^{19}$F NMR are used to investigate the folding network of calmodulin, a calcium-binding protein, which is biosynthetically enriched with 3-fluorophenylalanine. In calmodulin’s calcium-loaded state, CD experiments identify the existence of a folding intermediate along a heat-denaturation pathway. In comparison to the native state, $^{19}$F NMR solvent isotope shifts reveal decreased accessibility of water to hydrophobic core, whereas $^{18}$O paramagnetic shifts show increased hydrophobicity of this folding intermediate. $^{15}$N-$^1$H and methyl $^{13}$C-$^1$H HSQC NMR spectra demonstrate that this folding intermediate retains a near-native tertiary structure, whose hydrophobic interior is highly dynamic. $^{19}$F NMR CPMG relaxation dispersion measurements suggest that this near-native intermediate state is transiently adopted below the temperature associated with its onset. The folding network also involves an unproductive off-pathway intermediate. In contrast, calmodulin’s calcium-free state exhibits a simpler folding process which lacks discernible intermediates.
Acknowledgments

I would like to start off by thanking my supervisor, Prof. Scott Prosser, for his constant support and guidance over the course of my project. The knowledge and experience I have gained while under your tutelage, has been invaluable in my development as a working individual. Your great sense of humour has also made my time in the lab enjoyable and is something that I will miss.

I would like to thank members of the Prosser group, Sacha Larda, Rohan Alvares and Tae Hun Kim for insightful discussions and making a stimulating work environment. Many thanks also go out to former lab members of the Prosser group. I am grateful to William Thach for supervising me during my undergraduate project and teaching me the basics of protein expression. I also appreciate Sameer Al-Abdul-Wahid for all his help with NMR experiments. Thanks also to Julianne Kitevski-Leblanc whose work on calmodulin has made this project possible. I am also thankful to the Kanelis and Macdonald groups for allowing me use of their equipment.

Finally, I would like to thank my parents for their unwavering support throughout my entire time at the University of Toronto.
Table of Contents

Abstract ii
Acknowledgments iii
Table of Contents iv
List of Tables vi
List of Figures vii
Abbreviations x

Chapter 1: Introduction

1.1 Protein Folding 1
1.2 Calmodulin 3
1.3 NMR in Protein Studies 4
1.3.1 $^1$H-$^{15}$N Heteronuclear Single Quantum Coherence (HSQC) 5
1.3.2 Pulsed Field Gradient Stimulated Echo (PFG STE) 6
1.3.3 Transverse ($T_2$) and Longitudinal ($T_1$) Relaxation Time 8
1.3.4 Carr-Purcell-Meiboom-Gill (CPMG) Relaxation Dispersion 8
1.3.5 Spin Diffusion 10
1.4 Circular Dichroism (CD) 11
1.5 Fluorine-19 as a Tool to Study Protein Structure and Dynamics 12
1.5.1 Incorporation of Fluorine Labels 13
1.5.2 Biosynthetic Labeling of Calmodulin with 3-Fluorophenylalanine 14
1.5.3 Measuring Solvent Exposure Using Fluorine NMR 16
1.5.4 Examples of Fluorine NMR Applications with Proteins 17

Chapter 2: Materials and Methods

2.1 Expression and Purification of $^{15}$N-Enriched and 70% 3-Fluorophenylalanine Labeled Calmodulin 20
2.2 Isolating the Two Domains of Calmodulin (TR1C and TR2C) 21
2.3 Circular Dichroism Spectroscopy 21
2.4 NMR Spectroscopy 21
2.4.1 Solvent Induced Isotope Shifts 22
2.4.2 O₂ Paramagnetic Shifts 22
2.4.3 Pulsed Field Gradient Stimulated Echo 22
2.4.4 Spin Diffusion 23
2.4.5 ¹⁵N-¹H HSQC and ¹³C-¹H HMQC Spectra 23
2.4.6 CPMG Relaxation Dispersion Experiments 23
2.4.7 ¹⁹F Transverse (T₂) and Longitudinal (T₁) Relaxation Measurements 24

2.5 Estimating Populations of the Native and Near-Native States 24

Chapter 3: Results
3.1 Circular Dichroism and ¹⁹F NMR Evidence of a Folding Intermediate along a Heat Denaturation Pathway 26
3.2 Solvent Exposure and Hydrophobicity along the Temperature Denaturation Pathway 33
3.3 NMR Studies of Compactness of the Native State and the Desolvated Near-Native Intermediate 38
3.4 Folding Kinetics between the Near-Native Intermediate and Folded State 41
3.5 Identifying a Possible Off-Pathway Intermediate Within Calmodulin’s Folding Network 46
3.6 Possible Implications of the Near-Native Desolvated Intermediate in Misfolding 52
3.7 The Effect of Calcium-Binding on the Folding Pathway of Calmodulin 55

Chapter 4: Discussion and Conclusions
4.1 The Role of the Near-Native Desolvated Intermediate in Folding and Misfolding 60
4.2 The Effect of Off-Pathway Intermediates on Folding 62
4.3 The Effect of Calcium-Binding on Calmodulin’s Folding Network 63
4.4 Conclusions and Future Directions 64

References 66
List of Tables

Table 1. Enthalpies and entropies of partitioning of oxygen in the vicinity of the eight 3F-Phe reporters 37

Table 2. Extracted chemical exchange rates from $^{19}$F CPMG dispersion profiles of 3F-Phe CaM 47

Table 3. Extracted chemical exchange rates from $^{19}$F CPMG dispersion profiles of 3F-Phe TR1C and TR2C 52
List of Figures

Figure 1. Rugged free energy landscape diagram for a fast-folding helical protein 1
Figure 2. NMR and X-ray structures of calmodulin 3
Figure 3. Folding network of holo-CaM revealed by single molecule experiments 4
Figure 4. $^1$H-$^{15}$N HSQC spectra of a tumour suppressor protein 5
Figure 5. NMR pulse sequence for a PFG STE experiment 6
Figure 6. Signal intensity plot from a PFG STE experiment 7
Figure 7. Energy landscape model showing an excited state which is amendable to study by CPMG relaxation dispersion experiments 9
Figure 8. Basic overview of a CPMG relaxation dispersion experiment 9
Figure 9. Pulse sequence for a CPMG relaxation dispersion experiment 9
Figure 10. General scheme of a spin diffusion experiment 11
Figure 11. CD ellipticity profiles for various secondary and tertiary structure elements 12
Figure 12. General reaction pathway used to attach a fluorine label to the thiol group of a cysteine 13
Figure 13. X-ray structure of calmodulin highlighting the locations of the phenylalanine residues 14
Figure 14. $^{19}$F NMR spectrum of 70% 3F-Phe fractionally labeled CaM 15
Figure 15. $^{19}$F NMR spectra of 3F-Phe labeled CaM at various different labeling levels 15
Figure 16. $^{19}$F NMR spectra of 3F-Phe labeled CaM in 10% D$_2$O and 100% D$_2$O 16
Figure 17. $^{19}$F NMR spectra of 3F-Phe labeled ShaPrP 17
Figure 18. $^{19}$F NMR spectra of TET labeled β2AR bound to various ligands 18
Figure 19. $^{19}$F NMR spectra of fluorine labeled proteins in *E. coli* cells 19
Figure 20. Far-UV CD ellipticity profiles of 3F-Phe CaM at 35°C and 70°C 26
Figure 21. $^{15}$N-$^1$H HSQC spectra of 3F-Phe CaM 27
Figure 22. Near-UV CD profiles as a function of temperature for 3F-Phe CaM 27
Figure 23. Near-UV ellipticity signature as a function of temperature for 3F-Phe CaM 28
Figure 24. Near-UV ellipticity signature as a function of temperature for unlabeled CaM 29
Figure 25. $^{19}$F NMR spectra of 70% 3F-Phe CaM as a function of temperature 30
Figure 26. $^{19}$F NMR spectra peak line widths of 3F-Phe CaM as a function of
temperature

Figure 27. $^{19}$F chemical shifts of 3F-Phe CaM as a function of temperature

Figure 28. Fits of $^{19}$F chemical shifts to a two-state model

Figure 29. Relative fraction of the near-native intermediate state as a function of temperature

Figure 30. Temperature dependence of $^{19}$F solvent isotope shifts of 3F-Phe CaM

Figure 31. Temperature dependence of $^{19}$F paramagnetic shifts of 3F-Phe CaM

Figure 32. Temperature dependence of hydrophobicity of 3F-Phe CaM

Figure 33. Spin diffusion $^{15}$N-$^1$H HSQC spectra of 3F-Phe CaM

Figure 34. Spin diffusion measurements of 3F-Phe CaM as a function of temperature

Figure 35. Hydrodynamic radius measurements of 3F-Phe CaM as a function of temperature

Figure 36. $^{13}$C-$^1$H HMQC methyl spectra of 3F-Phe CaM

Figure 37. Graphical representation of the protein folding free energy landscape for CaM

Figure 38. $^{19}$F spectra of 3F-Phe CaM at various $v_{CPMG}$ values

Figure 39. $^{19}$F CPMG relaxation dispersion profiles for residues F68, F19, and F92 at 57°C

Figure 40. $^{19}$F CPMG relaxation dispersion profiles for 3F-Phe CaM with 5% TFE at 51°C

Figure 41. Order parameter values for 3F-Phe CaM

Figure 42. $^{19}$F CPMG relaxation dispersion profiles for residues F68, F19, and F92 as a function of temperature

Figure 43. Temperature dependence of the chemical exchange rates of 3F-Phe CaM

Figure 44. Folding and unfolding rates of 3F-Phe CaM

Figure 45. Schematic representation of the folding/unfolding pathway of CaM

Figure 46. $^{19}$F NMR spectra of 3F-Phe full length CaM, TR2C, and TR1C at 45°C

Figure 47. $^{15}$N-$^1$H HSQC spectra of 3F-Phe full length CaM, TR2C, and TR1C at 45°C

Figure 48. Temperature dependence of the chemical exchange rates of 3F-Phe TR1C and TR2C

Figure 49. An inset of the $^{19}$F NMR spectra of 3F-Phe CaM collected at various time points during the incubation period
Figure 50. Change in the integral of the aggregate peak over incubation period in the absence and presence of 5% TFE

Figure 51. Graphical representation of the complex folding network of CaM

Figure 52. $^{19}$F NMR spectra of 70% 3F-Phe apo-CaM at various temperatures

Figure 53. $^{15}$N-$^1$H HSQC spectra of apo-CaM at 16°C and 56°C

Figure 54. $^{19}$F chemical shifts of 3F-Phe apo-CaM as a function of temperature

Figure 55. $^{19}$F NMR spectra of 70% 3F-Phe apo-CaM at various $\nu$CPMG values

Figure 56. Graphical representation of water and oxygen accessibility in the native and the near-native intermediate states of CaM
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3F-Phe</td>
<td>3-fluorophenylalanine</td>
</tr>
<tr>
<td>β2AR</td>
<td>beta-2 adrenergic receptor</td>
</tr>
<tr>
<td>CaM</td>
<td>calmodulin</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CPMG</td>
<td>Carr-Purcell-Meiboom-Gill</td>
</tr>
<tr>
<td>GPCR</td>
<td>G protein-coupled receptors</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multiple quantum coherence</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>PFG STE</td>
<td>pulsed field gradient stimulated echo</td>
</tr>
<tr>
<td>RCO</td>
<td>relative contact order</td>
</tr>
<tr>
<td>TET</td>
<td>2,2,2-trifluoroethanethiol</td>
</tr>
<tr>
<td>TFE</td>
<td>trifluoroethanol</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction


1.1 Protein Folding

The mechanism by which proteins fold is extraordinarily complex and remains a focused research area in structural biology. Despite the astronomical number of possible conformations available to polypeptides, folding is remarkably fast.¹ Proteins accomplish this feat in part because of early nucleation events, occurring on the time scale of tens of nanoseconds, in which secondary structure and main-chain hydrogen bonds are rapidly established, thereby guiding the search toward a compact folded state.² Though there have been many proposed theories to describe the general process of protein folding, the energy landscape theory is the most widely accepted. The energy landscape theory states that the fully folded native state is achieved through the organization of partially folded structures.³ This idea of protein folding can be modeled by a rugged free energy landscape diagram as shown in Figure 1.

Figure 1 (Left). Schematic representation of a rugged free energy landscape for a fast-folding helical protein. The width of the funnel represents the configurational entropy while the depth of the funnel represents the free energy of an individual configuration. The parameter Q, represents the fraction of native contacts. High energy unfolded states occupy the top of this funnel while the low energy native state lies at the funnel minimum. Along the folding pathway, the protein can encounter kinetic traps, indicated by the small wells near the bottom of the funnel, which are a result of folding intermediates (eg. molten globule states).⁴ © Reprinted from Folding and Design, 1996, 1, 441-450 with permission from Elsevier.
The funneled organization of the energy landscape is responsible for the unique ability of proteins to fold quickly, as it narrows the configurational space of the conformational search.\textsuperscript{5} This view of protein folding also suggests that there are multiple pathways in the folding process where the protein can sample a large ensemble of conformers before achieving a fully folded state.\textsuperscript{6} Often, an intermediate state is acquired, where the ensemble of conformers are generally described as being near-native in structure. These intermediate states give rise to kinetic traps which affect the overall folding rate and add ruggedness to the energy landscape.\textsuperscript{7}

There is also the case of two-state folding. Many small proteins have been shown to fold via a two-state process where the folding pathway is populated by only unfolded and folded states.\textsuperscript{8,9} Folding rates of two-state folders can be reliably predicted from a relative contact order (RCO) parameter.\textsuperscript{10} The RCO is defined by the average sequence separation between all pairs of contacting residues, normalized by the number of residues, in the protein’s native state topology.\textsuperscript{10} Small proteins, stabilized by short-range interactions, are expected to fold faster than larger systems stabilized by long-range contacts. This observation was later generalized to 2- and 3-state folders in the sense that the intrinsic rate of formation of folded intermediates and native states were both found to be similarly correlated to native topology.\textsuperscript{11,12}

It has been established that the protein folding process is driven primarily by the hydrophobic effect.\textsuperscript{13} The “hydrophobic collapse” model, suggests that the initial step of protein folding involves polypeptide chain compaction, mediated by hydrophobic side-chain interactions, to form a collapsed state where secondary structure is formed.\textsuperscript{14} This is then followed by a second transition where water molecules are cooperatively squeezed out of the hydrophobic core as the protein proceeds to its final folded state. This model suggests that a molten globule intermediate state is often formed along the protein folding pathway. Many globular proteins have been shown to adopt intermediate states that are molten-globule-like.\textsuperscript{15,16,17} Their conformations, which are expanded and highly hydrated, retain native-like secondary structure yet lack significant tertiary structure. By contrast, in rare cases, the folding intermediate is believed to adopt a dry molten globule state, whose interior is best characterized as largely desolvated, liquid alkane-like and stabilized by configurational entropy.\textsuperscript{18,19} In this case, the final folding step is characterized by restriction of side chains and increased van der Waals interactions. Due to the fact that folding intermediates are typically short-lived and exist only transiently, characterization of these states still remains a large challenge.
In this study, we use $^{19}$F NMR to explore the protein folding/unfolding events of a water soluble protein, calmodulin. Experiments will be focused on identifying potential folding intermediates as well as measuring changes in solvent accessibility and hydrophobicity along the folding pathway. We also investigate conformational dynamics associated with the folding network.

1.2 Calmodulin

Calmodulin (CaM) is a ubiquitous calcium-binding protein which activates a variety of enzymes depending on the intracellular calcium levels.\(^{20}\) CaM is a well-structured protein (16,680 Da) made up of 148 amino acid residues which are organized into two structurally similar calcium binding domains (N-domain, residues 1-77; C-domain, residues 78-148) connected by a flexible tether.\(^{20}\) Each domain contains two EF-hand motifs which are made up of two alpha helices (E and F) connected by a 12-residue loop. The calcium ion binds to the loop region with a pentagonal bipyramidal symmetry.\(^{21}\) Since each EF-hand motif can coordinate one calcium ion, there are a total of four calcium ion binding sites per CaM molecule. CaM exists in two dominant conformational states, calcium-loaded (holo) and calcium-free (apo) states (Figure 2).

![Figure 2. NMR and X-ray structures of calmodulin. Calcium-free state (apo) conformation of calmodulin (left). Calcium-loaded state (holo) of calmodulin (middle). Calcium-loaded calmodulin bound to a binding partner (right). Calcium ions are represented by the yellow spheres. Hydrophobic patches involved in ligand binding are highlighted in red.\(^{22}\) © Reprinted from Proc. Natl. Acad. Sci. U. S. A., 2006, 103, 13968-13973 with permission from National Academy of Sciences, USA.](image-url)
Upon saturation of the calcium binding sites, apo-CaM undergoes large conformational changes. The binding of calcium induces these conformational changes due to the inclusion of the seven-coordinating ligands of the calcium ion. These conformational changes expose methionine-rich hydrophobic patches on CaM. The exposed hydrophobic patches form neat grooves which increase the binding affinity of CaM to a number of different binding partners thus, making CaM a versatile regulatory protein.

In the case of this thesis, our interest in studying CaM does not lie in the protein itself, but in using CaM as a model system for studying protein folding events. Recent single molecule force microscopy studies suggest that CaM exhibits a complex folding process, involving both on- and off-pathway folding intermediates (Figure 3).

![Figure 3. Folding network of holo-CaM as envisaged by single molecule experiments. Single molecule experiments reveal that the folding pathway of CaM consists of both on- and off-pathway intermediates.](image)

Figure 3. Folding network of holo-CaM as envisaged by single molecule experiments. Single molecule experiments reveal that the folding pathway of CaM consists of both on- and off-pathway intermediates. © Reprinted from Science, 2011, 334, 512-516 with permission from The American Association for the Advancement of Science.

However, a detailed microscopic description of the nature of the folding intermediates is still missing, as is a description of potentially faster folding steps. Focus will be placed on studying topology and dynamic changes within the hydrophobic core during the folding/unfolding process.

1.3 NMR in Protein Studies

NMR is powerful and vital technique for studying protein structure and dynamics. Today, there are almost 10,000 solution NMR structures available in the protein data bank. These structures
not only include the native states of proteins, but using advanced NMR methods the structure of a number of different weakly populated folding intermediates have also been solved.\textsuperscript{25,26} These intermediate states would otherwise go undetected by X-ray crystallography. The utility of NMR is not limited to only structural studies, but can also be used to acquire information on conformational equilibria and dynamics on a wide range of time scales in a protein system.\textsuperscript{27,28} In the following section, a number of solution NMR spectroscopy methods used in this study will be described.

1.3.1 $^1$H-$^{15}$N Heteronuclear Single Quantum Coherence (HSQC)

$^1$H-$^{15}$N HSQC experiments are routinely used to obtain structural information of proteins.\textsuperscript{29} $^1$H-$^{15}$N HSQC experiments are able to detect the one bond correlation between a backbone amide, $^{15}$N, and its corresponding hydrogen proton and thus, are essential for structure elucidation of the peptide backbone (secondary and tertiary structure).

![Figure 4](image.png)

**Figure 4.** $^1$H-$^{15}$N HSQC spectra of a tumour suppressor protein (p53C) at pH 7.2 (blue) and 5.0 (red). At pH 7.2, the observation of sharp and disperse peaks in the $^1$H-$^{15}$N HSQC spectrum is indicative of a well-folded protein. At pH 5, the loss of the number of peaks as well as the reduced dispersion suggests the formation of a molten globule state.\textsuperscript{30} © Reprinted from *J. Biol. Chem.*, 2010, 285, 2857-2866 with permission from The American Society for Biochemistry and Molecular Biology.

In a $^1$H-$^{15}$N HSQC spectrum, a single peak typically arises for each backbone and side chain $^1$H-$^{15}$N group in a protein, excepting prolines. The dispersion, intensity, and the number of observed
peaks in the spectrum correlates directly to the overall secondary and tertiary state of the protein. A well-folded protein would show intense, sharp and disperse peaks, whereas an unfolded protein would produce a $^1$H-$^{15}$N HSQC spectrum with reduced dispersion and broad peaks centered around 8.3 $^1$H ppm. $^{31}$ $^1$H-$^{15}$N HSQC experiments have been useful for the identification of molten globule folding intermediates. $^{30}$ Figure 4 shows the $^1$H-$^{15}$N HSQC spectra a tumour suppressor protein (p53C) at pH 7.2 and 5.0. Based on the disappearance of a number of resonance signals and the reduced dispersion of the peaks at pH 5.0, the onset of formation of a molten globule is detected at low pH.

1.3.2 Pulsed Field Gradient Stimulated Echo (PFG STE)

To gain a complete understanding of the structural properties of a protein, a description of its molecular dimensions is necessary. In particular, pulsed field gradient stimulated echo (PFG STE) NMR experiments can be used to characterize different protein conformations by measuring their effective hydrodynamic radius and diffusion coefficient. $^{32}$ The importance of hydrodynamic radius measurements of proteins lies in the fact that they can be used to easily distinguish the difference between native, intermediate, and unfolded states. Proteins are typically more expanded in their unfolded state. $^{33}$ Figure 5 shows the general NMR pulse sequence of a pulse field gradient stimulated echo experiment.

**Figure 5.** The NMR pulse sequence of a pulsed field gradient stimulated echo NMR experiment. $^{34}$ © Reprinted from *Biochim. Biophys. Acta: Biomembranes,* 2007, 1768, 1805-1814 with permission from Elsevier.

The PFG STE NMR pulse sequence consists of an initial $90^\circ$ pulse followed by a delay period, $\tau_2$. During this delay period, $\tau_2$, a gradient pulse with an amplitude, g, and a duration of, $\sigma$, is applied. This delay period is followed by a second $90^\circ$ pulse and a subsequent mixing time, $\tau_1$. A third $90^\circ$ pulse is then applied followed by another delay period, $\tau_2$, which consists of a second gradient pulse. The signal intensity of a PFG STE experiment depends on the
gradient pulse strength, \( g \), the diffusion time, \( \Delta = \tau_1 + \tau_2 \), and the diffusion coefficient, \( D \), of the protein. The intensity of the stimulated echo decays according to the following equation:

\[
I = I_0 \exp \left( \frac{-2\pi^2}{T_2} \right) \exp \left( \frac{-\pi^2}{T_1} \right) \exp(-\gamma^2 g^2 \delta^2 D \left[ \Delta \delta^3 \right])
\]

Equation 1

where \( T_1 \) and \( T_2 \) are longitudinal and transverse relaxation times.

This equation can be simplified to:

\[
I = I_0 \exp(-cg^2D)
\]

Equation 2

where \( c \) is a constant.

To extract the data, the signal intensity \( (I/I_0) \) is plotted as a function of \( g^2 \) where the exponential decay of the graph can be fit to determine a diffusion coefficient, \( D \). The hydrodynamic radius is then obtained by Stokes law, \( R_H = k_B T / 6\pi \eta D \), where \( k_B \) is the Boltzmann constant, \( T \) is the temperature in Kelvin, \( \eta \) is the viscosity of the solution, \( D \) is the diffusion coefficient and \( R_H \) is the hydrodynamic radius.\(^{35}\) Figure 6 shows an example of a signal intensity versus \( g^2 \) plot obtained from a PFG STE NMR experiment.

Figure 6. Graph of the signal intensity versus the square of the field gradient strength, \( g^2 \), obtained from a PFG STE experiment. Exponential decays were fit to Equation 2 to extract a diffusion coefficient for the four identified states (native, intermediate, denatured, and unfolded).\(^{36}\) © Reprinted from J. Mol. Biol., 2013, 425, 273-291 with permission from Elsevier.

By fitting the decays to Equation 2 and obtaining a diffusion coefficient, the authors were able to identify four different states of \( \beta_2 \)-microglobulin based on their differing hydrodynamic radiiuses.\(^{36}\) PFG STE NMR experiments were able to uncover a native state (19.6 Å), an intermediate state (23.6 Å), and denatured and unfolded states (33-35 Å).
1.3.3 Transverse ($T_2$) and Longitudinal ($T_1$) Relaxation Time

Relaxation is the process where coherence or non-equilibrium populations associated with nuclear spins return to their equilibrium state after an applied pulse in an NMR experiment. Relaxation drives the generated transverse magnetization to zero and the longitudinal magnetization to a steady state value. The process of transverse magnetization decaying to its equilibrium value of zero is called transverse or spin-spin relaxation. The time constant used to describe the rate of this decay is the transverse relaxation time ($T_2$). $T_2$ relaxation times can be measured using either a Hahn echo sequence ($90^\circ - \tau - 180^\circ - \tau$) or a Carr-Purcell-Meiboom-Gill experiment (discussed in the following section). In situations where inhomogeneous line broadening effects are small, the line width of a signal can also be related to $T_2$. Longitudinal relaxation, also known as spin-lattice relaxation, refers to the z-magnetization returning to its equilibrium value. The longitudinal relaxation time ($T_1$) is the rate constant for this process. $T_1$ relaxation times are often quantified using an inversion recovery sequence ($180^\circ - \tau - 90^\circ$). Transverse ($T_2$) and longitudinal ($T_1$) relaxation times are a reflection of molecular motions (e.g. bond librations, side chain rotations, and overall tumbling of the protein) occurring on very different time scales.

1.3.4 Carr-Purcell-Meiboom-Gill (CPMG) Relaxation Dispersion

Proteins are dynamic systems and can often exchange between different conformational states on a millisecond to microsecond time scale. These states are typically weakly populated and exist only transiently thus, are difficult to study. However, through the use of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments these minimally populated states can be characterized.

CPMG experiments are very sensitive to exchange processes occurring on the millisecond to microsecond time scale and can readily detect the presence of folding intermediate states. Figures 7, 8 and 9 show the basic overview of a CPMG relaxation dispersion experiment.
Figure 7 (Left). CPMG relaxation dispersion experiments are sensitive to exchange processes occurring between a ground state structure and a weakly populated excited state.\textsuperscript{41} © Reprinted from \textit{J. Biol. NMR}, \textbf{2008}, \textit{41}, 113-120 with permission from Springer.

Figure 8 (Left). The basic overview of a CPMG relaxation dispersion experiment. A) The effective transverse relaxation rate, $R_{2,\text{eff}}$, is measured as a function of $v_{\text{CPMG}}(1/2\tau)$ where $\tau$ is the delay time between refocusing pulses in the spin echo train. B) CPMG experiments can be used to determine the exchange rate between the two states ($k_{AB}$ and $k_{BA}$) as well as their relative populations ($p_A$ and $p_B$). C) The chemical shift difference, $\Delta\omega$, between the resonances of the two states can also be extracted.\textsuperscript{41} © Reprinted from \textit{J. Biol. NMR}, \textbf{2008}, \textit{41}, 113-120 with permission from Springer.

Figure 9 (Above). The pulse sequence for a CPMG experiment consists of an initial $90^\circ(x)$ pulse followed by a relaxation period $\tau$ and a subsequent spin echo train, $(\tau$-$180^\circ(y)$- $\tau)_n$, which can be refocused an arbitrary number of times depending on the fixed pulse sequence length, $T_{CP}$.

The pulse sequence for a CPMG experiment consists of an initial $90^\circ(x)$ pulse, which generates transverse magnetization, followed by a relaxation period, $\tau$, and a subsequent spin echo train ($\tau$-$180^\circ(y)$-$\tau)_n$. This spin echo train can be refocused an arbitrary number of times depending on the
overall pulse sequence length, $T_{CP}$, and the relaxation period time, $\tau$. These CPMG dispersion experiments monitor the transverse relaxation rate ($R_2$, the inverse of the transverse relaxation time $T_2$) as a function of the delay time, $\tau$, between refocusing pulses.$^{38,42}$ Longer delay times between refocusing pulses allows for a longer relaxation time thus, leads to decreased signal intensity in the case that there are conformational changes occurring on the respective time scale.$^{43}$ Monitoring peak intensities as a function of delay times between refocusing pulses allows for a relaxation dispersion curve, $R_2$ versus $\nu_{CPMG} (1/2\tau)$, to be obtained. From this dispersion curve, the chemical exchange rate, $k_{ex}$, between the two interconverting states and additional exchange parameters (populations of the two states, $p_A$ and $p_B$, and differences in chemical shift between the two states, $\Delta \omega$) can be obtained by fitting the curve to the equation shown below:

$$R_2 = R_2^0 + \frac{p_A p_B \Delta \omega^2}{k_{ex}} \left(1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right)\right)$$

Equation 3

The utility of the CPMG experiment is not limited to only obtaining dynamic information, but through the use of the extracted chemical shift difference between the two states, $\Delta \omega$, along with residual dipolar couplings and residual chemical shift anisotropies, the structure of weakly populated folding intermediates can be elucidated.$^{44}$ However, this type of analysis is generally applicable to situations where exchange is on a slow or intermediate timescale. The above expression is an approximation of a more general relaxation expression and is applicable to situations where $k_{ex}$ is much greater than $\Delta \omega$. Here, we use CPMG experiments to measure exchange processes associated with the folding network of CaM.

1.3.5 Spin Diffusion

Spin diffusion NMR experiments are often used to probe changes in compactness and inter-residue contacts within a protein.$^{45}$ The spin diffusion experiment is based on the nuclear Overhauser effect (NOE) which is a result of cross relaxation between two dipole-coupled spins in close proximity.$^{46}$ In a spin diffusion experiment, two spectra are collected, one where there is selectively applied saturation (on-resonance) in some region of interest containing protons, and one where there is no such saturation (off-resonance). The signal intensities from the saturated spectrum, $I_{SAT}$, and the unsaturated spectrum, $I_O$, are differenced. In the difference spectrum, $I_{DIFF} = I_O - I_{SAT}$, only signals which receive magnetization transfer, through spin diffusion, will
remain. Proteins consist of a network of $^1$H spins all of which are coupled to each other through a dipolar interaction. Saturation of aliphatic spins for example, typically irradiates a large number of hydrophobic residues clustered in the protein interior. By application of saturation pulse for a sufficient period of time (usually hundreds of milliseconds) it is possible to attenuate the vast majority of $^1$H resonances in the protein, at least in the absence of motion. Internal motions provide a different relaxation pathway and typically diminish the effect of spin diffusion. Thus, the extent to which magnetization is nulled by spin diffusion directly depends on the internal mobility or compactness of the protein. Figure 10 shows the basic scheme of the spin diffusion experiment.

**Figure 10.** General scheme of a spin diffusion experiment. Two spectra are collected where one of the two receives selective saturation. In this example, the receptor molecule is selectively saturated and then through spin diffusion, magnetization is transferred to the protons on the ligand. Differentiating the two spectra reveal protons which received magnetization. © Reprinted from *J. Chem. Educ.*, 2011. 88, 990-994 with permission from American Chemical Society.

By saturating the aliphatic interior of the protein and monitoring its effect on the backbone amide signal, changes in compactness of the protein can be visualized.

### 1.4 Circular Dichroism (CD)

In addition to NMR, circular dichroism (CD) is also a useful technique for studying the structural properties of proteins in solution. Circular dichroism is defined as the differential absorption of left-handed and right-handed circularly polarized light. An asymmetric molecule will absorb either left-handed or right-handed circularly polarized light to a greater extent to produce elliptically polarized light. The difference in absorption is typically reported in terms of an ellipticity value ($[\Theta]$ (deg cm$^2$·dmol$^{-1}$)). Well-folded proteins contain asymmetric secondary structural elements ($\alpha$-helices and $\beta$-sheets) which give rise to a characteristic CD signal (Figure 11).
Figure 11. Far-UV ellipticity profiles for various secondary structure elements (left).\textsuperscript{48} Near-UV ellipticity profile for type II dehydroquinase (right).\textsuperscript{49} Wavelength regions where phenylalanine, tyrosine, and tryptophan residues produce a CD signal are indicated. © Reprinted from Nat. Protoc. 2006. 1, 2527-2535 with permission from Nature Publishing Group. © Reprinted from Biochim. Biophys. Acta: Proteins and Proteomics. 2005. 1751, 119-139 with permission from Elsevier.

The overall secondary structure of a protein is monitored by measuring ellipticity in the far-UV region (190 to 240 nm). CD is useful for studying protein folding/unfolding pathways as conformational changes in secondary structures are clearly reflected by the CD signal in the far-UV regime.\textsuperscript{50} CD can also be used to acquire information on the tertiary structure of a protein.\textsuperscript{49} Each of the aromatic amino acid side chains (tyrosine, tryptophan, and phenylalanine) produces a characteristic profile in the near-UV region (260 to 320 nm) (Figure 11). In the case of calmodulin, since the protein contains eight phenylalanine residues, only two tyrosine residues, and no tryptophan residues, near-UV CD signal arises primarily from the phenylalanine residues located within the protein interior. Thus, CD measurements in this region would largely reflect the packing of the hydrophobic core.

1.5 Fluorine-19 as a Tool to Study Protein Structure and Dynamics

The usefulness of fluorine as a probe to study the structural and dynamic properties of a protein has been demonstrated in a number of different systems.\textsuperscript{51,52,53} Fluorine-19, which exists in 100% natural abundance, is a $\frac{1}{2}$ spin species making it amenable to NMR studies and is 83% as sensitive as $^1$H.\textsuperscript{54} In addition to the absence of fluorine signal in biological systems, $^{19}$F NMR
has several unique characteristics which give it an advantage over other nuclei (e.g. \(^{1}H, ^{13}C, ^{15}N\)). The \(^{19}F\) nucleus is an extremely sensitive NMR probe of local environment (e.g. electrostatic fields, van der Waals interactions) and hence, protein topology, through both chemical shifts and spin relaxation (\(T_1\) and \(T_2\)). Fluoroaromatics in particular, exhibit a significant range of relaxation rates and chemical shielding effects which depend sensitively on topology. \(^{19}F\) NMR also benefits from having a large chemical shift range which is hundreds of times larger than that of \(^{1}H\). The high sensitivity of the fluorine probe along with the absence of background signal is a significant advantage for both \textit{in vivo} and \textit{in vitro} protein studies.

1.5.1 Incorporation of Fluorine Labels

Fluorine probes are often incorporated into a protein biosynthetically or through the use of a fluorinated chemical modification reagent. Biosynthetic labeling involves adding a fluorinated amino acid analog to the growth medium followed by inducing protein expression. By also inhibiting the ability of the cell to synthesize that amino acid, it is possible to achieve high levels of labeling efficiency without drastically reducing protein yield. A disadvantage to this labeling method is the low specificity it provides. In this method, all sites of the given residue become fluorine labeled which can be problematic when dealing with large protein systems. In cases where a more specific labeling site is desired, a labile residue site on the protein can be chemically modified by using a fluorinated reagent. This labeling scheme often targets the thiol group on a cysteine due to its high nucleophilicity (Figure 12). Since there are typically low numbers of cysteines in a protein system, labeling specific locations becomes more manageable.

![Figure 12](image)

\textit{Figure 12.} A general reaction pathway used to attach a fluorine label to the thiol group of a cysteine. © Reprinted from \textit{Proc. Natl. Acad. Sci. U. S. A.}, \textbf{1999}, \textit{96}, 13744-13749 with permission from National Academy of Sciences, USA.
Though many fluorinated amino acids have been used in $^{19}$F NMR studies of proteins, mono-fluoroaromatics are most commonly used. Aromatic amino acids are typically situated in areas of high interest to structural protein studies (e.g. hydrophobic cores, dynamic loops). Thus, aromatic residues are the focus in many protein folding studies as they can often play a role in the formation and stabilization of folding intermediates. In our studies, we will biosynthetically incorporate 3-fluorophenylalanine (3F-Phe) into CaM, allowing changes in the interior of the protein to be monitored.

1.5.2 Biosynthetic Labeling of Calmodulin with 3-Fluorophenylalanine

Within CaM, there are eight phenylalanine residues which are all associated with the hydrophobic cores in their respective domains. There are five phenylalanines located in the N-terminal domain (F12, F16, F19, F65 and F68) and three phenylalanines in the C-terminal domain (F89, F92, and F141). Figure 13 indicates the location of eight phenylalanine residues within the two domains as well as their positions in the EF hand motifs. Previous research has shown the importance of these phenylalanine residues for stability of the hydrophobic core as well as for target binding and activation. Incorporating 3-fluorophenylalanine in CaM, biosynthetically, allows for a unique insight into the hydrophobic core of the protein. Figure 14 shows the 1D $^{19}$F NMR spectrum of 70% 3-fluorophenylalanine enriched CaM displaying the phenylalanine assignments for each peak.

![Figure 13](Left). X-ray structure of holo-CaM highlighting the location of the eight phenylalanine residues (left). Secondary structural map of the four EF hand motifs of calmodulin (right). The eight phenylalanine residues are highlighted. © Reprinted from J. Biol. NMR, 2010, 42, 113-121 with permission from Springer.
Figure 14. $^{19}$F NMR spectrum, collected at 50°C, of 70% 3F-Phe fractionally labeled CaM. A spectral deconvolution is shown to distinguish all eight fluorine resonances. Assignments of the eight observed resonances were previously established using $^{19}$F-edited $^1$H-$^1$H nuclear Overhauser effect spectroscopy (NOESY) experiments. 3F-Phe labeling of CaM is slightly unusual in that 100% labeling levels result in inhomogeneous line broadening and the appearance of minor conformer peaks (Figure 15). Based on the distorted $^{19}$F spectrum, 100% labeling levels were found to cause significant structural and dynamic perturbations to the protein.

Figure 15 (Left). $^{19}$F NMR spectra of 3F-Phe labeled CaM at various different labeling levels. 100% labeling levels was found to cause structural perturbations leading to the appearance of minor conformer peaks and line broadening. At 76% labeling levels, there are no detectable perturbations and sufficient fluorine signal. Reprinted from J. Biol. NMR, 2010. 42, 113-121 with permission from Springer.
However, structural perturbations from labeling could be minimized by lowering 3F-Phe enrichment levels to approximately 70%. Fractional labeling is achieved by supplementing the expression medium with the fluorinated amino acid analog as well as a fraction of the non-fluorinated amino acid. For our studies we will make use of 70% 3F-Phe enriched CaM.

1.5.3 Measuring Solvent Exposure Using Fluorine NMR

One of the most useful characteristics of fluorine as an NMR probe is its high sensitivity to the local environment. Fluorine chemical shifts are strongly affected by changes to the electrostatic field, van der Waals interactions, and hydrogen bonding. Based on this property, $^{19}$F probes can be used to assess solvent exposure through the use of solvent induced isotope shifts.

Solvent induced isotope shift experiments involve the substitution of D$_2$O for H$_2$O. Since fluorine is an extremely sensitive probe, a substitution of D$_2$O for H$_2$O is known to yield a chemical shift change whose magnitude directly correlates to the overall solvent exposure of the fluorine probe (Figure 16). Fluorine residues which are completely exposed to the solvent can experience chemical shifts as large as 0.25 ppm whereas the chemical shift of buried fluorine residues remain unaffected or show only a small shift.

![Figure 16](image)

**Figure 16.** Inset of the $^{19}$F NMR spectra of 3F-Phe labeled CaM in 10% D$_2$O (blue line) and 100% D$_2$O (red line). Upon substitution of D$_2$O for H$_2$O, the fluorine resonances experience a chemical shift change, $\Delta \delta_{19F}$, whose magnitude corresponds to the solvent accessibility of the specific site. Here, the most downfield peak at ~146.6 ppm experiences the smallest chemical shift change indicating that site is less solvent exposed than the two upfield peaks.

These solvent isotope shifts can be used to assess water penetration into the hydrophobic core along the CaM’s folding/unfolding pathway.
1.5.4 Examples of Fluorine NMR Applications with Proteins

$^{19}$F NMR spectroscopy is a widely used tool in studying the structure and dynamics of small soluble proteins. In a recent study, $^{19}$F NMR was used to study the dynamic equilibria between monomeric and oligomeric misfolded states of a mammalian prion protein (ShaPrP). By biosynthetically labeling the protein with 3F-Phe, the authors uncovered the coexistence of a monomer, an octamer, and a large oligomer state (Figure 17).

Figure 17. $^{19}$F NMR spectra of 3F-Phe labeled ShaPrP. A spectral deconvolution reveals the coexistence of a monomer, octomer, and large oligomeric states. © Reprinted from J. Am. Chem. Soc. 2013. 135, 10533-10541 with permission from American Chemical Society.

Experiments were focused on obtaining thermodynamic and kinetic parameters associated with the equilibrium between the three states. Thermodynamic analysis of the equilibrium revealed that formation of the octamer state is entropically driven while assembly of the large oligomer is enthalpically driven. The authors also investigated a Gerstmann–Straussler–Scheinker syndrome associated mutant version of the prion protein (F198S) and found that the mutation destabilized that native state while possibility increasing the susceptibility of the protein to form aggregates.

$^{19}$F NMR can also been applied to study larger systems such as G protein-coupled receptors (GPCRs). GPCRs, such as the beta-2 adrenergic receptor ($\beta_2$AR), can recognize and bind a number of different ligands. The chemical structure of the ligand determines its effect on the signaling activity of the GPCR. By covalently labeling one of the three accessible cysteines on $\beta_2$AR with 2,2,2-trifluoroethanethiol (TET), the effect of various classes of ligands
(agonist, neutral antagonist, and inverse agonist) on the conformational equilibria between various different states can be studied.

**Figure 18** (Left). $^{19}$F NMR spectra of TET labeled $\beta_2$AR$^{(\text{TET}C265, \text{C327S, C341A & C265A, TET}C327, \text{C341A})}$ bound to various ligands. A spectral deconvolution is shown to highlight the active (red) and inactive (blue) states. Changes in peak integrals were monitored upon binding of various ligands to examine the shifts in populations between active and inactive states.© Reprinted from *Science*, 2012, 335, 1106-1110 with permission from The American Association for the Advancement of Science.

Deconvolution of the $^{19}$F NMR spectra of mutants containing a single TET-labeled cysteine residue revealed the presence of inactive and active states (Figure 18). The authors monitor the $^{19}$F NMR signal arising from the active and inactive states in response to various ligands. The changes in peak integrals upon ligand binding allowed the population shift between active and inactive states to be quantified. Overall, the paper demonstrates the utility of $^{19}$F NMR as a tool for structure and dynamic studies of GPCRs.
Since $^{19}$F NMR benefits from the virtual absence of background signal, it is particularly useful for \textit{in vivo} experiments.$^{73}$ Figure 19 shows $^{19}$F NMR spectra of two small globular proteins, which are fluorine labeled, \textit{in vivo}.

\textbf{Figure 19.} $^{19}$F NMR spectra of 6-FTrp labeled B1 domain of protein G (GB1) and 3-FTyr labeled ubiquitin in \textit{E. coli} cells, lysate, and buffer.$^{73}$ © Reprinted from \textit{Chem. Eur. J}, \textbf{2013} with permission from John Wiley and Sons.

Even when in living cells, the fluorine probe on the proteins still produce a resolvable peak with sufficient signal and can be used to study the protein system. This would otherwise be impossible with $^1$H NMR due to the excessive amount of background proton signal in living cells. Here the authors used $^{19}$F NMR to probe cytoplasmic viscosity and protein-protein interactions in \textit{E. coli} cells.
2.1 Expression and Purification of $^{15}\text{N}$-Enriched and 70% 3-Fluorophenylalanine Labeled Calmodulin

Expression and purification of uniformly $^{15}\text{N}$-enriched, 3-fluorophenylalanine fractionally labeled CaM was performed as described previously with small modifications. A plasmid (pET21b) encoding Xenopus laevis calmodulin (residues 1-148) was transfected into BL21(DE3) competent cells. A 29 mL overnight culture of transformed E. coli BL21 (DE3) cells in LB was used to inoculate 1 L of M9 media (1 g/L $^{15}\text{NH}_4\text{Cl}$, 3 g/L D-glucose, 10 mg/L thiamine, 10 mg/L biotin, 100 mg/L ampicillin, 0.1 mM CaCl$_2$, 1 mM MgSO$_4$). The cell cultures were grown at 37°C with shaking until an OD$_{600}$ of 0.8 was reached. At this point, each 1 L 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). After 3 hours, cell cultures were harvested by centrifugation at 7000 rpm for 30 minutes at 4°C. Each 1 L pellet was 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. The cleared lysate was then separated from the cell debris and purified using Ni-NTA agarose resin (Qiagen, Mississauga, Ontario, Canada). 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 mix for 12 hours. The labeled protein was then purified using a phenyl sepharose column as described previously. Purified CaM NMR samples were stabilized in 20 mM Tris HCl, 100 mM KCl, 9 mM CaCl$_2$, H$_2$O/D$_2$O (90%:10%) or D$_2$O buffer adjusted to pH 8.0 and generally run either in the presence of equimolar 4-fluorophenylalanine or sodium fluoride as standards. Apo-CaM samples were prepared by dialyzing purified holo-CaM into 50 mM phosphate buffer adjusted to pH 7.4.
2.2 Isolating the Two Domains of Calmodulin (TR1C and TR2C)

Trypsin digestion of full length 3-fluorophenylalanine CaM was performed as previously described with several modifications. \cite{66, 75} 1 mM holo-CaM was dialyzed into 50 mM NH₄HCO₃, 50 mM NaCl, 5 mM CaCl₂ buffer adjusted to pH 7.9. Trypsin was added to a final concentration of 0.017 mM to the CaM sample followed by 1 hour incubation at 37°C. After the incubation period, 0.017 mM soybean trypsin inhibitor was added to stop the digestion reaction and the protein solution was left on ice for 15 minutes. Using a 50 cm sephadex G-50 column, the undigested protein was separated from the trypsin fragments. The protein was eluted with 50 mM NH₄HCO₃ and 50 mM NaCl buffer adjusted to pH 7.9. OD₂₈₀ readings show that the undigested protein and trypsin elutes first, followed by the two CaM fragments. Aliquots containing CaM trypsin fragments were pooled and separated using a phenyl sepharose column. The C-terminal fragment was first eluted using 2 mM Tris-HCl, 1 mM CaCl₂ buffer adjusted to pH 7.5 followed by elution of the N-terminal fragment using ddH₂O. Fragments were then stabilized in 20 mM Tris HCl, 100 mM KCl, 9 mM CaCl₂, H₂O/D₂O (90%:10%) buffer adjusted to pH 8.0.

2.3 Circular Dichroism Spectroscopy

CD spectra in the far-UV (200 nm to 250 nm) and near-UV (253 nm to 273 nm) regions were acquired on an Aviv CD spectrometer model 62DS. Spectra of 5 mg/mL 70% fractionally labeled 3F-Phe CaM, in 0.1 M KCl, 20 mM Tris HCl, and 9 mM CaCl₂ buffer adjusted to pH 8, were collected from 20°C to 100°C (5°C increments) (path length, 0.1 cm; steps, 0.05 nm; bandwidth, 1 nm; and averaging time 0.2 s). Secondary and tertiary structure thermal unfolding profiles were obtained by monitoring ellipticity at 222 nm (far UV) and averaging the ellipticity from 253 nm to 273 nm (near UV).

2.4 NMR Spectroscopy

NMR experiments were performed on a 500 MHz Varian Inova (¹⁵NH-detected aliphatic saturation spin diffusion), 600 MHz Varian Inova spectrometer (¹⁹F NMR solvent isotope shifts, ¹⁹F NMR O₂ shift experiments, ¹H stimulated echo diffusion NMR, ¹⁵N, ¹H spectroscopy, and ¹⁹F NMR CPMG studies), and a 700 MHz Agilent spectrometer (¹³C, ¹H SOFAST HMQC spectra) (Agilent Technologies, Santa Clara, CA), using a combination of room temperature and
cryogenic single gradient solution NMR probe, tunable to either $^1$H or $^{19}$F on the high frequency channel. All NMR data was processed using either NMRPipe$^{76}$ or Mestrenova software.

2.4.1 Solvent Induced Isotope Shifts

Solvent isotope shifts were evaluated by increasing the D$_2$O/H$_2$O fraction of the CaM sample from 10% to 100%. $^{19}$F 1D NMR experiments were collected with 8192 transients with a pulse width of 8.2 µs from 20-75°C in 5 degree increments. In our analysis, we define a normalized solvent isotope shift, $\Delta \delta^*_{H_2O}$, which is calculated by taking the relative difference between the chemical shifts of the peak of interest in D$_2$O and H$_2$O and dividing by the corresponding shift difference from an internal standard, 4-fluorophenylalanine, which is known to be fully exposed to the bulk solvent. Thus, the normalized shift is expressed as

$$\Delta \delta^*(H_2O) = \frac{(\delta_{D2O} - \delta_{H2O})}{(\delta_{4F-Phe} - \delta_{4F-Phe})}$$

2.4.2 O$_2$ Paramagnetic Shifts

Paramagnetic shifts from dissolved oxygen were obtained by equilibrating the protein sample at a partial pressure of 10 – 25 atm overnight in a 5 mm OD, 3 mm ID sapphire NMR sample tube (Saint Gobain – Saphikon Crystals, Milford, NH, USA). The pressure during the entire course of the NMR experiment was kept constant through the use of Swagelok (Swagelok, Solon, OH, USA) connections to a pressurized oxygen supply. $^{19}$F 1D spectra of ~1.5 mM CaM, in both the absence and presence of dissolved oxygen, were obtained as a function of temperature. As in the case of the solvent isotope shifts, we can define a normalized paramagnetic shift, $\Delta \delta^*_{O_2}$. This normalized shift is expressed as

$$\Delta \delta^*(O_2) = \frac{(\delta_{O2} - \delta_{O2})}{(\delta_{4F-Phe} - \delta_{4F-Phe})}$$

A van’t Hoff analysis of the temperature dependence of the partition coefficients of oxygen to the protein interior allowed for an estimate of the enthalpy and entropy of the partitioning of oxygen into the protein interior.$^{77}$

2.4.3 Pulsed Field Gradient Stimulated Echo

Diffusion measurements were performed using PFG STE HSQC.$^{78}$ Eighteen gradient amplitude experiments were acquired for each temperature, 28°C to 57°C (3°C increments), with gradient
strength varied from 6% to 92% of its maximum power. Translational diffusion coefficients were obtained from the slope of the natural logarithm of normalized peak intensity plotted against the corresponding square of the gradient strengths. Diffusion coefficients were converted to a hydrodynamic radius using the Stokes-Einstein equation and by normalizing to water (hydrodynamic radius = 1.4 Å) which was measured simultaneously.

2.4.4 Spin diffusion

Spin diffusion experiments were performed by selectively saturating the aliphatic resonances (~0.7 ppm) of the protein using a continuous wave saturation pulse with a B1 field of 5200 Hz, for a period of 1 sec. Detection of remaining signal was obtained via a 15N-1H HSQC and resulting peak intensities were then divided by the equivalent peaks obtained by an equivalent off-resonance (15 ppm) continuous wave excitation pulse.

2.4.5 15N-1H HSQC and 13C-1H HMQC Spectra

15N-1H HSQC spectra were collected via a cryogenic NMR probe, at a 1H Larmor frequency of 600 MHz, using 8 scans and 80 increments spanning 2000 Hz in the indirect dimension. 13C-1H SOFAST methyl HMQC spectra were acquired using a 700 MHz cryogenic probe with a repetition time of 0.5 sec, 96 increments, 128 transients, and a spectral width of 3520 Hz in the indirect dimension.

2.4.6 CPMG Relaxation Dispersion Experiments

The measurement of 19F transverse relaxation rates and relaxation dispersions was accomplished by a 1D-CPMG sequence (90 - [τcp - 180 - τcp]N) using a total of 12 τcp values, ranging from 150 µs to 6 ms, with a constant time delay, Tcpmg, of 12 ms (full length CaM) and 20 ms (TR1C and TR2C CaM). Peak intensities, I, were extracted and used to calculate effective transverse relaxation rates, \( R_{2,\text{eff}} = (-1/T_{\text{CPMG}})(\ln(I/I_o)) \), for the individual fluorine resonances where I_o is the corresponding peak intensity in the absence of a constant time delay. In the fast exchange limit, assuming the system can be approximated by two states (i.e. native and intermediate, N and I) separated by a chemical shift, \( \Delta \omega \), the exchange contribution to the transverse spin relaxation rate, is given by

\[
R_2 = R_2^0 + \frac{p_a p_b \Delta \omega^2}{k_{ex}} \left(1 - \frac{4v_{\text{CPMG}}}{k_{ex}} \tanh \left( \frac{k_{ex}}{4v_{\text{CPMG}}} \right) \right)
\]  

Equation 4

\[
k_{ex} = \frac{k_{IN}}{p_1} = \frac{k_{NI}}{p_2}
\]  

Equation 5
where $k_{ex}$ can be expressed in terms of the folding and unfolding rates (i.e. $k_{IN}$ and $k_{NI}$ respectively) and the fraction of each of the two states, $p_1$ and $p_2$. Chemical exchange rates, $k_{ex}$, were obtained by fitting relaxation dispersion curves ($R_{2,eff}$ vs $v_{CPMG}$ $(1/2\tau_{cp})$) to a fast exchanging two state model, Equation 4, using the open-source software package GUARDD. $^{80,81}$

2.4.7 $^{19}$F Transverse ($T_2$) and Longitudinal ($T_1$) Relaxation Measurements

The measurement of $^{19}$F longitudinal ($T_1$) relaxation times was accomplished using an inversion recovery sequence (180° - $\tau$ - 90°) using a total of 10 $\tau$ values, ranging from 1 ms to 0.7 seconds, and a repetition time of 4 seconds. $T_1$'s were calculated by fitting the intensity versus delay time plot to an exponential decay, $I = I_o \exp(-\tau/T_1)$. $^{19}$F transverse ($T_2$) relaxation times were calculated from the high refocusing CPMG experiment ($v_{CPMG} = 3350$ Hz). Using the ‘model-free’ approach given by Lipari and Szabo, $^{82}$ an order parameter value, $S_2$, was determined using $T_1$ and $T_2$ relaxation times.

2.5 Estimating Populations of the Native and Near-Native States

Using the temperature dependence of the $^{19}$F chemical shifts, an estimate of the relative populations of the native and near-native states was obtained. In our analysis, both states are assumed to exhibit a linear temperature dependence of chemical shift given by a slope, $m$, and a $y$-intercept, $\delta_o$. The native state can be denoted as $p_1(T)$ and the near-native intermediate as $p_2(T)$, where $p_1(T)+p_2(T)=1$. Further assuming the two states are in fast exchange, the observed shift, $\delta_{obs}$, is a weighted average of those associated with the native and near-native states,

$$\delta_{obs} = p_1\delta_1(T)+p_2\delta_2(T) = (1-p_2)(\delta_{1o}+m_1T)+p_2(\delta_{2o}+m_2T)$$  \hspace{1cm} \text{Equation 6}

The populations are assumed to be governed by a Sigmoidal function defining a transition temperature, $T_{NI}$, and a transition width, $\Delta$, such that $^{83}$

$$p_2 = \frac{1}{1 + \exp[-(T-T_{NI})/\Delta]}$$  \hspace{1cm} \text{Equation 7}
$T_{NI}$ and $\Delta$ were obtained via a non-linear least squares fit of the temperature dependence of the $^{19}$F chemical shifts of residues F12, F16, F19, and F92 to Equations 6 and 7, using, GNUPLOT (http://www.gnuplot.info/documentation.html). Fitted $T_{NI}$’s and $\Delta$’s from each of the residues were averaged to obtain a global transition temperature and transition width.
Chapter 3
Results

1D spectra, solvent isotope shifts, and oxygen paramagnetic shifts were obtained by Julianne Kitevski-Leblanc.
Circular dichroism profiles of 3-FPhe CaM were obtained with help from William Thach.
Spin diffusion data was obtained by Sacha Larda.

3.1 Circular Dichroism and $^{19}$F NMR Evidence of a Folding Intermediate along a Heat Denaturation Pathway

While folding events under non-denaturing conditions often occur on a microsecond timescale, and are thus, difficult to capture by NMR, it is possible to spectroscopically discriminate a hierarchy of events along a temperature denaturation pathway. In this study, we identify an intermediate by both heat denaturation and at lower temperatures, by CPMG relaxation dispersion measurements as discussed below. To begin, far-UV CD and $^{15}$N-$^1$H HSQC NMR spectra of CaM reveal that secondary structure is largely retained over the entire range of temperatures studied (25-75°C). The far-UV CD traces (200 to 250 nm) all exhibit the characteristic double minima, representative of $\alpha$-helical structure (Figure 20).

Figure 20. Far-UV CD ellipticity profiles (200 to 250 nm) of 3F-Phe CaM at 35°C and 70°C. Even at elevated temperatures, a strong CD signal is observed, signifying that the protein retains a large amount of secondary structure. As we increase temperature we see a progressive loss in secondary structure but even at 70°C, a significant amount of secondary structure is still retained. $^{15}$N-$^1$H HSQC spectra of the protein backbone tell largely the same story; the chemical shift dispersion over the range of temperatures...
is consistent with a single state whose secondary structure and overall topology is very near that of the native state (Figure 21).

**Figure 21.** $^{15}$N-$^1$H HSQC of 3F-Phe CaM as a function of temperature (left). Note that the boxed region is expanded on the right. The linear dependence of chemical shift, highlighted by the arrows, and line widths in the HSQC spectra are consistent with a single folded state over the temperature range examined. $^{65}$ $^{15}$N-$^1$H HSQC spectra show that approximately 30% of residues generally retain their native chemical shifts while the remaining residues experience a slight linear temperature dependence of chemical shift. However, the near-UV CD behavior, which is sensitive to tertiary structure and compactness, $^{49}$ reveals the presence of an intermediate state along the temperature denaturation pathway. Since CaM contains only two tyrosine residues and no tryptophan residues, the near-UV CD signature from 253 nm to 275 nm largely reflects the packing environment in the vicinity of the phenylalanines (Figure 22).

**Figure 22.** Near-UV CD profiles (253 to 275 nm) as a function of temperature for 3F-Phe CaM. Near-UV CD, which largely reflects tertiary structure and compactness, shows that with increasing temperature, the tertiary structure of the protein is moderately affected but still retains a distinguishable profile even at ~80°C. $^{65}$. 

...
As the approximate CD profile is also reproduced at all temperatures, we consider the average ellipticity in this wavelength regime, \(<\Theta>_{253-275}\), and plot this as a function of temperature (Figure 23).

**Figure 23.** Changes in the degree of tertiary structure of 3F-Phe CaM as a function of temperature as monitored by the near-UV ellipticity signature, \(<\Theta>_{253-275}\). As highlighted, a plateau in the degree of tertiary structure seen from 65–80°C signifies the onset of an intermediate state. At higher temperatures, beyond 80°C, the protein gradually acquires a more unfolded state.65

Generally, the average ellipticity gradually increases (toward zero) with increased temperature until 60°C, pointing to an overall loss of tertiary structure and compactness of the hydrophobic core. Between 60°C and 75°C, the ellipticity plateaus, which we associate with a folding intermediate. Above this regime, ellipticity again increases towards zero with temperature, signifying further loss of structure and compactness. Based upon near- and far-UV CD and \(^{15}\)N-\(^{1}\)H spectra, this folding intermediate, formed at ~60°C, is best represented by having a near-native conformation.

Of course, the possibility that this observed intermediate state only arises as a result of the incorporated fluorine atoms, which can possibly perturb the protein structure, must be addressed. Thus, near-UV CD profiles were collected for non-fluorinated CaM and the average
ellipticity in the wavelength regime, 253 to 275 nm, was monitored. Figure 24 reveals that a very similar result is observed for both non-fluorinated CaM as well as fluorinated CaM.

![Figure 24](image)

**Figure 24.** Near-UV ellipticity signature $<\Theta>_{253-275}$ of unlabeled CaM as a function of temperature. The plateau in ellipticity near ~55°C signifies the onset of an intermediate state. In the case of both unlabeled and 3F-Phe CaM, near-UV CD profiles with temperature hint at the presence of an intermediate state.

Comparing the two near-UV CD profiles concludes that this observed intermediate is not created as a result the incorporated fluorine atoms.

While near-UV CD gives a hint of a folding intermediate, $^{19}$F NMR provides a wealth of data regarding the equilibrium of folding states and their microscopic properties. As shown by the $^{19}$F spectra (Figure 25), a single Lorentzian line is observed for each labeled site at any one temperature, signifying either a single state at low temperatures or fast exchange between a native and near-native states at higher temperatures (~40-70°C). The existence of a single Lorentzian, arising from each 3-fluorophenylalanine species, also signifies rapid ring flipping (on a chemical shift scale) as has been noted previously, and underlies the extent of side chain dynamics inherent in the hydrophobic protein interior.
The temperature dependence of the line widths of the $^{19}$F peaks again also reveals the presence of an intermediate state (Figure 26). Line widths are good reporters of subtle changes in local dynamics and can reflect a distribution of unique sites and/or motions on a millisecond to microsecond timescale.\textsuperscript{87}

\textbf{Figure 25.} $^{19}$F NMR spectra of 70\% 3F-Phe fractionally labeled CaM as a function of temperature. The $^{19}$F NMR spectra exhibit the onset of line broadening near 50°C, indicative of an exchange process occurring. Above 70°C, very weak and broad $^{19}$F resonances arise, between -113.3 and -113.7 ppm, as shown above by the peaks denoted with an asterisk. This may reflect the onset of an unfolded or aggregated state.\textsuperscript{65}

\textbf{Figure 26.} $^{19}$F NMR spectra peak line widths of 70\% 3F-Phe fractionally labeled CaM as a function of temperature. Line widths progressively narrow from 20°C to 60°C. Above 60°C, line widths begin to broaden.
The temperature dependence of the $^{19}$F NMR spectra line widths show from 20°C to 60°C, the line widths progressively narrow which is consistent that the protein contains a single folded state. However, at 60°C a transition occurs where the line widths begin to broaden. The onset line width broadening at 60°C is suggestive of the presence of a folding intermediate.

The temperature dependence of the chemical shifts, $\Delta\delta_{^{19}F}$, of the dominant peaks is also consistent with a formation of folding intermediate near 60°C, based on a prominent non-linear dependence of chemical shift with temperature (Figure 27).

![Figure 27](image.png)

**Figure 27.** A temperature-dependent plot of the change in the $^{19}$F chemical shift ($\Delta\delta_{^{19}F}$), relative to the shift at 20°C, associated with each of the eight 3F-Phe residues.$^{65}$

While all eight resonances exhibit, to some degree, a non-linear dependence of shift with temperature, four resonances, originating from residues F12, F16, F19, and F92, were found to exhibit a significant non-linear dependence, sufficient to reliably extract the relative populations and estimates for a transition temperature, $T_{NI}$, and a transition width, $\Delta$. In this analysis, we assume that in the temperature range of interest (25-75°C) the protein is represented by a native state, whose population is given by $p_1(T)$, and by a near-native intermediate, whose population is given by $p_2(T)$, where $p_1(T)+p_2(T)=1$. We further assume that both the native states and near-native states alone are characterized by a linear dependence of chemical shift with temperature. Since the two states are expected to be in fast exchange, the observed chemical shift, $\delta_{\text{obs}}$, is a weighted average of those associated with the native and near-native states can be given by

$$\delta_{\text{obs}} = p_1\delta_1(T) + p_2\delta_2(T)$$
The populations are assumed to governed by a sigmoidal function defining $T_{NI}$ and $\Delta$ such that

$$p_2(T) = \frac{1}{1 + \exp\left(-\frac{(T - T_{NI})}{\Delta}\right)}$$

Figure 28 shows the fits of the imposed two-state model to the $^{19}$F NMR chemical shift data used to extract $T_{NI}$ and $\Delta$.

**Figure 28.** Fits of $^{19}$F chemical shifts, $\Delta\delta_{10F}$, to a two-state model in which the chemical shift of each state is assumed to depend linearly on temperature and the relative population of each state is expressed in terms of Equation 6. The transition from a native to near-native state is thus parameterized by a transition temperature, $T_{NI}$, and a width, $\Delta$. Four of the $^{19}$F resonances exhibit a pronounced inflection with temperature and can thus be used to estimate $T_{NI}$, and a width, $\Delta$. The corresponding fits and errors are shown in the figure. The overall average is given by estimate $T_{NI} = 66 \pm 15^\circ$C and a transition width, $\Delta$, of $10 \pm 7^\circ$C. The red and blue lines represent the linear dependence of chemical shift of the native and intermediate states with temperature respectively.

The final analysis produces a consistent description of the transition temperature, $T_{NI}$ ($66 \pm 15^\circ$C), and transition width, $\Delta$ ($10 \pm 7^\circ$C). Figure 29 provides a graphic illustration of the
estimates for the relative fraction of the near-native intermediate state as a function of temperature, based upon the fitted values of $T_{NI}$ and $\Delta$ for each $^{19}$F resonance used in the analysis.

Figure 29. Estimates of the temperature dependence of the relative fraction of the near-native intermediate state, $p_2(T)$, based upon individual fits of shift versus temperature for residues 12 (purple), 16 (blue), 19 (green), and 92 (orange). $p_2(T)$ is parameterized exclusively by the transition temperature and width ($T_{NI}$ and $\Delta$). An average transition temperature ($66 \pm 15^\circ C$) and width ($10 \pm 7^\circ C$) is also estimated and the resulting profile for $p_2(T)$ is represented by a solid line (red).\textsuperscript{65}

Though this analysis assumes the presence of only two states, it is likely that as we approach higher temperatures (~70°C), we also begin to weakly populate a fully unfolded state. For simplicity of this analysis we assume a two-state model, but it should be noted that there may be a small population of unfolded protein. Nevertheless, the two-state model analysis reveals that as we increase in temperature, the relative fraction of the near-native intermediate to the native state increases and points to a transition temperature of $66 \pm 15^\circ C$.

3.2 Solvent Exposure and Hydrophobicity along the Temperature Denaturation Pathway

From the temperature dependence of the $^{19}$F NMR 1D spectra, $^{15}$N-$^1$H HSQC spectra, and CD experiments it is clear that the folding network of CaM is more complex than a simple two-state model. We next consider experiments to monitor solvent exposure and hydrophobicity changes along the folding pathway. Specifically, we study the extent at which water penetrates the hydrophobic core and the changes in the hydrophobicity of the protein core over the temperature regime where the transition to the intermediate state was observed.
The presence of water within the hydrophobic core of the protein can be readily detected by the $^{19}$F phenylalanine probes through solvent induced isotope shifts. When exchanging the solvent from predominately water, H$_2$O, to 100% deuterium oxide, D$_2$O, fluorine resonances are known to experience a downfield shift, which may be as much as 0.25 ppm, as a result of deshielding of the signal from the water-soluble fluorine probe. The magnitude of the downfield shift reflects the degree of solvent exposure of the fluorine atom as this phenomenon requires the interaction between the hydrogen atoms of the solvent and the fluorine nucleus. Here we define a normalized solvent isotope shift, $\Delta \delta^*_{H2O}$, which is calculated by taking the relative difference between the chemical shifts of the peak of interest in D$_2$O and H$_2$O and dividing by the corresponding shift difference from an internal standard, 4-fluorophenylalanine, which is known to be fully exposed to the bulk solvent. Overall, this normalized shift can be expressed as:

$$\Delta \delta^*_{H2O} = (\delta^{D2O} - \delta^{H2O})/(\delta^{4-FPhe}_{D2O} - \delta^{4-FPhe}_{H2O})$$

Figure 30 shows the normalized solvent isotope shifts, $\Delta \delta^*_{H2O}$, as a function of temperature.

Figure 30. Temperature dependence of $^{19}$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. At 30°C, solvent isotope shifts of the majority of fluorophenylalanine probes (F16, F89, F68, F65, F19, F92) range between 25% and 55% of the fully exposed standard, suggesting significant accessibility of water. Note that while solvent isotope shifts are typically positive as defined, two residues exhibit negative solvent isotope shifts. A slight change in local structure around F141 and F12, caused by slight conformational perturbations resulting from substitution...
of H$_2$O for D$_2$O, may result in the solvent isotope shifts for these two residues to no longer accurately reflect solvent exposure. Water accessibility measurements require taking the relative difference between the chemical shifts of the peak of interest in D$_2$O and H$_2$O. In the case where a perturbation arises upon the substitution of D$_2$O for H$_2$O, the change in the chemical shift of the peak of interest in D$_2$O is a result of both the contact between the solvent and the probe and also the change in the probes local structural environment. Since the difference between the chemical shifts of the peak of interest in D$_2$O and H$_2$O cannot be solely attributed to the solvent exposure of the probe, the normalized solvent isotope shift is no longer an accurate reflection of solvent exposure. Nevertheless, as we increase the temperature from 30°C to 60°C, the decreasing normalized solvent isotope shifts indicate decreased solvent exposure of the phenylalanine fluorine probes. We therefore conclude that in the native state, water is gradually expelled from the hydrophobic core, driven at least in part by the relative gain in entropy of bulk water, upon heating. However, this trend is reversed near 65°C, the same temperature where the transition to an intermediate state was previously noted. At 65°C, the normalized solvent isotope shifts begin to increase, indicating hydration of the hydrophobic core. Thus, the near-native intermediate, formed near 65°C, is regarded as a desolvated state, relative to the native state. Since the eight individual solvent isotope shift profiles exhibit a similar trend, this reentry of water likely represents a cooperative process.

While solvent isotope shifts can be used to assess water penetration into the hydrophobic core, the measurement of paramagnetic shifts using dissolved oxygen provides a perspective of local density and hydrophobicity in the protein interior. Interaction of a paramagnetic species with a $^{19}$F nucleus induces a measureable chemical shift change in the fluorine resonance. The mechanism by which the paramagnetic species causes an observed shift effect is due to a Fermi contact interaction (a non-dipolar interaction between an electron spin and a nucleus). Here we introduce molecular oxygen to the protein sample by equilibrating it at a partial pressure of 10-25 atm. O$_2$ then partitions into the hydrophobic core, establishing concentration gradients, where a local O$_2$ concentration can be ascertained based on the magnitude of the paramagnetic shift. Generally, such shifts depend linearly on the local O$_2$ concentration, below pressures of 50 bar. Local O$_2$ concentrations are proportional to the paramagnetic shift, $\Delta\delta_{O_2}$, and as in the case of the solvent isotope shifts we define a normalized paramagnetic shift, $\Delta\delta^*_{O_2}$. This normalized shift can be expressed as:
\[ \Delta \delta^*_{O_2} = (\delta_{O_2} - \delta_{H_2O})/(\delta_{4^F\text{Phe}}^{O_2} - \delta_{4^F\text{Phe}}^{H_2O}) \]

Figure 31 shows the normalized paramagnetic shifts, \(\Delta \delta^*(O_2)\), due to dissolved oxygen, as a function of temperature.

**Figure 31.** 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.\(^{65}\)

Paramagnetic shifts exhibit a prominent increase with temperature, to the point where \(\Delta \delta^*(O_2)\), is between 1 and 18 times that observed with the 4F-Phe standard at a temperature nearly coinciding with \(T_{NI}\) associated with onset of the near native intermediate. This is in sharp contrast to the expected norm for compact proteins, where the average oxygen concentration in the interior of the protein is known to be less than that in water, while that in the interior of a micelle is an order of magnitude higher than water.\(^{77,93}\) Oxygen accessibility is influenced by both hydrophobicity and local density. As temperature increases and protein fluctuations become more prevalent, local void volumes are created in the protein interior, which facilitates solubilization of oxygen. This increased partitioning of oxygen to the protein interior with temperature is reversed above 65°C, where oxygen solubility is seen to decrease and the interior becomes hydrated. A van’t Hoff analysis of the temperature dependence of the partition coefficients of oxygen to the protein interior, \(K_p = [O_2]_{in}/[O_2]_{out}\), between 25°C and 65°C, reveals that partitioning of oxygen to the hydrophobic core is strongly entropically driven (\(\Delta S = 88.1\) J K\(^{-1}\) mol\(^{-1}\), Table 1), and remarkably similar to that measured recently in micelles.\(^{77}\) The results
suggest that the environment of the protein interior is disordered and more liquid alkane-like near $T_{NI}$ (i.e. highly dynamic and hydrophobic).

**Table 1.** Experimentally determined enthalpies and entropies of partitioning of oxygen in the vicinity of the eight 3-FPhe reporters.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Enthalpy (kJ mol$^{-1}$)</th>
<th>Entropy (J K$^{-1}$ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F12</td>
<td>16.5</td>
<td>67.0</td>
</tr>
<tr>
<td>F16</td>
<td>36.1</td>
<td>111.0</td>
</tr>
<tr>
<td>F89</td>
<td>35.2</td>
<td>108.5</td>
</tr>
<tr>
<td>F141</td>
<td>13.8</td>
<td>63.9</td>
</tr>
<tr>
<td>F68</td>
<td>25.7</td>
<td>89.6</td>
</tr>
<tr>
<td>F65</td>
<td>33.4</td>
<td>107.5</td>
</tr>
<tr>
<td>F19</td>
<td>20.0</td>
<td>79.7</td>
</tr>
<tr>
<td>F92</td>
<td>19.3</td>
<td>77.7</td>
</tr>
<tr>
<td>Global Average</td>
<td>25.0</td>
<td>88.1</td>
</tr>
</tbody>
</table>

Although it is well known that hydrophobic forces play a greater role at higher temperatures, the significant role of entropy and the similarity from the perspective of access by O$_2$, is consistent with previous descriptions of hydrophobic forces in the protein interior.$^{17}$

The solvent isotope shifts and paramagnetic shifts, which measure local oxygen accessibility, may be combined in a quotient, $\Delta \delta^*$(O$_2$)/$\Delta \delta^*$(H$_2$O), reflecting the local hydrophobicity. Most definitions of hydrophobicity, incorporate a partition coefficient, associated with the transfer of a molecule of interest from a polar to a non-polar environment.$^{94}$ In our case, we make use of the different chemical potentials of water and O$_2$ to study hydrophobicity of a point of interest in the protein interior. Effectively, we measure the ratio of partition coefficients associated with H$_2$O and O$_2$ between water and the protein interior. Assuming the oxygen paramagnetic shifts and solvent isotope shifts are proportional to the respective probe concentrations, the experimentally determined quotient, $\Delta \delta^*$(O$_2$)/$\Delta \delta^*$(H$_2$O), will equal the above hydrophobicity quotient, as long as the geometric considerations of accessibility of water and O$_2$ are regarded as equal.$^{68,69,95}$ Thus, we define a hydrophobicity parameter at a specific site, $i$, in terms of the respective ratio of concentrations of probe species in the bulk water phase and the protein interior

$$\text{Hydrophobicity} = \Delta \delta_{O2}^* / \Delta \delta_{H2O}^* = (\delta_{O2} - \delta_{H2O}) / (\delta_{D2O} - \delta_{H2O})$$

Figure 32 depicts the hydrophobicity quotient, $\Delta \delta^*$(O$_2$)/$\Delta \delta^*$(H$_2$O), as a function of temperature.
Figure 32. Hydrophobicity as measured by the relative partitioning of oxygen and water, described through a ratio of normalized oxygen-induced paramagnetic shifts to normalized solvent isotope shifts. Note that the solvent isotope shifts of F12 and F141 have been renormalized to values representative of the average solvent exposure of the remaining six residues ($\Delta \delta^*(H_2O) (30^\circ C) = 0.3$).

Hydrophobicity thus exhibits a maximum at the temperature associated with the transition to the folding intermediate, whereupon increased temperature results in gradual unfolding and loss of hydrophobicity.

3.3 NMR Studies of Compactness of the Native State and the Desolvated Near-Native Intermediate

A number of additional NMR experiments can be used to further characterize the identified intermediate state: 1) Spin diffusion from the aliphatic side chains to the protein backbone amides, 2) Diffusion measurements which monitor the protein hydrodynamic radius, and 3) $^{13}$C-$^1$H Heteronuclear multiple quantum coherence (HMQC) experiments focused on the methyls located in the hydrophobic core of the protein.

As noted in the introduction, compactness of a protein can be monitored by using spin diffusion NMR experiments. Here we selectively saturate the aliphatic interior of the protein and monitor the effect on the $^{15}$N-$^1$H HSQC spectrum of CaM with increasing temperature. The intensity of the majority of amide $^1$H signals, upon saturating the aliphatic region of the spectrum, generally decreases. With increased temperature, this effect is diminished, signifying that the aliphatic side chain isomerizations within the hydrophobic core are becoming less
restricted (Figures 33 and 34). The gradual loss of rigidity of the hydrophobic core results in less efficient spin diffusion.

**Figure 33.** $^{15}$N-$^1$H HSQC spectrum of CaM with selective saturation (red) overlaid with another $^{15}$N-$^1$H HSQC spectrum not containing selective saturation (black) at 30°C (left) and 60°C (right).

**Figure 34.** Spin-diffusion measurements along the temperature denaturation pathway. Amide intensities from an $^{15}$N-$^1$H HSQC are first normalized by dividing the observed intensity, after selectively saturating the aliphatic region for a period of 1 second, with the intensity resulting from a saturation pulse far off-resonance. A low normalized intensity reflects a high rigidity factor. The majority of backbone amides exhibit a very similar trend in loss of rigidity with temperature, toward the onset of the near native intermediate, estimated at 66°C. The red dotted line shows a fit associated with residues expected to be most affected by spin diffusion based on their position within the hydrophobic core. The black dotted line shows a fit associated with residues which are expected to be less affected by spin diffusion. $^{65}$
The hydrodynamic radius of a protein can be measured using PFE STE experiments. Hydrodynamic radius measurements are particularly useful as they can be used to distinguish whether the protein is indeed in a near-native state as opposed to an aggregated or unfolded state.\textsuperscript{36,73} Unfolded proteins are substantially more expanded than their native counterpart whereas folding intermediates are only slightly more expanded.\textsuperscript{33} NMR-based diffusion measurements of CaM reveal an increase in the hydrodynamic radius with increasing temperature, followed by a modest plateau near the temperature regime where the folding intermediate exists (Figure 35).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure35.png}
\caption{Temperature-dependence of the hydrodynamic radius, based on PFE STE experiments of 70\% enriched CaM. Diffusion measurements were normalized with respect to that of water.\textsuperscript{65}}
\end{figure}

At 28°C, where the protein exists predominately in a native state, a hydrodynamic radius of 27.7 ± 1.5 Å was measured. This result is coincides with CaM’s hydrodynamic radius literature values (~30 Å) obtained through dynamic light scattering experiments.\textsuperscript{96} At 57°C, near the transition temperature, the protein adopts only a slightly larger hydrodynamic radius of 34.4 ± 1.3 Å which rules out the possibility that the observed intermediate is an aggregated or unfolded state.

\textsuperscript{13}C-\textsuperscript{1}H HMQC experiments, focused on methyls in the hydrophobic core, are useful in revealing changes in structure and dynamics in a protein.\textsuperscript{97,98} Natural abundance \textsuperscript{13}C-\textsuperscript{1}H HMQC methyl spectra of enriched CaM corroborate the idea that the hydrophobic core is more dynamic, showing a clear convergence of methyl resonances with temperature (Figure 36).
At 37°C the methyl resonances, most of which arise from the hydrophobic core, are well resolved and exhibit a prominent dispersion. This dispersion is diminished at 65°C signifying a more dynamic hydrophobic interior.

Taken together, these experiments suggest that the desolvated near-native intermediate adopts a more dynamic state with an expanded hydrophobic interior.

3.4 Folding Kinetics between the Near-Native Intermediate and Folded State

While the $^{19}$F NMR solvent isotope shifts and O$_2$ paramagnetic shifts identify a folding intermediate and give some sense of compactness and the role of water in this near native state,
we next consider experiments which identify exchange rates between native and near-native states. Proteins are not static entities and often rearrange themselves into various conformations. One might ask if the observed transition between the native and near-native state can be detected dynamically (Figure 37).

**Figure 37.** Graphical representation of the protein folding free energy landscape for CaM. From the initial CD and $^{19}$F NMR experiments, we identified that along the folding/unfolding pathway of CaM, the protein adopts a desolvated near-native intermediate state. CPMG experiments allow fluctuations between the native and near-native states to be monitored.

CPMG-based relaxation dispersion NMR spectroscopy provides a quantitative measure of millisecond timescale fluctuations taking place between a highly populated, ground state, and a weakly populated, excited state. As stated in the introduction, CPMG experiments monitor transverse magnetization relaxation rates, $R_2$, as a function of separation between 180° refocusing pulses, $\tau_{cp}$. Figure 38 shows the effect of varying $v_{CPMG}$ ($1/2\tau_{cp}$) values on the intensity of the $^{19}$F spectrum.
Figure 38. $^{19}$F spectra of 3F-Phe CaM at various $\nu_{\text{CPMG}} (1/2\tau_{\text{cp}})$ values (3350 Hz (red), 850 Hz (black) and 85 Hz (blue)). The decrease in spectral intensity at lower $\nu_{\text{CPMG}}$ values is indicative of chemical exchange occurring on the millisecond time scale. Peak intensities of residues F68, F19, and F92 were extracted and used analyze exchange rates.

As $\nu_{\text{CPMG}}$ decreases and the CPMG delay time, $\tau_{\text{cp}}$, is increased, we see a gradual decrease in signal intensity which is evidence of conformational exchange processes occurring on a millisecond timescale. While all residues exhibit dispersions, those of residues F19, F68, and F92, are most pronounced and most easily resolved. Peak intensities can then be extracted and used to calculate a transverse relaxation rate ($R_2$). Figure 39 shows relaxation dispersion profiles, $R_2$ vs $\nu_{\text{CPMG}}(1/2\tau_{\text{cp}})$, for F68, F19, and F92.
Fitting these dispersion profiles to a fast two-site exchange model, Equation 4, reveals that at 57°C CaM undergoes two-site exchange on a timescale of 2400 Hz. At the same time, CPMG relaxation dispersion measurements, based on $^{15}$N transverse relaxation, reveal no millisecond timescale dynamics, from which we conclude that the conformational exchange involves a subtle change, associated primarily with the hydrophobic interior of the protein and not the protein backbone.

Two results connect the observed fluctuations to the transition between the native and near-native states: 1) The addition of the cosolvent, trifluoroethanol (TFE), which is known to stabilize the folded state, completely removes the observed dispersion (Figure 40), and 2) The dispersions can only be observed between 40°C and 65°C, coincident with the temperature regime in which the near native state is believed to be populated.
Figure 40. $^{19}$F CPMG relaxation dispersion profiles for 3F-Phe CaM with 5% TFE at 51°C. The lack of dispersion here indicates the addition of 5% TFE prevents excursions to the near-native intermediate state.

While CPMG experiments provide a measure of dynamics occurring on the millisecond timescale, a Lipari Szabo model-free analysis can be used to characterize faster motions on a nanosecond to picosecond timescale. Side-chain motions on this timescale can be interpreted in the form of a Lipari and Szabo order parameter value, $S^2$, which is derived from NMR relaxation data ($T_1$ and $T_2$). The amplitude of internal local reorientations can be described by the order parameter, $S^2$, which ranges from 1, for a restricted motion, to 0, for a more flexible motion. The order parameters for F68, F19, and F92 display a clear trend with temperature as shown in Figure 41.

Figure 41. Order parameter values, $S^2$, for 3F-Phe CaM as a function of temperature. The ‘model-free’ approach, given by Lipari and Szabo, was used to calculate order parameters. The progressive decrease of the order parameter value with temperature indicates motions of the phenylalanine side chains become less restricted at elevated temperatures.
At 45°C, the order parameter for the three residues ranged from 0.25 to 0.31. These order parameter values correlate well to those previously measured by methionine residues in CaM.\textsuperscript{99} Though order parameters for methyl groups in hydrophobic side chains can often range from 1 to as low as 0.1,\textsuperscript{100} order parameters in the range $0.1 \leq S^2 \leq 0.3$ are considered to represent high amplitude motions. Elevated temperatures resulted in a marked decrease in order parameter values which ranged from 0.14 to 0.17 at 57°C. The lower order parameter values at higher temperatures signify loosening of the hydrophobic core where the side chain motions are now less restricted. This result also corroborates the previous spin diffusion and $^{13}$C-$^1$H HMQC experiments.

### 3.5 Identifying a Possible Off-Pathway Intermediate Within Calmodulin’s Folding Network

While the single $^{19}$F CPMG experiment at 57°C shows that CaM undergoes millisecond timescale excursions to the near-native intermediate state, a CPMG temperature dependence study reveals that the conformational exchange occurring, is not as simple as a two-state process. Figure 42 shows relaxation dispersion profiles for F68, F19, and F92 from 45°C to 57°C and Table 2 shows the extracted exchange rates, $k_{ex}$.

![Figure 42. $^{19}$F CPMG relaxation dispersion profiles for residues F68, F19, and F92 as a function of temperature (Red-57°C, Green-51°C, Blue-45°C).](image-url)
Table 2. Extracted chemical exchange rates, $k_{ex}$, from $^{19}$F CPMG dispersion profiles of 3F-Phe enriched CaM.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chemical exchange rates, $k_{ex}$ (Hz), from CPMG relaxation dispersions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temperature</td>
</tr>
<tr>
<td>F68</td>
<td></td>
</tr>
<tr>
<td>F19</td>
<td></td>
</tr>
<tr>
<td>F92</td>
<td></td>
</tr>
</tbody>
</table>

For residues F68 and F19 (located in the N-terminal domain), the temperature dependence of $k_{ex}$ shows that transitions between the native and near-native states become increasingly faster at elevated temperatures (Figure 43). However, residue F92 (located in the C-terminal domain) exhibits non-Arrhenius behaviour where transitions between the native and near-native states become slower at higher temperatures (Figure 43).

Figure 43. Temperature dependence of the chemical exchange rates, $k_{ex}$, derived from $^{19}$F CPMG dispersion profiles of 3F-Phe CaM. Residues F68 and F19 (blue and green) exhibit exchange kinetics which are faster at higher temperatures while residue F92 (red) exhibits exchange kinetics which become slower with increasing temperature.
Furthermore, using Equation 5 and the populations extracted from the $^{19}$F chemical shift fits, it is possible to interpret the exchange rate in terms of a folding rate, $k_f$, and an unfolding rate, $k_u$ (Figure 44).

![Figure 44](image)

**Figure 44.** Folding and unfolding rates of 3F-Phe CaM as a function of temperature. Folding and unfolding rates were calculated using the populations extracted $^{19}$F chemical shift fits and the exchange rates from the CPMG experiments.

For residues F68 and F19, folding and unfolding rates both increase with increasing temperature. For residue F92, unfolding rates show a subtle decrease with increasing temperature while folding rates show a more profound decrease with temperature. We attribute this non-Arrhenius behavior exhibited by F92 to the presence of an off-pathway intermediate.

As mentioned in the introduction, recent single molecule experiments have suggested that along CaM’s folding/unfolding pathway, a number of different non-native off-pathway intermediates may be populated. Highly structured off-pathway intermediates are often able to accumulate due to prematurely formed interactions. These intermediates are considered off-pathway as they never fold towards the native state; that is, the protein must return to the on-pathway intermediate conformation before continuing along the folding funnel towards the
native state. As we populate the near-native intermediate state by elevating temperatures, it increases the likelihood that the protein converts to a trapped off-pathway intermediate state. This results in an overall slowing of the measured exchange rate between the native and the on-pathway intermediate states. We conclude that this off-pathway intermediate exchanges with the near-native intermediate, and not the native state, based on the observation that folding rates, as opposed to unfolding rates, become increasingly slower upon population of the near-native intermediate. Figure 45 presents a model of the folding network of CaM based on the observations in the CPMG experiment.

![Figure 45](image)

**Figure 45.** Schematic representation of the folding/unfolding pathway of CaM. $^{19}$F CPMG experiments reveal that the folding network of CaM involves both on- and off-pathway intermediates.

There are a number of possible structural explanations of the off-pathway intermediate. It is important to note that the non-Arrhenius behaviour is only exhibited in the C-terminal domain. This narrows the possibility that this off-pathway intermediate could be a result of: 1) Intramolecular conformational changes within the C-terminus, 2) Intermolecular interaction between the C-terminal domains of two different CaM molecules, and 3) Intramolecular conformational change where the N-terminal and C-terminal of a single CaM molecule interact. The last possibility could go undetected by residues in N-terminal (F19 and F68), if the protein structure around these residues undergoes only minor changes in the off-pathway intermediate. To investigate these hypotheses, the N- and C-terminal domains of CaM (TR1C and TR2C, respectively), separated through a trypsin digest, are individually examined.
Figure 46 shows $^{19}$F NMR spectra of CaM TR1C and TR2C. The $^{19}$F 1D NMR spectra of the separated domains reveal three resolved fluorine resonances for the TR2C domain (F89, F141, F92) and five fluorine resonances for the TR1C domain (F12, F16, F68, F65, F19).

![Figure 46. $^{19}$F NMR spectra of 70% 3F-Phe labeled full length CaM (blue), TR2C(green), and TR1C(red) at 45°C.](image)

The $^{15}$N-$^1$H HSQC spectra of the two domains, shown in Figure 47, show that the two domains still retain a similar tertiary structure when compared to the full length protein.

![Figure 47. $^{15}$N-$^1$H HSQC spectra of 70% 3F-Phe labeled full length CaM (black) overlaid with an analogous spectrum of TR2C(red) and TR1C(blue) fragments at 45°C.](image)
$^{19}$F CPMG experiments were repeated on the individual domains of CaM and the relaxation dispersion profiles and extracted exchange rates are shown in Figure 48 and Table 3, respectively. Again, profound dispersions are seen for F68, F19, and F92, indicative of chemical exchange occurring on the millisecond timescale. For TR1C, the chemical exchange rate shows a similar temperature dependence to that observed in full-length CaM (Figure 48). TR1C exhibits exchange kinetics which are faster at higher temperatures. However, the trend in exchange rate for TR2C is now in contrast to what was observed in full-length CaM (Figure 48).

**Figure 48.** Temperature dependence of the chemical exchange rates, $k_{ex}$, derived from $^{19}$F CPMG dispersion profiles of the isolated domains of CaM, TR1C and TR2C. As opposed to the non-Arrhenius behavior observed in full-length CaM, the isolated domains now exhibit exchange rates which are faster at higher temperatures.
Table 3. Extracted chemical exchange rates, $k_{ex}$, from $^{19}$F CPMG dispersion profiles of 3F-Phe TR1C and TR2C.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Chemical exchange rates, $k_{ex}$ (Hz), from CPMG relaxation dispersions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temperature</td>
</tr>
<tr>
<td></td>
<td>35°C</td>
</tr>
<tr>
<td>F68 (TR1C)</td>
<td>6200 ± 950</td>
</tr>
<tr>
<td>F19 (TR1C)</td>
<td>6200 ± 950</td>
</tr>
<tr>
<td>F92 (TR2C)</td>
<td>6200 ± 950</td>
</tr>
</tbody>
</table>

In the absence of the N-terminal domain, the C-terminal domain now shows Arrhenius behavior where the exchange rate is faster at higher temperatures. This result suggests that the off-pathway folding/unfolding events in full-length CaM, involves an intermediate in which both the N- and C-terminal domains interact. This is a similar conclusion noted in a recent single molecule study.23 Through single molecule experiments, Stigler et al. found that along CaM’s folding/unfolding pathway, the protein can populate a non-native intermediate with mis-paired EF hands belonging to the two different domains. The mis-paired EF hands had to be unpaired before the protein could continue along the folding funnel towards the native state thus, slowing the transition between native and on-pathway intermediate states. Paramagnetic relaxation enhancement experiments on calmodulin have also provided evidence that the protein can transiently adopt compact state where the two domains come in contact.101 The transient interaction between the N- and C-terminal domains may not detected by the probes in the N-terminal due to the local environment around these probes not being significantly altered in the off-pathway intermediate.

3.6 Possible Implications of the Near-Native Desolvated Intermediate in Misfolding

The role of a dry near-native folding intermediate in misfolding is also intriguing. The rate-limiting step in protein aggregation is often desolvation.102 While CaM admittedly has no connection to amyloidosis, we use it as a model system to probe the possible implication of a desolvated intermediate state to aggregate formation.

To investigate the possible role of the desolvated intermediate in aggregation, CaM was incubated, over the period of three weeks, at temperatures near the transition to the near-native
intermediate state (\(\sim 65^\circ C\)). \(^{19}\text{F}\) NMR spectra were recorded in two day intervals to monitor aggregate formation. After 1.5 weeks of incubation, the formation of an aggregated state becomes clear through the \(^{19}\text{F}\) NMR spectrum (Figure 49).

**Figure 49.** An inset of the \(^{19}\text{F}\) NMR spectra of 70% 3F-Phe labeled CaM collected at various time points during the three week incubation period at 65°C. A deconvolution is shown to highlight the peak arising from F65 (red) and the broader peak which we associate with an aggregated state (blue).

\(^{19}\text{F}\) spectra of the incubated sample show the formation of an additional broad peak which we associate with the onset of an aggregated state. Based on the changes to the integral of the peak, the population of the aggregated state was found to steadily increase over the three week incubation (Figure 50). To correlate the desolvated intermediate state to the formation of the aggregated state, we incorporate 5% TFE, which is known to stabilize the native state and prevent exchange to the desolvated intermediate, to a second CaM sample and repeat the experiment. Under the same conditions, incubation of CaM with 5% TFE considerably reduces the formation of the aggregated state.
Figure 50. Change in the integral of the aggregate peak, shown in Figure 49, over the three week incubation in the absence (blue) and presence (green) of 5% TFE.

By preventing excursions to the desolvated intermediate state the assembly of aggregated states is drastically reduced. This result clearly implicates the near-native intermediate to have a role in aggregate formation. A constructed model, shown in Figure 51, proposes that conformational excursions can convert the native state into the desolvated near-native intermediate, which then renders the protein susceptible to aggregate formation.

Figure 51. Graphical representation of the complex folding network of CaM. Excursions to the near-native intermediate are believed to make the protein more susceptible to aggregate formation. For simplicity of the figure, aggregated states are shown to be on-pathway. It is also possible that aggregated states represent an off-pathway kinetic trap.
3.7 The Effect of Calcium-Binding on the Folding Pathway of Calmodulin

Holo-CaM is somewhat atypical of globular proteins in that it is held together not only by hydrophobic forces but by additional ionic forces, via the four tightly bound calcium ions within the EF hands. The binding of calcium to CaM also exposes hydrophobic patches to the surrounding solvent. Furthermore, the hydrophobic interior is recognized to be somewhat more dynamic than other globular proteins, which may explain the anomalously high degree of water and oxygen penetration observed in the vicinity of the phenylalanines in the protein interior. On the other hand, apo-CaM has an overall more compact structure and has a hydrophobic core more representative of a typical globular protein. Here we study this apo state to investigate the effect of calcium binding on the folding pathway of CaM.

Apo-CaM not only differs from holo-CaM in structure, but also has significantly lower thermal stability (melt temperature of ~55°C). Figure 52 shows 3F-Phe apo-CaM \(^{19}\)F NMR spectra at 16°C, 34°C, and 54°C.
Figure 52. $^{19}$F NMR spectra of 70% 3F-Phe labeled apo-CaM at various temperatures. Residue fluorine signals which are believed to arise from the N-terminal (blue) and C-terminal (red) are highlighted.

Due to the low melt temperature of apo-CaM, significant chemical shift changes are seen in the $^{19}$F spectra at elevated temperatures. As temperature is increased from 16°C to 54°C, we observe the gradual convergence of five separate peaks to a small ppm range (-162.5 to -163.3) and also broadening of three other peaks. Though no assignments have been made for the fluorine peaks of apo-CaM, we associate the five peaks which converge at 54°C to be those from the N-terminus (F12, F16, F19, F65, and F68) and the remaining three peaks to be from the C-terminal (F89, F92, F141). The reasoning behind this is that the two domains of apo-CaM are known to have separate melt temperatures which are ~ 15°C apart. The convergence of the five peaks at 54°C is thought to represent an unfolded N-terminal domain while the broadening of the other three remaining peaks represents partially folded C-terminal domain. Tertiary structure changes can be further monitored using $^{15}$N-$^1$H HSQC experiments (Figure 53).
Figure 53. $^{15}$N-$^1$H HSQC spectra of wt apo-CaM at 16°C and 56°C. The dramatic decrease in dispersion at 56°C signifies the onset of an unfolded state.

$^{15}$N-$^1$H HSQC spectra of wt apo-CaM at 16°C and 54°C reinforce the conclusions drawn from the $^{19}$F spectra. At 16°C, well resolved and disperse peaks are observed in the $^{15}$N-$^1$H HSQC spectrum, indicative of a well-folded state. Elevated temperatures cause the loss and clustering of peaks in the spectrum, suggesting the onset of an unfolded state. Though previous research has noted the N-terminus to be more thermally stable than the C-terminal, we attribute the discrepancy observed here to a small perturbation introduced as a result of the incorporated fluorine atoms. This perturbation, which is avoided by holo-CaM, may only arise in apo-CaM due to its more compact structure. The N-terminal domain structure may be slightly more affected than the C-terminal domain because of the positions of the phenylalanine residues within the N-terminal as well as the fact that the N-terminal contains two additional fluorine substitutions. This extra perturbation in the N-terminal domain may be enough to shift the melt temperature to one which is below that of the C-terminal domain.

The temperature dependence of the chemical shifts, $\Delta \delta_{19F}$, give further insight into apo-CaM’s folding/unfolding pathway (Figure 54).
**Figure 54.** A temperature-dependent plot of the change in the $^{19}$F chemical shift ($\Delta \delta_{^{19}\text{F}}$), relative to the shift at 16°C, associated with each of the eight 3F-Phe residues in apo-CaM. The non-linear dependence of the chemical shifts here is believed to signify the onset on an unfolded state.

As observed in holo-CaM, the chemical shifts of apo-CaM also exhibit a non-linear dependence of the chemical shifts with temperature. However, in the case of apo-CaM, the non-linear dependence of the chemical shifts is believed to signify the formation of an unfolded state and not an intermediate state. We rule out the possibility of an intermediate here since the chemical shifts only deviate from linearity near the melt temperature of 3F-Phe apo-CaM (~54°C, as indicated by the $^{19}$F spectra), where a significant population of unfolded state would be expected to exist. The unfolded state would exhibit a linear temperature dependence of chemical shift different from that of the folded state thus, result in a non-linear chemical shift change at elevated temperatures. We conclude from this that the folding/unfolding pathway of apo-CaM occurs via a simple two-state process which lacks a folding intermediate.

As in the case of holo-CaM, $^{19}$F CPMG experiments on apo-CaM are used to monitor the possibility of any chemical exchange processes. Figure 55 shows $^{19}$F apo-CaM spectra at various $\nu_{\text{CPMG}} (1/2\tau)$ values.
Figure 55. $^{19}$F NMR spectra of 70% 3F-Phe labeled apo-CaM at various $\nu_{\text{CPMG}}$ (1/2$\tau$) values. The dotted black lines are shown to highlight that varying $\nu_{\text{CPMG}}$ values have no noticeable effect on peak intensities.

As highlighted by the dotted lines in Figure 55, the peak intensities show no discernable decay at lower $\nu_{\text{CPMG}}$ values. In contrast to the calcium-loaded state, the absence of dispersion in the CPMG experiment suggests that apo-CaM has a less dynamic hydrophobic interior and does not undergo exchange with any other conformational states. This result also supports the notion that apo-CaM lacks any observable intermediates. Taken together, studies on apo-CaM suggest that calcium binding not only dramatically alters the folding network of CaM, but also affects the dynamic nature of the protein.
Chapter 4
Discussion and Conclusions

4.1 The Role of the Near-Native Desolvated Intermediate in Folding and Misfolding

The classic notion of protein folding posits that the process is largely driven by hydrophobic collapse in which the release of waters of hydration from the unfolded polypeptide results in a net gain in entropy, thereby propelling the protein into a folded state. More recent descriptions emphasize the role of backbone hydrogen bonds in the early stages of folding. In this case, secondary structure and overall topology is expected to be established prior to hydrophobic collapse. Our data shows that an intermediate, consistent with a dry near-native state, is established along the temperature denaturation pathway. Upon heating from lower temperatures associated with the native state, desolvation barriers decrease, reaching a point where water is released from the protein and the hydrophobic interior adopts a liquid alkane-like state. Moreover, the observed trends of solvent isotope shifts, oxygen solubility, and hydrophobicity from 30-65°C are consistent with theoretical descriptions which predict hydrophobic interaction strengths to increase with temperature as envisaged in Figure 56.

![Figure 56](image)

**Figure 56.** Graphical representation of water and oxygen accessibility in the native (N) and the near-native intermediate (I) states of CaM, based upon solvent isotope shifts (Δδ*(H2O)) and paramagnetic shifts (Δδ*(O2)) associated with 3-fluorophenylalanine probes in the protein interior. Sites having higher water or oxygen accessibility are shown with spheres having a larger radius and darker shade of blue or pink, respectively. Note that residues F12 and F141 exhibit solvent isotope shifts, which are too small to be observed using the current scale. The depiction of the near native intermediate structural model was constructed by slightly modifying the known native CaM structure to correspond to a slightly larger hydrodynamic radius.
Others have also observed that hydrophobic forces and in some instances (through changes in side chain pKa) internal salt linkages, tend to increase with higher temperatures. The near-native intermediate state precedes protein compaction upon cooling, and is characterized by similar backbone structure to that of the native state and minimal solvent penetration. Moreover, the presence of greater dynamics within the hydrophobic interior provides a gain in configurational entropy, thereby stabilizing the intermediate. This would also represent an advantage in the folding process, since diffusive searches toward the compact folded state would presumably be very fast. Others have noted the importance of configurational entropy in proteins such as calmodulin for purposes of ligand binding, while at the same time establishing the presence of conformers within the ensemble, which are consistent with the description of a dry molten globule. The final step in the folding process is expected to be cooperative and facilitated by closer association of the hydrophobic residues and reentry of a few water molecules of hydration, as the interior becomes close packed. Cheung and others have noted the importance of water in “lubricating” the protein folding process. Indeed, there are many examples in the literature in which folding intermediates are demonstrated to be hydrated and a “wet” molten globule state has been suggested. In our case, however, a dry near-native folding intermediate is clearly observed along the temperature denaturation pathway, and there are potential consequences of such states to both function and to disease through protein aggregation and misfolding.

In protein folding, desolvation may represent the rate-limiting step. Upon establishing the dry near-native intermediate, the increased dynamics effectively provides greater sampling of the ensemble and a lowering of the barrier toward more fully folded conformers. The increased disorder inherent in the intermediate effectively accelerates the folding process, which involves collapse of the (desolvated) hydrophobic core and stabilization of specific inter-residue contacts through van der Waals, hydrogen bonds, and salt links. While we do not imagine the folded state to be highly solvated, there are clearly specific waters of hydration involved in stabilizing the folded state, at least in the vicinity of the phenylalanines. These waters of hydration may also arise from higher energy conformers that are more solvent exposed but associated with the native ensemble. The strength and density of water interactions with the phenylalanine residues, in the protein interior, is diminished in the dry folding intermediate. The main point is simply that the trajectory of solvent exposure along the temperature
folding/unfolding pathway involves a desolvation step and the establishment of a dry near-native state, whose hydrodynamic radius is increased.

Desolvation may also represent the rate-limiting step in protein aggregation. In studies of amyloidosis, protein aggregation and fibril formation is achieved once a partially unfolded metastable state is reached. If the intermediate is more dynamic one might expect rare conformational excursions, which lead irreversibly to misfolded states. In the case of a desolvated intermediate, distinct and likely transient states may be more predisposed to aggregation through non-native hydrophobic contacts. While CaM is admittedly a model protein with no connection to amyloidosis, we have observed the formation of aggregated states upon incubation of CaM over periods of weeks, at temperatures near the transition to the near native state (~60°C). Moreover, the incorporation of 5% TFE, which is known to stabilize the native state and prevent excursions to the near native intermediate, results in no detectable aggregates, under the same conditions.

4.2 The Effect of Off-Pathway Intermediates on Folding

The ability of proteins to efficiently adopt their native conformation is important for biological systems. While the presence of on-pathway near-native intermediates is often productive for folding, off-pathway intermediates can slow the overall folding rate of the protein and lead to irreversible aggregated states. These off-pathway intermediates act as kinetic traps and must undergo extensive structural reorganization to achieve the native state. Our results show that the folding network of CaM involves an off-pathway intermediate with non-native interactions between the two calcium-binding domains. This off-pathway intermediate extends from the desolvated near-native intermediate and competes with the productive folding pathway. Upon population of the near-native intermediate, and subsequently the off-pathway intermediate, folding rates were considerably slowed. The effect of off-pathway intermediates in CaM’s folding network is also highlighted in single molecule studies. Stigler et al. note that while the individual domains of CaM can fold within microseconds, complete folding of CaM has a long average folding time of 5 seconds due to the presence of off-pathway intermediates. Studies on apoflavodoxin also demonstrate the extent at which off-pathway intermediates can affect the folding process. In a GuHCl-induced unfolding/folding study, approximately 90% of apoflavodoxin molecules were observed to encounter an off-pathway intermediate.
Interestingly, the unfolding of this off-pathway intermediate was found to represent the rate-limiting step in folding to the native state.

Off-pathway intermediates have also been implicated with misfolding events.\textsuperscript{121,122} In the initial steps of folding, the unfolded state often samples various folding pathways which occasionally lead to dead-end off-pathway intermediates. These intermediates are typically molten globule-like and contain exposed hydrophobic groups making the protein more predisposed to aggregation. However, in the case of CaM, very little structural information of the off-pathway intermediate can be obtained. As such, it is difficult to determine whether this intermediate is involved in the formation of aggregated states.

### 4.3 The Effect of Calcium-Binding on Calmodulin’s Folding Network

Investigations of apo- and holo-CaM revealed that calcium binding has a dramatic impact on the folding network and dynamics of the protein. In its calcium-free state, CaM exhibits a rather simple folding pathway with no discernible intermediates. In contrast, the folding pathway of calcium-loaded CaM was found to be more complex, consisting of both an on-pathway intermediate and an unproductive off-pathway intermediate. The differing results can best be attributed to the large conformational changes the protein undergoes upon calcium binding.

While the two domains of apo-CaM have a typical compact globular structure, the domains of holo-CaM are less tightly packed and contain exposed hydrophobic patches. Native state topology has been thought to be the most important factor in determining how a protein folds.\textsuperscript{10} Simple point mutations have been shown to switch some proteins from a simple two-state folder, to a multi-state folder.\textsuperscript{8} Folding rates also correlate remarkably well with topology of the folded structure and can be predicted using the relative contact order (RCO) parameter.\textsuperscript{10,11} The RCO is the average sequence separation between all pairs of contacting residues, normalized by the number of residues, and is inversely proportional to the natural logarithm of the intrinsic folding rate. The RCO is small in proteins which are stabilized by short-range interactions and large in systems where there is an abundance of long-range contacts. A smaller RCO would result in a faster folding rate. Based on this empirical model, the two domains of apo-CaM are predicted to fold within microseconds where a stable intermediate may not even be necessary for efficient folding of the protein. Holo-CaM having a slightly unusual globular structure as well as
inclusion of calcium ions within its system, is expected to have an overall slower folding rate which may lead to its complex folding network. A similar result on the effect of calcium binding on the folding pathway of CaM has also been observed in recent single molecule studies. It is also interesting to note the change in the CaM’s dynamic nature upon calcium binding. CPMG dispersion experiments revealed that holo-CaM contains a very dynamic hydrophobic interior whereas apo-CaM was less susceptible to conformational exchange events. This result suggests that holo-CaM is not static, but often exchanges between different conformations on a millisecond to microsecond timescale. The dynamic nature of holo-CaM has previously been shown to be directly correlated to its overall function. By being able to sample various conformations, it is possible that holo-CaM can access many functionally relevant states allowing the protein to bind a wide variety of ligands. Paramagnetic relaxation enhancement experiments on holo-CaM have also provided direct evidence of this. Through the use of a nitrooxide spin-label, holo-CaM, in its native state, was shown to sample an ensemble of states whose structure closely matches those of the ligand bound state. The sampling of apo-CaM is restricted to much smaller conformational space.

4.4 Conclusions and Future Directions

This study considers the topology of a well-known soluble protein, calmodulin, along the protein folding/unfolding pathway established by temperature. In CaM’s calcium loaded state, around 66°C, we detect a folding intermediate whose backbone structure is native-like, as judged by CD and 15N-1H NMR spectroscopy. 19F NMR solvent isotope shifts show that water accessibility is clearly reduced in the hydrophobic interior upon attaining the folding intermediate, thus addressing a longstanding debate regarding the existence of dry near-native intermediates. The combination of solvent isotope shifts and O2 paramagnetic shifts give a detailed microscopic perspective of both solvent exposure and hydrophobicity. Moreover, CPMG measurements point toward the existence of the near-native state at physiological temperatures though we cannot say if such a state is on-pathway between native and unfolded states at physiological temperatures. Based on NMR observations of holo-CaM as a function of temperature, we would conclude that folding can occur by a rapid and early acquisition of secondary structure and hydrogen bonds until a “dry” near-native intermediate is attained. Given the high extent of disorder associated with this intermediate, it can then diffusively explore conformations, which allow it to adopt a
fully folded compact state. We also identify the existence of an off-pathway intermediate consisting of non-native contacts between the two domains. The presence of this intermediate within the folding network considerably slows the overall folding rate of the protein. Studies on apo-CaM revealed that calcium binding has a significant impact on the folding network of CaM. In contrast to holo-CaM, apo-CaM is thought to represent a two-state folder lacking observable intermediates.

In future studies, we will look to employ the methods presented in this report to other proteins systems. $^{19}$F NMR provides unique insight into a protein’s structure, dynamics, and conformational equilbiria. In particular, focus will be placed on amyloidogenic proteins. Our hypothesis is that dry near-native states are connected with off-pathway misfolded states which in turn, connect to aggregated states. There may be a parallel in amyloidogenic proteins.
References


