EXPERIMENTAL INVESTIGATION OF PROTEIN STABILITY AND PROTEIN FOLDING UTILIZING THE GLASSY STATE

Theresa Yuek-Chiu Cho

A DISSERTATION
PRESENTED TO THE FACULTY OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF CHEMICAL AND BIOLOGICAL ENGINEERING

Advisor: Professor Pablo G. Debenedetti

April 2012
Abstract

Proteins are one of the main components of living systems and have a wide diversity of structure and function. Aside from being biology’s workhorses, proteins are also utilized in industrial, agricultural and environmental applications. Although proteins are very useful in many different applications, the number of applications is limited by the fact that the folded and functional state of a protein is only marginally stable. Increasing the stability of proteins in general, and enzymes in particular, is of major technological importance. The first part of this dissertation research is directed at studying phenomena and factors that impact protein stability. Experiments using differential scanning calorimetry, infrared spectroscopy, ultraviolet spectroscopy and enzymatic assays were conducted to study the effect of trehalose on lysozyme aggregation when subject to thermal treatment. A high concentration of lysozyme (200 mg/ml) in a high concentration of trehalose (80 wt%) results in aggregation. On the other hand, the use of lower trehalose concentrations (0 to 40 wt%) prevents or inhibits aggregation at both high and low protein concentrations. A higher initial protein concentration increases the rate of aggregation. The rates of aggregation and inactivation of lysozyme in trehalose follow first-order kinetics.

Despite the attractive features of stabilized functional proteins for biotechnological applications, proteins in vivo can misfold resulting in loss of functionality or becoming toxic to the cell. Diseases associated with misfolding include Alzheimer’s disease, cystic fibrosis, and Parkinson’s disease. Correct folding of proteins, the process in which their characteristic and functional three-dimensional function
structure forms from a random coil, is one of nature’s most ubiquitous phenomena. Understanding the mechanisms underlying protein folding is important for providing insight into protein misfolding and aggregation as well as for the rational modification and design of novel proteins. One important step in unraveling the protein folding mechanisms is to characterize intermediate or partially structured states, in terms of their structural, kinetic and thermodynamic properties. However, uncovering the mechanisms underlying the folding process has always been a challenge due to the very short time scales for proteins to fold. The second part of this dissertation focuses on the development of a new experimental technique to study protein folding mechanisms by capturing intermediate structures during protein refolding using a quench-and-refold approach with infrared spectroscopy.

The experimental procedure developed involves first thermally denaturing a protein, then trapping the unfolded biomolecule by rapidly quenching the system into a glassy state, and finally heating the system slowly across its glass transition. The main advantage of this technique is the opportunity to observe subtle structural changes as they evolve from a denatured state back to the native state over a continuum in time, due to the slow kinetics found in the viscous liquid state near the glass transition temperature. Infrared spectroscopy is employed to observe secondary structural changes throughout the procedure. This method requires careful consideration of an appropriate protein solution medium or solvent, such that the solvent does not denature the protein nor permit aggregation. The solvent should also prevent ice crystallization upon cooling or heating. A solution consisting of sucrose, ethylammonium nitrate, and water (or heavy water) was eventually chosen and used for subsequent quench and refold experiments.
Lysozyme was used as the model protein for initial experiments. Further studies were then conducted with Ribonuclease A. By comparison of differential scanning calorimetry and infrared spectroscopy measurements, early energy-neutral structure formation events are revealed.
Acknowledgments

First and foremost I would like to thank my advisor Pablo Debenedetti. I am very grateful to have worked with him on this challenging experimental project. I also admire him for his extensive knowledge in the research field, excellent teaching skills and his aptitude in administrative appointments. I have learned a lot from his guidance.

I greatly appreciate the opportunity to work with Brian Pethica, who provided immeasurable help and insight into experimental details and analyses, as well as interesting conversations about life experiences on anything ranging from dance to research. I also enjoyed my collaborations with David Moore on IR experiments and with Austen Angell and Nolene Byrne on quench-and-refold experiments. I would also like to thank Jay Benziger, Ilhan Aksay and Jannette Carey for allowing me to use their laboratory equipment for my many experiments.

I thank my group members, especially Lawrence, Scott, Brian, Tom, Swaroop, Andy, and Nicolas for all the work discussions, group lunches and fun times both inside and outside of the office or lab. I also thank my friends, particularly Erin, Antti-Pekka, Jessica, Pan and Brian, as well as Triple 8, for making my life so much more enjoyable at Princeton.

I would very much like to thank my family. Special gratitude goes to my parents, Dr. Lunkit Frank Cho and Alice, and my siblings, Jennifer and Chris, for their continuous love and support in all my decisions in life. Last but not least, a big thank you to my husband, Stefan, for his never-ending energy, creativity, optimism, support and encouragement, especially towards the finishing of my thesis in Sweden. Tack så jätte mycket min älskling.

vi
# Table of Contents

Abstract.............................................................................................................iii

Acknowledgements..........................................................................................vi

List of Figures...................................................................................................x

List of Tables..................................................................................................xvii

1 Introduction.................................................................................................1
   1.1 Glassy State Properties.................................................................8
   1.2 Protein Stability..............................................................................11
      1.2.1 Proteins in Solution...............................................................12
      1.2.2 Proteins in the Glassy State....................................................15
   1.3 Protein Folding Mechanism...........................................................16
      1.3.1 Driving Forces of Protein Folding ........................................16
      1.3.2 Classic Versus New View of Protein Folding.......................18
      1.3.3 Structural Mechanisms of Protein Folding..........................23
   1.4 Current Experimental Techniques for Protein Studies.................25
      1.4.1 Differential Scanning Calorimetry........................................26
      1.4.2 Ultraviolet Spectroscopy......................................................29
      1.4.3 Infrared Spectroscopy.........................................................30
      1.4.4 Other Experimental Techniques..........................................34
   1.5 Thesis Outline...................................................................................36
   1.6 References.........................................................................................39
2 Experimental Investigation of Lysozyme Aggregation in Trehalose.............52

2.1 Experimental methods.................................................................57

2.1.1 Materials..........................................................57

2.1.2 Differential Scanning Calorimetry (DSC)..............................58

2.1.3 Infrared Spectroscopy (IR)..................................................59

2.1.4 Ultraviolet Spectroscopy (UV).............................................61

2.1.5 Enzymatic Activity Assay....................................................61

2.2 Results and Discussion..............................................................62

2.2.1 Lysozyme in 80 wt% Trehalose...........................................62

2.2.2 Effect of Trehalose Concentration.......................................65

2.2.3 Effect of Protein Concentration..........................................73

2.2.4 Kinetics..............................................................................78

2.3 Conclusion..............................................................................84

2.4 References..............................................................................86

3 Structure-Energy Relations in Hen Egg White Lysozyme Observed During
Refolding From a Quenched Unfolded State........................................94

3.1 Experimental Methods.............................................................101

3.1.1 Materials..........................................................101

3.1.2 Differential Scanning Calorimetry....................................102

3.1.3 Enzymatic Activity Assay...............................................104

3.1.4 Infrared Spectroscopy....................................................105

3.1.5 Isothermal Titration Calorimetry.....................................106
3.2 Results and Discussion

3.2.1 Approach Formulation

3.2.1.1 Protein Selection

3.2.1.2 Solvent Formulation

3.2.2 IR Qualitative Analysis

3.2.3 IR Quantitative Analysis

3.2.4 Comparison of DSC and IR

3.2.4.1 Protein Concentration

3.2.4.2 Solvent Composition

3.2.4.3 Solvent Evaluation

3.2.4.4 Kinetics

3.3 Conclusion

3.4 References

4 Structure-Energy Relations in Bovine Pancreatic Ribonuclease A Observed During Refolding From a Quenched Unfolded State

4.1 Experimental Methods

4.2 Results and Discussion

4.2.1 RNase A in EAN, sucrose and H₂O/D₂O

4.2.2 Effect of EAN Concentration

4.3 Conclusions

4.4 References

5 Concluding Remarks and Suggestions for Future Work

5.1 References
List of Figures

Figure 1.1: Isobaric relationship between volume and temperature in the liquid, glassy and crystalline state………………………………………………………………………9

Figure 1.2: Strong and fragile liquids characterized by their Arrhenius and non-Arrhenius behavior………………………………………………………………………………..11

Figure 1.3: Schematic outline of a folding energy landscape………………………………………………………22

Figure 1.4: Representative free energy landscape for an idealized two-state and a multi-state folding protein…………………………………………………………………….23

Figure 1.5: Visual representation of four structural models of protein folding………………24

Figure 1.6: Schematic DSC curve illustrating the common features of a thermogram………………………………………………………………………………………………………27

Figure 1.7: Representative DSC scan showing the thermal information provided by DSC technique……………………………………………………………………………28

Figure 1.8: Representative second derivative spectrum and amide I curve fitting into Gaussian components……………………………………………………………………32

Figure 2.1: Secondary structure ribbon and amino acid sequence bead visual representation of lysozyme……………………………………………………………………………….56

Figure 2.2: IR spectra of 30 mg/ml lysozyme in 80 wt% trehalose-D$_2$O…………………………………….63

Figure 2.3: DSC heating scans for 30 mg/ml lysozyme in various concentrations of trehalose………………………………………………………………………………………..66
Figure 2.4: Dependence of denaturation temperature and enthalpy on trehalose concentration for 30 mg/ml lysozyme…………………………………...67

Figure 2.5: Stacked plots of second derivative and deconvoluted amide I’ spectra of 30 mg/ml lysozyme in 0 wt% and 40 wt% trehalose-D2O………………70

Figure 2.6: IR spectra of amide I’, amide II and amide II’ regions for 30 mg/ml lysozyme in D2O………………………………………………………71

Figure 2.7: Absorbance intensity for the 1652 cm⁻¹ band for 30 mg/ml lysozyme in various trehalose concentrations……………………………………73

Figure 2.8: Aggregation percentages determined from DSC of 30 mg/ml lysozyme and 200 mg/ml lysozyme in various concentrations of trehalose……..74

Figure 2.9: Effect of initial protein concentration on aggregation percentage for lysozyme heated above Td in the absence and presence of trehalose……75

Figure 2.10: Effect of initial protein concentration on irreversible enzymatic inactivation of lysozyme heated above Td in the absence and presence of trehalose………………………………………………………76

Figure 2.11: Lysozyme aggregation of 30 mg/ml and 0.1 mg/ml lysozyme after different thermal anneal times above Td at various trehalose concentrations………………………………………………………79

Figure 2.12: Lysozyme enzymatic inactivation of 30 mg/ml and 0.1 mg/ml lysozyme after different thermal anneal times above Td at various trehalose concentrations…………………………………………………………81

Figure 2.13: Linear kinetic fits to lysozyme aggregation percentages of 30 mg/ml and 0.1 mg/ml lysozyme in various trehalose concentrations…………………82
Figure 3.1: Comparison of the heating up scans of 21 wt% lysozyme, 29 wt% EAN, 30 wt% sucrose and 20 wt% water sample cooled in the DSC from 30 °C and quenched with LN2 from 86 °C………………………………………………99

Figure 3.2: Representative DSC scan illustration integrated areas to determine unfolding and refolding energies of a quench-and-refold scan………100

Figure 3.3: State diagram for trehalose-water binary mixture…………………………113

Figure 3.4: Illustration of the proposed explanation for EAN’s ability to prevent aggregation during thermal denaturation…………………………115

Figure 3.5: Photos of turbidity of 200 mg/ml lysozyme dissolved in EAN and water solutions at different temperatures……………………………………117

Figure 3.6: DSC scans to test the reversibility of thermal cycling of 200 mg/ml lysozyme in EAN-H2O solutions…………………………………………119

Figure 3.7: Determination of the reversible and irreversible aggregation temperature of 200 mg/ml lysozyme in a 20 mol% EAN solution upon multiple scans……………………………………………………………………120

Figure 3.8: Active units per mg of lysozyme dissolved in D2O, H2O and 20 mol% EAN-80 mol% H2O……………………………………………………122

Figure 3.9: DSC scans of 200 mg/ml lysozyme in 60 wt% EAN-water and 40 wt% EAN, 27 wt% H2O and 33 wt% sucrose, showing ice crystallization in the absence of sucrose…………………………………………………………123

Figure 3.10: Successive DSC heating curves of 200 mg/ml in the presence and absence of EAN showing reversibility of lysozyme…………………………124
Figure 3.11: IR spectra of the percent transmission with respect to wavenumber of D$_2$O and H$_2$O………………………………………………………………………………126

Figure 3.12: IR spectrum of deuterated lysozyme at 25 °C for 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O…………………………………………………………127

Figure 3.13: Original background unsubtracted IR spectra of 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O…………………………………………………………129

Figure 3.14: Raw IR spectra of 40 wt% EAN, 33 wt% sucrose and 27 wt% D$_2$O at different temperatures after quenching in LN$_2$…………………………….130

Figure 3.15: Spectra of folded lysozyme (2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O) at 25 °C before thermal treatment and at -135 °C after slow cooling and quenching in LN$_2$…………………………………………………………131

Figure 3.16: IR spectra of lysozyme (2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O) from quench-and-refold approach…………………………………………………………132

Figure 3.17: Absorbance level at 1652 cm$^{-1}$ and 1627 cm$^{-1}$, associated with α–helix and β-sheets respectively, for initial states before treatment and after heating and quenching……………………………………………………………133

Figure 3.18: Comparison of the second derivative spectrum and the curve fitting of spectra for 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O at 25 °C and -65 °C after thermal denaturation, quench and slow heating across T$_g$……………………………………………………………137

Figure 3.19: Example of quantitative analysis procedure breakdown seen at very low temperatures………………………………………………………………………………139
Figure 3.20: Secondary structure content percentage at 1652 cm\(^{-1}\) (α–helix) and 1627 cm\(^{-1}\) (β-sheets) for lysozyme.

Figure 3.21: Changes in absorbance level at 1652 cm\(^{-1}\) (α–helix) for thermal unfolding of lysozyme in 40 wt% EAN, 33 wt% sucrose and 27 wt% D\(_2\)O.

Figure 3.22: Comparison of DSC up scans of solutions quenched from 30 °C and 86 °C of various lysozyme concentrations.

Figure 3.23: DSC quench-and-refold scans of 25 wt% lysozyme in different ionic liquid and carbohydrate mixtures.

Figure 3.24: Comparison of the viscosity of different mixture of water, sucrose, glucose and EAN.

Figure 3.25: RAW ITC and normalized excess heat release of lysozyme and EAN binding.

Figure 3.26: DSC heat scans after DSC quench and LN\(_2\) quench of [A] 40 wt% EAN, 27 wt% H\(_2\)O and 33 wt% sucrose and [B] 60 wt% sucrose and 40 wt% H\(_2\)O.

Figure 3.27: DSC heat scans after DSC quench and LN\(_2\) quench of [A] 20 wt% lysozyme, 32 wt% EAN, 21 wt% H\(_2\)O and 27 wt% sucrose [B] 40 wt% EAN, 27 wt% H\(_2\)O and 33 wt% sucrose [C] 47 wt% EAN, 19 wt% H\(_2\)O, 34 wt% sucrose.

Figure 3.28: Visual comparison of the qualitative secondary structure for lysozyme in two different solvent composition of EAN, sucrose and D\(_2\)O.

Figure 3.29: Effect of annealing quenched samples of 20 wt% lysozyme, 38 wt% EAN, 27 wt% sucrose, and 15 wt% water.
Figure 3.30: Qualitative changes for the refolding of α–helical secondary structure for 2 wt% lysozyme, 39 wt % EAN, 33 wt% sucrose and 26 wt% D2O after thermal anneals.................................................................157

Figure 4.1: DSC heat up scans of 20 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose.................................................................172

Figure 4.2: DSC solvent background subtracted heat scans for 20 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose after LN2 quench of folded and unfolded protein.................................................................173

Figure 4.3: Second derivative and IR fitted spectrum of native state RNase A: 2 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose..............174

Figure 4.4: IR spectra of 2 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose before quench, after fast quenching with liquid nitrogen and slow quenching at 20 °C/min from room temperature.................................176

Figure 4.5: Successive quench-and-refold spectra of 2 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose.............................................177

Figure 4.6: Qualitative and quantitative IR analyses for α–helical and β-sheet secondary structural changes of 2 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose.........................................................179

Figure 4.7: DSC heat scans for 20 wt% RNase A in 10 mol EAN and no EAN matrix compositions.................................................................182

Figure 4.8: DSC solvent background subtracted heat scans for 20 wt% in 10 mol EAN and no EAN matrix compositions, LN2 quenched from folded and denatured protein..........................................................184
Figure 4.9: Native state IR spectra for RNase A in no EAN and 10 mol EAN solution matrices at 25 °C……………………………………………………………………………186

Figure 4.10: Qualitative IR analyses for α–helix (1652 cm⁻¹) and β-sheet (1632 cm⁻¹) secondary structural changes of RNase A in varying compositions of EAN, sucrose and D₂O……………………………………………………………………………187

Figure 4.11: Quantitative IR analyses for α–helix (1652 cm⁻¹) and β-sheet (1632 cm⁻¹) secondary structural changes of RNase A in varying compositions of EAN, sucrose and D₂O……………………………………………………………………………188

Figure 4.12: DSC heating and cooling scans of 20 wt% RNase A in 20 mol EAN, 10 mol EAN, and no EAN solvent compositions………………………………………………………191

Figure 4.13: Changes in absorbance level of β-sheet (1632 cm⁻¹) and DSC scans for thermal unfolding of RNase A in 20 mol EAN, 10 mol EAN and no EAN solution matrices……………………………………………………………………………193
List of Tables

Table 1.1: Amide I band frequencies and assignments for secondary structure components of protein in D$_2$O and H$_2$O solutions........................................33

Table 1.2: Other protein folding experimental techniques........................................35

Table 2.1: Quantitative percentage estimates of secondary structure content for lysozyme using different experimental techniques.................................56

Table 2.2: Density of various trehalose and water solutions.......................................58

Table 2.3: Aggregation percentages for 30 mg/ml lysozyme in various trehalose concentrations determined by enthalpy changes from DSC scans in Figure 2.3.................................................................66

Table 2.4: Secondary structure content for lysozyme in various trehalose water concentrations at 25 °C.................................................................68

Table 2.5: Comparison of percent unaggregated protein to percent of recovered activity.................................................................77

Table 2.6: Aggregation and inactivation rate constants for lysozyme in various trehalose concentrations.................................................................83

Table 3.1: Measured density of solutions utilized in this research.............................103

Table 3.2: Melting temperature, enthalpy and aggregation percentage from Figure 3.7 DSC heating scans.................................................................121

Table 3.3: Comparison of protein secondary structures determined by Fourier transform infrared spectroscopy to previous studies.................................128

Table 3.4: Amide I’ band frequencies and assignment for protein secondary structure in D$_2$O.................................................................135
Table 3.5: Constraints for subsequent quantitative curve fitting of quench-and-refold spectra taken from native state curve fitting for lysozyme………………136

Table 3.6: Sample secondary structure percentages determined by curve fitting of quench-and-refold spectra………………………………………………139

Table 3.7: Lysozyme unfolding and refolding energies from annealing DSC experiments shown in Figure 3.29…………………………………………………154

Table 4.1: Comparison of secondary structure composition of RNase A in 40 wt% EAN, 27 wt% D2O and 33 wt% sucrose at 25 °C to literature values……………………………………………………………………………….175

Table 4.2: Constraints for subsequent quantitative curve fitting of quench-and-refold spectra taken from native state curve fitting for RNase A……………178

Table 4.3: Comparison of thermal properties obtained from DSC scans for RNase A in varying solvent compositions…………………………………183

Table 4.4: Secondary structure composition of RNase A in various solvent matrices at 25 °C………………………………………………………………….185

Table 4.5: Summary of secondary structural content for RNase A in various solvent compositions during quench-and-refold……………………………190
Chapter 1

Introduction

Four classes of molecules constitute the main components in living systems: proteins, carbohydrates, lipids, and nucleic acids. Of these components, proteins have the widest diversity in structure and function. They are amazingly versatile molecules that make chemical reactions to form the basis for life, transmit signals in the body, identify and kill foreign invaders, form engines that make us move and record visual images, among other biological roles. The name protein was proposed by Swedish scientist Jöns Berzelius in 1838, establishing the importance of protein as the word protein originates from the Greek word proteios, which means “standing in front” or “in the lead” [Eisenberg, 2002].

In addition to being biology’s workhorses, proteins are utilized in industrial, biomedical, agricultural and environment applications. Protein biotechnology is an ever-increasing field and has evolved and diversified. Researchers Walsh and Headon have stated that protein biotechnology is concerned with the commercial production and isolation of specific proteins from plant, animal or microbial sources, and/or the
subsequent utilization of these proteins in order to achieve a pre-defined biological event [Walsh and Headon, 1994]. The introduction of DNA recombinant technology has led to the production of large amounts of desired proteins.

In many industrial applications, the effective catalytic property of enzymes is the main desired feature and the majority of enzymes utilized in industry are hydrolytic in action [Kirk et al., 2002]. Examples of such proteins include the use of proteases in detergent and dairy industries and carbohydrates in starch, textile, detergent and baking industries. In 2010, industrial enzymes had an estimated world market value of $3.3 billion [BCC Research, 2011].

In medicine, proteins are employed for therapeutic purposes in treating human diseases, diagnostic purposes, as well as catalysts for the synthesis of small molecule drugs [Delagrave and Murphy, 2003]. Examples of therapeutic recombinant proteins employed in medicine include hormones, cytokine, vaccines, blood factors and monoclonal antibodies. Although proteins purified from tissue have been used to treat disease for a longer time, it was not until 1982 that the first recombinant protein was approved for medical use. As of 2008, more than 130 protein therapeutics are used and approved in the world [Leader et al., 2008].

Biotechnology in agricultural science is more often associated with the change of a crop’s genotype to make it more resistant, for example to pesticides, than the production of proteins. However the development of nuclear transfer technology provides the ability to make precise genetic changes for a better and more efficient production of pharmaceutically active proteins in transgenic animals [Wilmut et al., 1997; Bulfield, 2000].
The use of enzymes is in itself environmentally friendly. Proteins are naturally occurring molecules and are degradable by nature. Traditional chemical-technical processes often produce a wide range of undesirable or harmful by-products and/or waste [Willke and Vorlop, 2004]. However, due to the specificity of enzymes, fewer by-products are produced in processes that employ enzymes. Additionally, the enzyme’s functional properties can be applied to the actual breakdown or withdrawal of harmful compounds [Gupta et al., 2005].

Although proteins are very useful in many different applications, the number of applications employing proteins is limited by the fact that the folded and functional state of a protein is only marginally stable. In fact, the free energy difference between the protein’s native and unfolded states, $\Delta G_{N \rightarrow U}$, is typically of the order of 10 kcal/mol, which is equivalent to two to three hydrogen bonds [Pace et al., 1996]. Therefore, even small changes in solution conditions, such as temperature and pH, can be detrimental to the conformation and functionality of proteins.

Thus, processes that employ enzymes need to operate at moderate temperatures and other solution conditions. Processes run at higher temperatures are favorable, as increasing temperature increases enzymatic activity, however, a process temperature that is above the enzyme’s unfolding temperature decreases its activity dramatically [Fullbrook, 1996]. Increasing the stability of enzymes is of major technological importance as it would increase the number of possible applications and allow for existing process to operate under more favorable conditions. The shelf-life and stability of many new pharmaceutical products are major concerns for drug manufacturers because of the sensitivity of proteins to environmental conditions. For economic
viability, a typical pharmaceutical product requires a shelf life of 18-24 months [Cleland et al., 1993]. Over time, these proteins are prone to denature by hydrolysis, unfolding, aggregation or chemical reaction [Wang, 1999]. Thus, one of the most challenging tasks is to develop more effective and economical methods for preserving labile biomolecules against physical and chemical degradation.

Researchers have attempted to study and understand how to stabilize proteins against different environmental factors including pH [Kapetaniou et al., 2006], organic solvents [Zhen, 2005], and pressure [Meersman et al., 2006]. In this dissertation, focus is placed on the thermal stability of proteins. Researchers have adopted two strategies, both inspired by nature, to increase the stability of proteins against changes in temperature. The first strategy is to use site-directed mutagenesis, which alters the amino acid sequence of the protein. Understanding and knowledge from studies of organisms that are capable of living at extreme temperature [Vielle an Ziekus, 2001, Jaenicke and Bohm, 1998] and proteins that function normally at temperature near or above 100°C are utilized [Huber et al., 1989, Adams, 1993]. Several different mechanisms for the stability of hyperthermophilic enzymes have been proposed, such as the burial of more hydrophobic surface area upon folding. One example is the stabilization of lysozyme by increasing the amount of buried hydrophobic surface area [Tankano et al., 1998]. However, caution must be used to ensure that protein activity is not adversely affected by changes in amino acid sequencing.

The second approach to increase protein stability is to alter the protein’s solvent environment. Researchers have found that the addition of certain small molecules to the solution can act as a folding enhancer and stabilize the native state of the protein and/or
suppress aggregation by reducing protein-protein interactions [Taravati and Ebadi, 2007]. Examples of such additives researched and currently used include: sucrose, sorbitol, arginine, glycerol and spermine [Hirano et al., 2007; Lee et al., 2006; O’Connor et al., 2004; Petersen et al., 2004; Arakawa and Tsumoto, 2003; Ueda et al., 2001; Valente et al., 2005]. Sugars and polyols are the most commonly used non-specific protein stabilizers [Wang, 1999]. Though previous research has shown that the addition of osmolytes can increase protein stability, continuing research is being conducted to fully understand the mechanisms.

Aside from the positive uses of stabilized and functional proteins for biotechnical applications, proteins \textit{in vivo} (in the biological environment in which they perform a certain task) can misfold, resulting in loss of functionality or becoming toxic to the cell [Kelly, 2002; Sorgjerd et al., 2008]. Proteins are synthesized in ribosomes as linear chains of amino acids forming polypeptides in a specific order dictated from the information encoded in the cellular DNA. In order to function, it is important for the polypeptide chain to fold into a unique native three-dimensional structure that is characteristic for each protein. Also, only correctly folded proteins have long-term stability in crowded biological environments. Therefore, it is well known that misfolded proteins or proteins that fail to remain in their correctly folded state tend to form aggregates and/or interact improperly with the cell, leading to cellular stress and cell death. Diseases associated with misfolding include genetic, sporadic and even infectious ailments. Examples include Alzheimer’s disease, cystic fibrosis, Creutzfeldt-Jakob disease and familial amyloidotic polyneuropathy [Dobson, 2002; Hammarstrom et al., 2002; Thomas et al., 1995].
Experimental research conducted to elucidate how proteins fold has mainly focused on answering the following fundamental questions: what is the structure of the protein intermediates, what implications are there for folding pathways, and what are the barriers to the folding process, typically measured through mutational and kinetic studies. In this dissertation, we focus on the study of the protein folding mechanism in vitro. It should be noted that protein folding in vitro and in vivo are quite different. In the living cell there are a number of proteins and systems that assist the folding and take part in the rescue and turnover of misfolded proteins. Examples of such proteins include peptidyl prolyl isomerase (PPI), protein disulfide isomerase (PDI) and various chaperones [Bukau et al, 1996]. However, even if auxiliary factors assist protein folding in vivo, understanding unassisted folding is needed to properly evaluate the roles of such factors, thus the focus on in vitro experiments. Though the folding mechanism of proteins in vitro are still not fully understood, in vitro experiments performed on many different proteins have revealed several possible folding mechanisms, which will be outlined later in the chapter.

The long-term goal of this research is to provide further insight into protein stability by excipients and to capture experimentally the structural changes and intermediates that occur during protein folding. Understanding how proteins fold is key to being able to predict structure from sequence, which is critical for interpretation and utilization in genome sequencing work. This knowledge should also help towards understanding the factors that stabilize a protein’s three-dimensional structure, which could provide a basis to rational design and modification of new proteins with novel and exciting functions that are not provided by nature.
This dissertation contains two main studies. The first part focuses on the investigation of protein aggregation when subject to thermal denaturation. Irreversible protein aggregation in the absence and presence of sugar was determined by a number of experimental techniques to provide further understanding of protein stability by excipients. The study also includes a high protein concentration, which is desirable for economic reasons in pharmaceutical drug production. Such high protein concentrations are not used in practice at present, as aggregated protein is then typically present and generally not accepted for product approval [Wang, 1999].

The second part is the introduction of a new experimental technique to study protein folding mechanisms by capturing intermediate structures during protein refolding using a quench-and-refold approach with infrared spectroscopy. We present an experimental technique for studying structural changes during protein folding in real time. In contrast to conventional protein folding studies, we replace the independent variable time with temperature. Ideally, protein folding in real time should be studied at fixed temperature, however the fast folding times of proteins limit detection by conventional experimental techniques [Eaton et al., 2004]. The solution we applied is to conduct measurements of the time-evolution of protein structure at low temperatures with controlled annealing. A technique was developed for trapping a protein in an unfolded, high entropy, high enthalpy state, at low temperatures. Taking advantage of kinetic traps or local minima in the system energy landscape and the decrease in entropy of the system in the glassy state, trapping of the protein is achieved by quenching the sample. Infrared (IR) spectroscopy is then used to observe normally occurring structural changes at temperatures above and below the glass transition. This is done by collecting spectra of
an unfolded protein undergoing controlled annealing across the matrix’s glass transition temperature. This experimental technique will permit researchers to study specific refolding intermediates, or even misfolded intermediates, on arbitrarily slow time scales.

In the remainder of this chapter, we review some background information relevant to this dissertation and to the study of protein stability and protein folding mechanisms. These reviews include the properties of a glassy state, protein stability both in solution and in the glassy state, driving forces of protein folding, classic models and new viewpoints of protein folding mechanisms, and current experimental techniques used to study proteins. We conclude this chapter with an overview of this dissertation.

1.1 Glassy State Properties

A glass is an amorphous material with mechanical properties similar to that of a solid, but without the associated long-range crystalline order. The most common method for producing glassy materials is by lowering the temperature or quenching a viscous liquid fast enough to avoid crystallization. This supercooling causes molecular mobility to decrease. Eventually, at sufficiently low temperature, the mobility of the molecules is not detectable on an experimental timescale, where characteristic relaxation times are on the order of 100 seconds. Due to rapid cooling, there is insufficient time for molecules to order themselves into a long range ordered crystalline configuration, resulting in the formation of a glass.

In the transition of a sample from a liquid to a glassy state, the rate of changes for volume and enthalpy with respect to temperature (the thermal expansion coefficient ($\alpha$)
and the isobaric heat capacity ($C_p$), respectively) exhibit a large but continuous decrease towards a value close to that of a crystalline solid. This is shown in Figure 1.1.

![Diagram](image)

**Figure 1.1:** Top diagram reveals the isobaric relationship between volume and temperature in the liquid, glassy and crystalline state with $T_m$ as the melting temperature and $T_{ga}$ and $T_{gb}$ as the glass transition temperature corresponding to (a) low and (b) high cooling rates. The lower diagram shows the abrupt but continuous change of the thermal expansion coefficient, $\alpha_p$, at $T_g$ [Debenedetti, 1996].

The intersection of the liquid and vitreous portions of the volume versus temperature curve provides one definition of the glass transition temperature, $T_g$. In general, this temperature occurs around $2T_m/3$, where $T_m$ is the melting temperature. The
temperature observed at the onset of endothermic heat flow increase is another way to
determine the $T_g$. Another definition of $T_g$ is the temperature where the shear viscosity
reaches $10^{13}$ poise. The behavior, as illustrated in Figure 1.1, is not a true phase
transition since it does not involve discontinuous changes in any physical property.

$T_g$ is considered an important characteristic of many materials. $T_g$ is found to
decrease at slower cooling rates, as shown in Figure 1.1, because there is more time
available for configurational sampling at each temperature, however the slower cooling
rate is still fast enough to bypass crystallization. This indicates that the properties of a
glass depend on thermal history. However, for practical purposes, the $T_g$ dependence on
cooling rate is modest, as $T_g$ changes only by 3-5 °C, even when cooling rates change by
an order of magnitude [Debenedetti and Stillinger, 2001; Ediger et al, 1996].

Supercooled liquids can be classified as strong or fragile liquids. The scheme
proposed by Angell, shown in Figure 1.2, plots the temperature dependence of viscosity
at atmospheric pressure in a scaled Arrhenius fashion [Angell, 1991]. The temperature is
scaled at the point where viscosity reaches $10^{13}$ Poise [Laughlin and Uhlmann, 1972].
For strong liquids, such as germanium dioxide ($\text{GeO}_2$), the viscosity exhibits an
Arrhenius behavior, whereas fragile liquids show significant deviation from Arrhenius
behavior. Strong liquids are said to have persistent short-intermediate order when heated
across the glass transition, and this is reflected in small changes in the thermal expansion
coefficient, $\alpha$. Fragile liquids, on the other hand, exhibit a large change in $\alpha$ due to
largely non-directional, dispersive intermolecular forces.
Figure 1.2: Strong and fragile liquids characterized by their Arrhenius and non-Arrhenius behavior, respectively, in the temperature dependence of their structural relaxation time (here represented by the viscosity). Insert depicts changes in heat capacity for various liquids across the glass transition [Angell, 1991]

### 1.2 Protein Stability

Proteins can be described in terms of four hierarchical levels of structure: primary, secondary, tertiary and quaternary. The primary structure is the sequential order of amino acids, containing both amino and carboxylic functional groups, in the polypeptide chain. The secondary structure is composed of repetitive 3-dimensional conformational units. A key to this structure is hydrogen bonding between the amide protons and carbonyl groups of specific pair residues. The secondary structure includes
most frequently α-helix or β-sheet structures. In an α-helix, the structure is stabilized by a series of hydrogen bonds from C=O of residue n to N-H of residue n+4, while in the β-sheet, the structure is stabilized by hydrogen bonds between C=O and N-H groups of two adjacent β-strands. Very often, the structural units between elements of secondary structures constitute well-defined and non-repeating turn structures. The tertiary structure is the overall 3-dimensional configuration of the polypeptide chain, which is stabilized by interactions between the side chains of the residues. The tertiary structure can be divided into one or more subdomains that constitute local parts with specific tertiary structure elements. The quaternary structure is the overall configuration adopted by proteins consisting of more than one polypeptide, and is stabilized by interaction between the polypeptides.

The unique structure and stability of a protein is an important factor to its proper biological functionality. Unstable proteins can unfold and assemble into aggregates that contain nonnative protein structure, often containing high levels of nonnative, intermolecular β-sheet structure [Dong et al., 1995]. Aggregation is often irreversible. As mentioned earlier, stable and correctly folded proteins are essential for biology, industrial, biotechnical and medical applications. In this section a brief exploration of how matrix conditions affect protein stability in both the aqueous and glassy state is provided.

### 1.2.1 Proteins in Solution

A protein in aqueous solution is in equilibrium between the native, folded conformation and the ensemble of its denatured, unfolded conformations. It is assumed
in what follows that the latter can be lumped into a single denatured “state”. The stability of the native state is based on the difference in magnitude of the Gibbs free energy (ΔG) of the system between the native and denatured state; the temperature dependence of the protein’s stability is determined by the corresponding entropy change, ΔS. A positive ΔG indicates that the native state is more stable than the denatured state. More positive ΔG indicates greater stability in the native state. For a protein to unfold, stabilizing forces need to be broken. Conformational entropy overcomes stabilizing forces allowing the protein to unfold at temperatures where entropy becomes dominant [Dill, 1990].

It is well known that incubating protein solutions at high temperatures can perturb the native protein confirmation leading to unfolding and physical degradation [Carpenter et al., 1999]. Though thermally induced degradation for some proteins is reversible, there are many cases in which high temperatures lead to aggregation or irreversible degradation. Some examples include lysozyme [Ueda et al., 2001], streptokinase [Azuaga et al., 2002] and albumin [Rondeau et al., 2010]. One interesting observation is that aggregation starts at temperatures below a protein’s denaturation temperature, suggesting that aggregates are not necessarily formed from fully unfolded proteins [Dong et al., 1995]. The Lumry-Eyring framework has been used to describe the aggregation pathway of many proteins [Lumry and Eyring, 1954; Sanchez-Ruiz, 1992]. This framework consists of a conformational change of a protein that is reversible, followed by an irreversible assembly of nonnative structures to form aggregates. Consistent with expectations from reaction kinetics, experiments have shown that temperature greatly
affects aggregation kinetics, as rate constants increase exponentially with temperature [Atkins, 1994].

The solution pH also has a strong influence on aggregation rate, as proteins typically have a narrow pH range of stability against aggregation. This is due to the fact that the solution pH determines the charge of the protein, which affects electrostatic interactions. Nonspecific repulsions from charged groups on a protein as well as specific charge interactions, such as salt bridges can affect the protein conformational stability [Dill, 1990]. Aside from pH effects on protein confirmation, changes in protein molecule charges could give rise to highly attractive protein-protein interactions, which could in turn make aggregation energetically favorable [Chi et al., 2003]. Insulin [Nielson et al., 2001] and relaxin [Nguyen and Shire, 1996] are just two examples where studies show the effects of pH on protein aggregation.

The effects of cosolutes and ligands on protein stability can be explained by the Wyman linkage function and other theories, such as preferential exclusion [Timasheff, 1998]. In a two-state equilibrium, the Wyman linkage function states that differential binding of a cosolute or ligand will shift the equilibrium towards that of greater binding. For example, binding of zinc to human growth hormone increases the free energy of unfolding, thus favoring the native state [Cunningham et al., 1991]. Despite the weak and short lived interactions between protein denaturants, such as urea, and proteins, previous research revealed a larger number of binding sites exposed in the unfolded state, thus favoring the denatured state [Liepinsh and Otting, 2004]. Protein stabilizers, on the other hand, are preferentially excluded from the protein surface and can be said to exhibit negative binding. The greater the solvent-exposed surface area, as in the unfolding of
protein, the greater degree of preferential exclusion and negative binding, thus the native state is favored [Timasheff, 1998]. Many studies have shown that ligands and cosolutes can also influence the rate of aggregation. For example, the addition of sucrose inhibits aggregation of immunoglobulin, whereas the addition of urea accelerates aggregation [Kim et al., 2000].

The salt type and concentration in a protein solution have varying effects on protein structural stability, equilibrium protein solubility and the rate of aggregation. One example is the effect of salts on bovine serum albumin. At high ionic strength, albumin is stabilized by kosmotropic salts, but destabilized by chaotropic salts [Yamasaki et al., 1991]. Salts are known to bind to proteins, where ions interact with unpaired side chains on the protein surface as well as with peptide bonds. As mentioned previously, the protein state with the most salt binding is favored according to the Wyman linkage theory. The strength of electrostatic interactions between charged groups within the protein and between protein molecules are modulated by electrolytes. Intramolecular charge-charge interactions tend to affect the stability of the protein structure and intermolecular electrostatic interactions are known to affect the aggregation rate [Chi et al., 2003].

1.2.2 Proteins in a Glass Matrix

Another approach to stabilizing proteins is to place them in a glassy matrix. This is mainly achieved by lyophilization (freeze-drying), vacuum drying, and or quenching of proteins in a carbohydrate mixture [Franks, 1998; Wang, 1999]. Since glasses have immeasurably slow molecular mobility on a laboratory timescale, we expect these
materials to significantly slow down or even inhibit chemical reaction and diffusion [Crowe et al., 1998].

Water has a strong plasticizing effect within a glassy matrix, thus enhancing molecular mobility. Increase in water content in a glass also causes the glass matrix to swell due to a weakening of intermolecular forces and an increase in free volume [Angell, 1995]. As a result, the glass transition temperature of the matrix is depressed as a lower temperature is now required to arrest molecular mobility. This is referred to as devitrification. For example, the glass transition temperature of 80 wt% trehalose in water is -27 °C, whereas a matrix of 60 wt% trehalose in water has a glass transition temperature of -80 °C [Miller et al., 1997].

1.3 Protein Folding Mechanism

Today, we have a thorough understanding of how cells construct sequences of amino acids using the DNA as a template. There are also methods to analyze the amino acid contents in specific proteins, and methods to determine the sequence of the residues. However, there are still major uncertainties on how polypeptide chains are able to fold into well-defined structures necessary for proteins to perform their specific functions. In this section, we will briefly review the literature on the driving forces and proposed mechanisms of protein folding.

1.3.1 Driving Forces of Protein Folding

Extensive scientific research has been done to understand the driving forces behind protein folding. Hydrophobic effects (aversion for water of the nonpolar
residues) have been found to be the dominant driving force in folding and are supported by several experimental observations. First, nonpolar solvents are found to denature proteins [Singer, 1963] According to the hydrophobic mechanism, nonpolar solvents reduce the free energy of the unfolded state by solvating the exposed nonpolar amino acids. Secondly, protein stability is weakened at both high and low temperatures, just as nonpolar solutes exhibit solubility minima in water at intermediate temperatures [Privalov and Gill, 1988]. In addition, protein crystal structures reveal that non-polar residues are present predominantly inside the core of the folded protein and polar residues are located at the surface [Dill, 1990].

While the free energy of unfolding due to hydrophobicity was determined to be 100-200 kcal/mol, the actual protein free energy of unfolding was measured at 5-20 kcal/mol [Privalov and Gill, 1988; Pace et al., 1996]. This discrepancy indicates that there is an opposing force nearly equal to the hydrophobic driving force. This opposing force is entropic, since unfolding leads to an increase in entropy [Northrop, 1932; Mirsky and Pauling, 1936]. Marginal stability of globular proteins is a balance between driving forces and opposing forces. While hydrophobic driving forces are responsible for overcoming the entropic opposing force, other interactions define the confirmation [Dill, 1990].

Interactions/driving forces other than hydrophobic are classified as either long or short range. Electrostatic forces constitute one such long-range interaction. Electrostatic energy is dependent on the square of the net charge. When changing the pH of a solution increases the net charge on a native protein, the increased charge repulsions will ultimately destabilize the folded protein because the charge density is greater on a folded
protein than an unfolded one [Kauzmann, 1954; Tanford, 1961]. Thus, unfolding of the protein leads to a state of lower electrostatic free energy. Other interactions are ion pairing and salt bridging, both of which can increase protein stability. These forces come into play when amino acids of opposite charges are close enough to each other [Mirsky and Pauling, 1936]. Hydrogen bonding is a short-range attractive force, and occurs between an electropositive hydrogen atom and an electronegative atom to which the hydrogen is not covalently bonded, such as between C=O and NH groups of amino acids. Hydrogen bonding affects helix-coil transitions, which are characteristic features found in globular proteins. However, none of these interactions by itself plays the dominant role in protein folding [Dill, 1990].

1.3.2 Classic Versus New View of Protein Folding

Interest in protein folding was motivated by Anfinsen’s work in the early 1960’s, which showed that ribonuclease could refold reversibly into its unique and functional native state [Anfinsen et al., 1961, Haber and Anfinsen, 1962]. Subsequent experiments gave rise to a thermodynamic hypothesis, which states that the three-dimensional native structure of a protein, in its normal physiological environment regarding pH, ionic strength, temperature, etc., is a thermodynamically stable equilibrium state with the lowest Gibbs free energy [Epstein et al., 1963].

For a long period of time, ideas behind protein folding were dominated by the Levinthal paradox, which concluded that proteins must fold by specific “pathways”. According to the paradox, it is too time consuming for the unfolded protein to randomly search through and/or try the large number of possible confirmations in order to find its
lower free energy structure [Levinthal, 1968]. Using Levinthal’s argument, feasible protein folding mechanisms were divided into two classes. The first class consists of thermodynamically controlled (path independent) folding mechanisms, while the second consists of kinetically controlled (path dependent) mechanisms. Levinthal postulated a kinetic control of protein folding, thus prompting a search for kinetic folding pathways [Levinthal, 1968].

At present, there are two main views regarding protein folding: a classical and a new view. The classical view is based on simple kinetic models. These simple kinetic models are determined from single or multi-exponential time decays of protein structural changes between the native (folded) and the denatured (unfolded) state. Kinetics are labeled as “two-state” when a single exponential decay is observed in the folding and unfolding directions, therefore leading to an assumption that there are only two states, a native state N and a denatured state D. Typically, small single domain proteins, with 100 amino acids or less, have been found to fold without an accumulation of detectable intermediates [Jackson, 1998]. “Multi-state” folders exhibit additional exponential time decays, due to the presence of intermediate conformations. Intermediates are normally found for proteins that are relatively large (more than 100 amino acids) and have molten globule state characteristics. The molten globule state, which is considered a key kinetic intermediate, is quite compact with native-like secondary structure [Arai and Kuwajima, 2000; Ptitsyn, 1995]. The following are the three most important classical models [Dill and Chan 1997].
20

\[ U \leftrightarrow N \]
\[ \downarrow \]
Off-pathway model:
\[ X \]

On-pathway model:
\[ U \leftrightarrow X \leftrightarrow N \]

Sequential model:
\[ U \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow \ldots \leftrightarrow N \]

where \( U \) is the fully unfolded state, \( N \) is the native state and \( X \) or \( I \) represent intermediate states. The off-pathway model correlates to two-state kinetics, since intermediates that are formed are considered uninteresting dead ends and do not affect the single exponential decay observed in the folding and unfolding directions between the native and unfolded state. On the other hand, both the on-pathway and sequential models correlate to multi-state kinetics due to one or more intermediate confirmations present in the folding and unfolding directions between the native and unfolded state. Models are chosen based on the “best fit” of experimental rate constants and amplitudes. For example, Demetrius found that single domain proteins, which lack disulfide bridges, are described by two-state kinetics. Also, for the proteins that exhibit two-state kinetics, a positive correlation was revealed between folding rates and thermodynamic stability of the folded state [Demetrius, 2002].

The contemporary view of protein folding replaces the sequential event pathway with a funnel concept of parallel events [Dill and Chan, 1997; Wolynes et al 1995]. This new view arises from advances in both experiment and theory. Experimental advances such as high resolution proton and deuterium NMR, mass spectroscopy, mutational
studies and fast laser-triggered methods provide much more detailed information about protein structure, even down to the atomic level. Moreover, these techniques are able to detect early events in folding. New theories based on highly simplified statistical mechanics, mostly lattice-based, also contribute to this new view [Chan and Dill, 1996]. While the classical view represented specific states (N, D, I), statistical mechanical models recognize that protein states consists more of a distribution, an ensemble of individual chain conformations.

The folding funnel is a conceptual mechanism for folding that avoids the Levinthal paradox. The top of the funnel consists of proteins in random states. These states have a relatively high enthalpy and entropy. This funnel has opposing forces due to maximization of entropy (keeping protein unfolded at the top) and minimization of enthalpy (pulling the protein down the funnel towards the native state). These funnels are also termed as free energy landscapes, a hyperspace representation of conformation free energy as a function of the degrees of freedom (e.g. dihedral bond angles). Figure 1.3 provides a schematic outline of a folding energy landscape.

The vertical axis denotes the free energy. This free energy includes contributions from hydrogen bonds, ion pairs, and hydrophobic energies, among others. These individual energies are also affected by external conditions (e.g. temperature). The lateral axis denotes the conformational coordinates. Each point on the surface corresponds to a microstate of the system. Two points that are close to each other are geometrically similar.
Figure 1.3: Schematic outline of a folding energy landscape. The width of the funnel represents the entropy, and as the protein approaches the native state the funnel becomes narrower, and the entropy decreases. Denatured molecules at the top of the funnel may fold by a myriad of different routes, some of which involve transient intermediates (local energy minima), whereas others involve significant kinetic traps (misfolded states) [Brooks et al., 1998].

As Figure 1.4 shows, energy landscapes have many features, including hills and valleys. Hills correspond to high-energy confirmations, while valleys are local minima or traps. The features of the surface determine the native state, which is given by the global minimum. According to this new view, a two state folding protein’s energy landscape would consist of a smooth funnel with a single minimum corresponding to the native state, while a multi-state protein folder may have a “bumpy” landscape with numerous local minima corresponding to different intermediates and a single global minimum for the native state. Figure 1.4 below represents a hypersurface for a two-state and multi-state folding protein.
However, continuing experiments are necessary to fully characterize the energy landscape of folding, and to identify intermediates during the folding process. Further research would provide insight into the reasons for incompletely folded proteins. Also, there still remain questions regarding the possibility of combining the classical view and the new view of protein folding.

1.3.3 Structural Mechanism of Protein Folding

Scientists over the past 40 years have been searching for a model describing how the structural parts in the three-dimensional native state of a protein are formed and assembled during the folding process. Several different structural mechanisms have been proposed and are discussed below and it is unlikely that a single mechanism for protein folding exists. Figure 1.5 shows a visual representation of four models of protein folding: framework, hydrophobic collapse, classic nucleation-growth and nucleation-condensation.
The framework model proposes that hydrogen bonded secondary structure could form first, independently of tertiary structure [Ptitsyn and Rasin, 1975]. These elements would then diffuse and collide until successfully adhering and coalescing to form the tertiary structure in the native state [Karplus and Weaver, 1976].

The classic nucleation-growth model postulates that several neighboring sequence residues would form native secondary structure that would act as a nucleus from which the native structure would propagate in a step-wise manner. Tertiary structure would form as a necessary consequence from secondary structure [Wetlaufer, 1973].

In the hydrophobic collapse model, folding is thought to be driven by water molecules being squeezed out from the hydrophobic parts of the polypeptide chain. This
phenomenon will cause the chain to collapse rapidly around its hydrophobic side chains and rearrange into secondary structural elements from the restricted conformational space. In this case, the secondary structure would be directed by native-like tertiary structure [Dill, 1985].

According to the nucleation-condensation model, the folding process starts with a rapid random search of conformation in the unfolded states. Native-like secondary structures in the unfolded peptide become stabilized by long-range interactions as folding proceeds. When sufficient interactions have been established, the transition state is reached and rapid formation of the final structure ensues. Folding of the protein would slow down if conformations are formed with non-native interactions, which have to be disrupted before the transition state is reached due to the loss of stabilization energy [Neira and Fersht, 1999]. This model contrasts with the classic nucleation-growth model as this model has a folding nucleus that is rather diffuse.

1.4 Current Experimental Techniques for Protein Studies

Many different experimental techniques are applied to study the characteristics and properties of proteins. These same experimental techniques can also be utilized to test and gain insight into the stability and reversibility (folding) of proteins when subjected to different environmental conditions, such as changes in temperature, pH and solution medium. Below is a general introduction to the experimental techniques utilized often in our research: differential scanning calorimetry, ultraviolet spectroscopy and infrared spectroscopy. A brief overview of other experimental techniques and approaches for protein studies is also provided.
1.4.1 Differential Scanning Calorimetry (DSC)

DSC is a thermal analysis technique. The power compensated differential scanning calorimeter contains two electrically heated cells, where one cell holds the sample, while the other holds an inert reference material, or is left empty. DSC measures the heat required to maintain the temperature difference between the two cells, as the two cells are subjected to identical heating and cooling rates and the sample undergoes a phase transition. The amount of heat flow required for the sample is dependent on whether the process is exothermic or endothermic. For example, as a solid melts to a liquid, it will require more heat flowing to the sample in order to increase its temperature at the same rate as the reference cell. This is due to the sample absorbing heat as it undergoes this endothermic phase transition. Data results from a DSC experiment present a curve of heat flux versus temperature or time. Figure 1.6 reveals the general characteristic features of a DSC curve. A positive peak generally represents endothermic processes, and exothermic processes generally give rise to a negative peak. Subtle changes, for example a glass transition, are represented by a continuous step change.

The DSC curve provides transition temperatures, which are determined at the lowest or highest point of the peaks for exothermic or endothermic transitions, respectively. The DSC curve can also be used to calculate and measure the enthalpy, $\Delta H$, and heat capacity, $\Delta C_p$, change for the transition. The enthalpy is calculated by integrating the area under the peak.
DSC is a powerful method in studying protein stability and reversibility. For proteins, the thermally induced process detectable by DSC is the structural melting, unfolding or denaturation of the molecule. The transition from the native state to a denatured conformation is represented by an endothermic peak, and is due to the rupture or inter- and intra- molecular bonds [Ma and Harwalkar, 1991]. DSC can determine two important protein parameters: the protein denaturation temperature, $T_d$, and the enthalpy of denaturation, $\Delta H_d$. 

The denaturation transition midpoint, $T_d$, is the temperature where 50% of the protein is in its native conformation and the other 50% is denatured, and correlates with the content of ordered secondary structure. At $T_d$, the change in Gibbs free energy between native and unfolded protein is null, $\Delta G = 0$, which also implies that $\Delta H = T\Delta S$, where $\Delta S$ is the change in entropy [Cooper, 1999]. The denaturation temperature is a measure of the thermal stability of proteins and has been found to be influenced by the
DSC heating scan rate [Ruegg et al., 1977] and protein concentration [Wright, 1984]. The higher the $T_d$, the more stable the molecule [Kaushik and Bhat, 2003]. An assessment of the calculation of $\Delta H_d$ is provided in [Cooper, 1999; Privalov and Kchechinashvili, 1972]. $\Delta H_d$ is shown by the shaded area under the curve in Figure 1.7, a representative DSC denaturation curve for lysozyme.

Figure 1.7: A representative DSC scan showing the thermal information provided by DSC technique. The shaded area represents the unfolding enthalpy [Byrne and Angell 2008].

This $\Delta H_d$ value is a net value from a combination of endothermic, such as the disruption of hydrogen bonds [Privalov and Kchechinashvili, 1972], and exothermic processes, such as protein aggregation and the breakup of hydrophobic interactions [Arntfield and Murray, 1981; Jackson and Brandts, 1970]. From the same experiment DSC also measures the change in heat capacity ($\Delta C_p^d$) for denaturation. Heat capacity changes associated with protein unfolding are primarily due to changes in hydration of side chains that were buried in the native state, but become solvent exposed in the denatured state [Privalov and Makhatadze, 1992].
1.4.2 Ultraviolet Spectroscopy (UV)

When an atom or molecule absorbs energy, the electrons are promoted from the ground state to an excited state. Many molecules absorb ultraviolet light, which tends to excite the valence electrons. A simple correlation between measurable quantities that is extensively used in analyzing UV spectroscopic data for dilute solutions is Beers Law. This can be expressed as $A = abc$. Here, the absorbance, $A$, is directly proportional to the path length, $b$, and the concentration, $c$, of the absorbing species. $a$ is a constant of proportionality, otherwise known as absorptivity. Provided the absorptivity is known for the protein being studied, UV spectroscopy can be used to determine the protein concentrations of prepared solutions [Mach et al., 1995]. This detection of native protein concentration provides a useful approach to study protein reversibility and stability in various solutions or when subject to thermal treatment by comparing any changes in concentration.

UV spectroscopy is also used to determine the enzymatic activity of proteins. Enzymes are catalysts, and in a catalytic reaction, the amount of enzyme units present in the sample is proportional to the rate of product formation or reactant disappearance. In the reaction mixture, there is at least one component in the reaction mixture that absorbs a certain wavelength. For example, lysozyme is an enzyme that cleaves *micrococcus lysodeikticus* cells. Lysozyme activity is measured by the change in absorbance at 450 nm, $\Delta A_{450\text{nm}}$, using a suspension of *micrococcus lysodeikticus* as the substrate. One lysozyme unit will produce a $\Delta A_{450\text{nm}}$ of 0.001 per minute in a 2.6 ml reaction mixture at pH 6.24 and 25 °C [Shugar, 1952].
1.4.3 Infrared Spectroscopy (IR)

IR spectroscopy is an absorption method that employs radiation in the 1-100 μm wavelength range. Infrared radiation is not sufficient to induce valence electron transitions; instead it excites vibrational and rotational motions in molecules. The resultant spectrum obtained from the IR spectrometer is usually represented as the absorbance versus the wavenumber in cm⁻¹.

The main use of IR spectroscopy is to elucidate the secondary structure of a protein. Secondary structure of a protein or peptide, such as α-helix and β-strands are stabilized by hydrogen bonding between the NH and CO groups which influences the energy of three main vibrations: (1) the amide A band at 3300 cm⁻¹, which characterizes the N-H stretch vibration, (2) the amide I band at 1650 cm⁻¹, which characterizes the C=O stretch and (3) the amide II band at 1550 cm⁻¹ which characterizes the N-H bending/deformation [Surewicz et al., 1993]. IR spectroscopy can be used to identify the number of stabilizing hydrogen bonds, and the functional groups engaged in hydrogen bonding.

To ensure that the spectrum captures only variation in secondary structure of a protein, careful subtraction of the background solution peak is necessary. The intensity and the shape of protein IR bands can be greatly affected by the presence of water [Pevsner and Diem, 2001]. An option, to avoid the band shifts or intensity difference due to water (H₂O), is to use heavy water (D₂O). Since the IR spectrum is also dependent on molecular mass, the peak for D₂O is shifted away from principal protein bands (e.g. amide I) [Solomon and Lever, 1999], thus making detection and subtraction of protein peaks easier and more accurate.
Qualitative and quantitative analyses can be performed on the IR spectra; however, quantification is not straightforward and can present conceptual and practical problems. The IR spectrum of a protein usually consists of many overlapping component bands that represent different structural elements (e.g. alpha-helix and beta-sheets). The fundamental difficulty is due to contributing component bands that are usually broader than the separation between adjacent peak’s maxima. Therefore, several techniques are applied to enhance the resolution. One computational technique is Fourier deconvolution, which narrows the bands, without distorting the integrated intensities (areas) of individual component bands. However, one must be aware that deconvolution also amplifies the noise. Over-deconvolution (where a noisy peak is misinterpreted as a component band) should be avoided; therefore there is a limit to the resolution enhancement factor [Surewicz et al., 1993]. Second derivatives of the spectrum, by nine-Point Savitsky-Golay smoothing function, aids in revealing component peak position [Carpenter et al., 1998]. Once these peaks are determined, a Gaussian curve-fitting method can be used to determine a quantitative evaluation of the protein’s secondary structure. Figure 1.8 provides a visual sample of an IR spectrum with Gaussian curve-fitting and its respective secondary derivative spectrum [Natalello et al., 2005].

Fractional areas of the component bands are assumed to represent the percentage of their respective assigned structures in the protein. This has been shown to provide comparative estimates with other structural techniques, such as NMR and X-ray crystallization [Byler and Susi, 1986; Kong and Yu, 2007].
Figure 1.8: A representative second derivative spectrum and an amide I curve fitting into Gaussian components [Natalello et al., 2005].

IR is a powerful technique for structure determination. It is not limited by the requirement of a well ordered crystal as needed for X-ray crystallography, or the size of the protein, which for NMR studies can only be applied to small proteins due to the complex nature in interpretation of NMR spectra. IR can also be applied to the study of proteins in both aqueous and non-aqueous media [Haris and Severcan, 1999]. To date, the best-studied and most sensitive spectral region is the amide I band. Table 1.1 lists the assigned secondary structures within the amide I band [Kong and Yu, 2007].
Table 1.1: Amide I Band Frequencies and Assignments for Secondary Structure of Proteins in D$_2$O and H$_2$O Solutions$^a$.

<table>
<thead>
<tr>
<th>H$_2$O</th>
<th>D$_2$O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Frequencies</td>
<td>Assignment</td>
</tr>
<tr>
<td>1624±1.0</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1627±2.0</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1633±2.0</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1638±2.0</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1642±1.0</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1648±2.0</td>
<td>Random</td>
</tr>
<tr>
<td>1656±2.0</td>
<td>α-Helix</td>
</tr>
<tr>
<td>1667±1.0</td>
<td>β-Turn</td>
</tr>
<tr>
<td>1675±1.0</td>
<td>β-Turn</td>
</tr>
<tr>
<td>1680±2.0</td>
<td>β-Turn</td>
</tr>
<tr>
<td>1685±2.0</td>
<td>β-Turn</td>
</tr>
<tr>
<td>1691±2.0</td>
<td>β-sheet</td>
</tr>
</tbody>
</table>

$^a$Kong and Yu, 2007

IR can provide a sensitive and diagnostic tool for monitoring structural changes of a protein, particularly under thermal denaturation. When monitored by a spectroscopic signal, the thermal unfolding transition of protein can often be modeled with two-state thermodynamics. However, that does not imply that a protein that exhibits spectral two-state equilibrium folding-unfolding behaviors folds without any detectable intermediates [Wang et al., 2005]. Infrared spectroscopy can also avoid limitations in saturation and subtraction of spectral absorbance and is also a widely used technique in protein aggregation studies [Kong and Yu, 2007]. In theory, infrared spectroscopy could be used to monitor folding and unfolding process of the protein under chemical denaturation, however obtaining good IR spectra has met with limited success as high concentrations
of chemical denaturants is needed to ensure protein unfolding [Bowler et al., 1993]. Nonetheless, IR is an important tool for the structural study of protein stability and folding.

1.4.4 Other Experimental Techniques

In recent years, the development of a wide range of techniques has led to further insight into protein characteristics and the biophysics of protein folding, particularly for small, model proteins *in vitro*. This has been realized by development of experimental approaches, such as temperature jump, pressure jump, ultra-fast mixing and stopped flow mixing, with faster time scales of measurement and enhanced sensitivity. Approaches combining both experiment and simulation have also provided valuable insights, however there is still much to be learned, as comparison of results from simulations and experimental approaches remains challenging [van Gusteren et al., 2008]. Further enhancements are needed for both experimental techniques and theoretical tools in order to better refine and test their respective outputs for studies of larger proteins, chemically modified proteins and membrane proteins. Table 1.2 provides a compact list of other experimental techniques applied to study the structural diversity and properties of non-native states, folding pathways and mechanisms, and/or barriers to correct protein folding. Time scales, measurement details, as well as references are also listed for the different experimental techniques in Table 1.2.
<table>
<thead>
<tr>
<th>Technique</th>
<th>Time scale</th>
<th>Measurement Comments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorbance</td>
<td>≥ ns</td>
<td>Chromophore environment</td>
<td>Roder et al., 2006</td>
</tr>
<tr>
<td>Anisotropy</td>
<td>≥ µs</td>
<td>Shape and size of molecule from correlation time data</td>
<td>Royer, 2006</td>
</tr>
<tr>
<td>ANS (1-anilo-8-naphthalene sulfonic acid) binding</td>
<td>≥ µs</td>
<td>Aromatic surface area exposure</td>
<td>Royer, 2006</td>
</tr>
<tr>
<td><strong>Circular Dichroism (CD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Far UV CD</td>
<td>≥ µs</td>
<td>Secondary structure content</td>
<td>Kelly et al., 2005</td>
</tr>
<tr>
<td>Near UV CD</td>
<td>≥ µs</td>
<td>Aromatic residue packing</td>
<td></td>
</tr>
<tr>
<td><strong>Fluorescence correlation spectroscopy (FCS)</strong></td>
<td>≥ ps</td>
<td>Size and shape of molecule from diffusion time</td>
<td>Haustein and Schwille, 2007</td>
</tr>
<tr>
<td>Fluorescence resonance energy transfer (FRET)</td>
<td>≥ ps</td>
<td>Molecular ruler measuring distance between two fluorophores</td>
<td>Roder et al., 2006</td>
</tr>
<tr>
<td>Intrinsic tryptophan fluorescence</td>
<td>≥ ns</td>
<td>Tryptophan environment</td>
<td>Royer, 2006</td>
</tr>
<tr>
<td>Native-state hydrogen exchange</td>
<td>h</td>
<td>Metastable state detection and overall global stability</td>
<td>Krishna et al., 2004</td>
</tr>
<tr>
<td><strong>Nuclear Magnetic Resonance (NMR)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMR relaxation methods</td>
<td>ms</td>
<td>Nonrandom structure of denatured states and conformation exchange between different proteins</td>
<td>Dyson and Wright, 2005</td>
</tr>
<tr>
<td>Pulsed H/D exchange</td>
<td>≥ ms</td>
<td>Folding populations hydrogen-exchange properties or protection</td>
<td></td>
</tr>
<tr>
<td>Real-time NMR</td>
<td>&gt; min</td>
<td>Structural information by measure of chemical shifts</td>
<td></td>
</tr>
<tr>
<td>Protein Engineering</td>
<td>Depend on probe</td>
<td>Rate of folding and stability indicate role of individual residue in protein</td>
<td>Zarrine-Afsar and Davidson, 2004</td>
</tr>
<tr>
<td>Small-angle X-ray scattering</td>
<td>≥ µs</td>
<td>Radius of gyration where information of 3-D structure can be obtained by modeling</td>
<td>Roder et al., 2006</td>
</tr>
<tr>
<td>Raman spectroscopy</td>
<td>≥ µs</td>
<td>aromatic residue conformation and solvent accessibility</td>
<td>Balakrishnan et al., 2008</td>
</tr>
</tbody>
</table>

*a Bartlett and Radford, 2009*
1.5 Thesis Outline

This dissertation contains an experimental study on the effects of trehalose concentration, protein concentration, and thermal treatment on the kinetic, structural and thermodynamic stability of lysozyme; the development of a new experimental technique for detecting protein secondary structural intermediates upon folding; and experimental studies in which the technique is applied to both lysozyme and ribonuclease A. The development and initial results from the new protein folding experimental technique are found in the following publication:


The remainder of this section outlines the experimental methods employed for the studies conducted in this dissertation.

In Chapter 2, we used differential scanning calorimetry (DSC) and infrared spectroscopy (IR) to study the thermal and structural behavior of lysozyme, respectively, when subject to thermal treatment in various trehalose solution matrices ranging from 0 wt% trehalose to 80 wt% trehalose in water. Thermal treatments of lysozyme with different initial protein concentrations ranging from 0.1 mg/ml to a considerably higher protein concentration of 200 mg/ml were conducted as well. Ultraviolet spectroscopy (UV) was utilized to determine the initial and final concentration of native protein as well as the activity of the protein before and after thermal treatment. These UV experiments also provided a kinetic outlook on the effect of trehalose on lysozyme aggregation. The
focus of this study is to determine the effect of trehalose and protein concentration on lysozyme aggregation when it is thermally treated above its denaturation temperature.

In Chapter 3, Fourier transform infrared spectroscopy (IR) with a novel quench-and-refold approach was developed to experimentally detect the continuous secondary structural changes of lysozyme upon folding from its fully thermally denatured state. This technique is able to capture such structural changes due to the arrested time scales available from quenching, by liquid nitrogen in these studies, the protein in a solution matrix to its glassy state. Development of this new technique required careful selection of a solution matrix that would allow for full protein reversibility after heating above its thermal denaturation temperature and quenching to below the glass transition temperature of the solution matrix. This was conducted by first studying the aggregation properties of lysozyme in different formulations that included ethylammonium nitrate (EAN), sucrose / trehalose and water using DSC and IR. Enzymatic activity assays were also conducted to determine if lysozyme is still active in the solution matrix. Isothermal titration calorimetry was used to elucidate any enthalpy changes of interaction between lysozyme, EAN and water that could provide information on the ability of EAN to prevent or inhibit lysozyme aggregation. Once a suitable solution matrix was found, quench-and-refold experiments were conducted using both DSC and IR. Upon comparison of DSC and IR results, an interesting difference provided insight into energetics and structure formation relations and prompted further DSC quench-and-refold studies on kinetics and with other solution matrices.

In Chapter 4, we extend the same DSC and IR quench-and-refold experimental techniques and approaches applied to lysozyme to another protein, bovine pancreatic
ribonuclease A (RNase A). The main idea of this study was to determine if the energy-structure relation and observations found in Chapter 3 is applicable to another protein or if it is only a special property of lysozyme. An additional study on the effect of EAN concentration on the folding of RNase A using DSC and IR and the quench-and-refold approach was also explored, providing further insight into the stability of RNase A.

Conclusions and suggestions for future work are presented in Chapter 5.
1.6 References


Roder H, Maki K, Cheng H. 2006. Early events in protein folding explored by rapid


Chapter 2

Experimental Investigation of Lysozyme Aggregation in Trehalose

During unfolding and refolding processes *in vitro* and *in vivo*, proteins may form undesirable aggregates. Protein aggregation is a major problem in biotechnology and is also considered a cause of many fatal diseases such as Alzheimer’s or Huntington’s disease [Dobson, 2003]. Protein aggregation is often irreversible, and aggregates often contain nonnative protein structures [Dong et al., 1995]. It is generally known that the rate of reaction for protein aggregates to form is controlled by both thermodynamics and kinetics. Irreversible protein aggregation generally proceeds first from a denaturation step and then through an association of two or more protein chains. The aggregation rate and morphology of a protein aggregate have also been found to depend strongly on the protein’s solution environment, such as temperature, cosolutes, surfactants, pH, salt type and concentration [Carpenter et al., 1999; Randolph and Jones, 2002; Wang et al., 1995], the relative intrinsic thermodynamic stability of the native state [Kim et al., 2000], as well as the initial protein concentration [Georgalis et al., 1997]. Under conditions where
the protein’s native state is thermodynamically favored, denaturation is rate limiting. In solution conditions where the colloid or associated protein stability is high and the folded state is unstable, association is rate limiting [Chi et al., 2003].

One major approach to the prevention of protein aggregation is the addition of small molecules. Researchers have found that the addition of certain small molecules to the solution can act as a folding enhancer and stabilize the native state of the protein or suppress aggregation by reducing protein-protein interactions [Taravati and Ebadi, 2007]. Examples of such additives researched and used include carbohydrates, sorbitol, arginine, glycerol and spermine, among other protein stabilizing excipients [Hirano et al., 2007; Lee et al., 2006; O’Connor et al., 2004; Petersen et al., 2004; Arakawa and Tsumoto, 2003; Ueda et al., 2001; Valente et al., 2005]. In pharmaceutical research, it is highly desirable to find small molecules that are safe for use in therapeutics. One highly convenient additive is trehalose (α-D-glucopyranosyl-α-D-glucopyranoside). Trehalose, a glass-forming disaccharide molecule composed of two glucose units, is a well-known bioprotectant system, capable of protecting many organisms by preserving life under extremes of cold and dehydration [Lee et al., 1992; Crowe and Crowe, 2000]. Trehalose is naturally produced in the body and found in common foods, and is therefore “generally regarded as safe” by the U.S. Food and Drug Administration [Arakawa et al., 2001]. Trehalose prevents or reduces aggregation by protecting the native state against thermal denaturation. Denaturation is unfavorable in the presence of trehalose due to the increase in free energy difference between the native and denatured state, in essence increasing the stability of the protein’s native state [Kaushik and Bhat, 2003]. The change in free energy is a result of preferential exclusion of carbohydrate from the vicinity of the
protein [Lee and Timasheff, 1981; Arakawa et al., 1990]. This effect reduces the concentration of partially denatured states, which are prone to aggregation [Kendrick et al., 1998; Kerwin et al., 1998; Chi et al., 2003]. The addition of trehalose also increases a protein’s thermal denaturation temperature [Santoro et al., 1992]. The increase in denaturation temperature allows pharmaceutical drugs to be stable and active at higher temperatures during transport or storage.

In pharmaceutical drug production, protein aggregation occurs when active proteins are subject to heat and other stress conditions. Only very small protein concentrations are used to avoid or reduce aggregation and the consequent loss of activity. Several experiments have shown that increasing protein concentration increased aggregation [Georgalis et al., 1997; Ueda et al., 2001]. However, higher protein concentration in production is desirable for economic reasons. Recent studies have shown that at higher excipient concentrations, the rate of aggregation decreased due to higher solution viscosity [Munishkina et al., 2004; Tavarti et al., 2007]. With an increase in solution viscosity, unfolded proteins are inhibited in finding other molecules to interact with. Most research with high concentrations of protein study the effect of denaturants such as guanidine and urea on aggregation, but comparatively less work has been done on high protein concentrations in the presence of protein stabilizers [Rudolph and Lilie, 1996].

A particular interest of our research goal in developing a new experimental technique for observing protein folding mechanisms is to determine if protein aggregation is prevented upon heating the protein above its denaturation temperature in a solution with high protein and trehalose concentration. The solvent to be used for the
quench-and-refold experiments, which will be explained in detail in Chapter 3, should not promote aggregation and should allow protein reversibility when subject to thermal treatment. Here, in Chapter 2, we present infrared spectroscopic, ultraviolet spectroscopic, calorimetric and enzymatic assay results obtained in coming to the conclusion that another additive is needed in the solvent to prevent protein aggregation during thermal treatment. This additive is ethylammonium nitrate and will be described later in Chapter 3.

Hen egg-white lysozyme, a common model protein, was used to study aggregation at high temperatures [Ahern and Klibinov, 1985; Tomizawa et al., 1995]. Lysozyme is a 129 amino acid residue enzyme hydrolase and consists of 2 domains, one mostly \(\alpha\)-helical and the other predominantly \(\beta\)-sheet. Lysozyme also has 4 disulfide bonds and a \(pI\) of 11. Lysozyme catalyzes 1,4- \(\beta\)-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and between N-acetyl-D-glucosamine residues in chitodextrins [Vocadlo et al., 2001]. It has a molecular weight of 14.7 kDa. Lysozyme is abundant in a number of secretions, such as tears, saliva, human milk and mucus [Jolles, 1996]. It is also present in cytoplasmic granules of the polymorphonuclear leukocytes and in large amounts in egg white [Hindenburg et al., 1974]. A visual representation of lysozyme is shown in Figure 2.1. Lysozyme is also the first enzyme whose X-ray structure was determined [Johnson et al., 1965; Phillips, 1967]. Table 2.1 summarizes secondary structure information on lysozyme. Included in the table are the \(\alpha\)-helix, \(\beta\)-sheet, turn and random quantitative content percentages for lysozyme that were determined by different techniques: infrared spectroscopy (IR), X-ray diffraction and circular dichroism (CD).
Figure 2.1: Ribbon (left) visual representation of lysozyme secondary structure [http://lysozyme.co.uk]. Bead (right) visual representation of lysozyme amino acids and disulfide bond connections [Kimball, 1994].

Table 2.1: Quantitative Percentage Estimates of Secondary Structure Content for Lysozyme Using Different Experimental Techniques.

<table>
<thead>
<tr>
<th>Method</th>
<th>(\alpha)-helix</th>
<th>(\beta)-sheet</th>
<th>Turn</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR(^1)</td>
<td>40</td>
<td>19</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>X-ray(^2)</td>
<td>45</td>
<td>19</td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td>CD(^3)</td>
<td>38</td>
<td>30</td>
<td>22</td>
<td>10</td>
</tr>
</tbody>
</table>

\(^1\) Dong et al., 1990  
\(^2\) Levitt and Greer, 1977  
\(^3\) Manavalan and Johnson, 1987
Upon thermal denaturation, lysozyme exhibits two-state behavior, which makes this particular protein attractive for differential scanning calorimetric studies [Privalov and Khechinashvili, 1974].

The rest of the chapter is structured as follows. Section 2.1 describes the experimental techniques used to determine the extent of aggregation. Section 2.2 presents the results and discussion of the systematic study. Finally, Section 2.3 provides the main observations concluded from this study.

2.1 Experimental Methods

2.1.1 Materials

Trehalose dihydrate (90210) and hen egg white lysozyme (L-6876) were purchased from Sigma Chemical Co. (St. Louis, USA) and were used without further purification. Labels in parentheses correspond to the catalog numbers. The buffers, salts and all materials required for activity assays for lysozyme were also obtained from Sigma Chemical Co. In preparation of the samples used in these experiments, solvent solutions were prepared by weighed amounts of trehalose (Tre) and water (H\textsubscript{2}O or D\textsubscript{2}O). Lysozyme samples were prepared by dissolving protein in its lyophilized form mass appropriately into 1 ml of solution. Solution densities at 25 °C were used to determine the correct weight for 1 ml of the Tre-water solution and listed in Table 2.2 [Elias and Elias, 1999]. If 200 mg of lysozyme was added to 1 ml of solution, the protein concentration was labeled as “200 mg/ml”. Samples in this study ranged from 0.01 mg/ml to 200 mg/ml.
Table 2.2: Density of Various Trehalose and Water Solutions\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Trehalose wt %</th>
<th>Density of Trehalose –H\textsubscript{2}O g/ml</th>
<th>Density of Trehalose –D\textsubscript{2}O g/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.998</td>
<td>1.105</td>
</tr>
<tr>
<td>10</td>
<td>1.034</td>
<td>1.134</td>
</tr>
<tr>
<td>20</td>
<td>1.075</td>
<td>1.168</td>
</tr>
<tr>
<td>30</td>
<td>1.119</td>
<td>1.204</td>
</tr>
<tr>
<td>40</td>
<td>1.166</td>
<td>1.242</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Elias and Elias, 1999

### 2.1.2 Differential Scanning Calorimetry (DSC)

DSC experiments were conducted at lysozyme solution concentrations of 30 mg/ml and 200 mg/ml in various weight percent trehalose in H\textsubscript{2}O solutions using a Perkin-Elmer Pyris 1 differential scanning calorimeter equipped with an Intracooler 2P cooling accessory. This particular DSC is very useful for high protein concentrations. Indium was used to calibrate the enthalpy of the DSC instrument, while both indium and cyclohexane were used to calibrate for the temperature. An empty sealed stainless sealed pan was used as a reference. Approximately 40 µl of a sample solution was placed in a stainless steel pan. To prevent any water evaporation during the scans, the pans were hermetically sealed with an O-ring. All DSC thermal scans started at 25 °C and samples were heated to temperatures above the protein’s denaturation temperature, T\textsubscript{d}. The denaturation temperature was determined as the temperature at the maximum heat flow of the protein denaturation peak. This was considered as “scan 1”. Depending on the particular study, the samples were held for varying time lengths at 15 °C above the sample’s T\textsubscript{d}. Holding a sample at a particular temperature is referred to as annealing.
The samples were then cooled back to room temperature, and reheated. The reheated scan is considered as “scan 2”. Both heating and cooling scans were performed at scan rates of 20 °C/min. DSC scan results were normalized by the protein mass. Change in protein denaturation enthalpy between scan 1 and scan 2 provided a quantitative estimate of protein aggregation. Data for each sample was collected 3 times. Data were analyzed using Micocal™ Origin™, version 2.1.

2.1.3 Infrared Spectroscopy (IR)

IR spectroscopy is an excellent tool to observe and probe protein secondary structural changes [Surewicz et al., 1993]. Of particular interest in this study were the changes that occurred as lysozyme solutions of 30 mg/ml in 0 wt% trehalose, 20 wt% trehalose and 40 wt% trehalose in D2O were heated above their respective protein denaturation temperatures, annealed, and then cooled back to 25°C. IR spectra were recorded at various temperatures using a Nicolet 6700 FTIR spectrometer equipped with a broad-band liquid nitrogen-cooled mercury cadmium telluride (MCT) detector and a temperature-controlled demountable cell accessory from Harrick Scientific Products, Inc. Each spectrum was generated by co-addition of 64 interferograms collected at a 4 cm⁻¹ resolution. The spectrometer was also under a continuous dry air purge. For the liquid samples, a lysozyme sample of approximately 20 µl was sandwiched between two CaF2 windows using a 6 µm Teflon spacer. The sandwiched windows were then placed in the temperature-controlled cell. Reference spectra under identical scan conditions with only the solvent present were also recorded in the same cell. All protein IR spectra of the liquid samples were corrected by subtracting out the contribution of the solvent [Dong et
The IR spectrum of water vapor was subtracted from the sample spectrum. IR data was collected with 3 different samples.

To analyze and gain quantitative data from IR spectra, Fourier self deconvolution (FSD) [Byler and Susi, 1986; Dong et al, 1990; Arraondo et al., 1993; Pevsner and Diem, 2001; Kong and Yu 2007], secondary derivation and Gaussian curve fitting of the amide I’ region (1600-1700 cm$^{-1}$) were conducted. The amide I band is notated with a prime, such as amide I’ band, for samples in D$_2$O instead of H$_2$O. The amide I’ region arises primarily from the peptide carbonyl stretch vibration and has a high sensitivity to small variations in molecular geometry and hydrogen bonding patterns. This region is extensively characterized in protein studies [Carpenter et al., 1998, Kong and Yu, 2007], and is the optimal region to determine protein secondary structure. FSD was first used to enhance the resolution of all IR spectra. The full width at half-maximum (fwhm) was set at 15 and the enhancement factor (k) was set at 2.5; both values are similar to literature values [Byler et al., 1986; Gornetschelnokow et al., 1993; Carrasquillo et al., 2000]. The IR spectra was then analyzed by second derivative calculation and smoothed with an eleven-point Savitsky-Golay function. Second derivative spectra provided information on the number of peaks and peak positions for each sample [Susi and Byler, 1983; Susi and Byler, 1986], which were then used as starting parameters for the Gaussian curve fitting of the amide I’ region for quantification of lysozyme secondary structures. All IR spectra were baseline-corrected and normalized by area in the amide I’ region.
2.1.4 Ultraviolet Spectroscopy (UV)

Sample solutions prepared for studies using a UV spectroscopy consisted of lysozyme concentrations ranging from 0.01 mg/ml to 30 mg/ml. Lysozyme was placed in solvents containing 0 wt% trehalose, 20 wt% trehalose and 40 wt% trehalose in H2O. Sample solutions of 1 ml were placed in 1.5 ml disposable micro centrifuge tubes and heated to temperatures 15 °C above Td and annealed for various time periods in a VWR Scientific 1160 A water bath. After annealing, samples were then cooled back to room temperature. Once cooled, the samples were centrifuged using an Eppendorf 5424 microcentrifuge, courtesy of Professor David Wood’s lab, at 14,500 rpm for 20 minutes at 25°C (room temperature). The supernatant from the centrifuge samples was collected and then used to determine the concentration of soluble protein after thermal treatment. High protein concentration samples required dilution to ensure that the UV absorbance peaks of soluble protein were not saturated and were below the instrument maximum absorbance of 3. The protein concentration was monitored with a Spectronic Genesys2 UV Spectroscopy located at and used courtesy of Professor Benziger’s lab. The blank-subtracted absorbance at 280 nm (I280) and an extinction coefficient of 2.63 cm/mg ml was used to calculate soluble protein concentration [Sophianopoulus et al., 1962]. Samples were repeated 3 times.

2.1.5 Enzymatic Activity Assay

Thermally treated samples of lysozyme in varying protein concentrations ranging from 0.1 mg/ml to 30 mg/ml and varying trehalose concentrations of 0 wt%, 20wt% and 40 wt%, were assayed at 25 °C using procedures provided by Sigma [Shugar, 1952].
Assays were repeated 3 times. The activity assay for lysozyme is based on the ability of the enzyme to lyse or cut \textit{Micrococcus lysodeikticus} cells. Thermally treated solutions of varying initial protein concentrations were diluted to final protein concentrations of 6 to 20 µg/ml, depending on the thermal treatment the sample was exposed to and the solvent composition. Then 0.1 ml of the protein solution was subsequently diluted with 2.5 ml of 0.015% (w/v) \textit{Micrococcus lysodeikticus} cells suspended in 66 mM Potassium Phosphate Buffer, pH 6.24, and placed in a 1 cm path quartz cuvette. The decrease in absorbance at 450 nm was measured in an ultraviolet spectrometer (Spectronic Genesys2 in Professor Benziger’s lab) for approximately 5 minutes. The initial linear slope of the absorbance vs. time curve was multiplied by a dilution factor (from initial protein concentration to 20 µg/ml) to determine the activity of lysozyme.

\section*{2.2 Results and Discussion}

\subsection*{2.2.1 Lysozyme in 80 wt\% Trehalose}

Infrared spectroscopy was used to determine the reversibility of lysozyme unfolding in highly concentrated trehalose solution subjected to thermal denaturation. A sample of 30 mg/ml lysozyme in 80 wt\% trehalose was prepared. Due to the high trehalose concentration, the sample was pasty, highly viscous, and not a homogenous solution until it was heated above 80 °C. To achieve IR spectra at room temperature (25 °C), the paste was smeared onto the IR windows. The sample, sandwiched between two IR windows, was then heated at 20 °C/min to 98 °C. In Figure 2.2, the infrared spectra of 30 mg/ml lysozyme in 80 wt\% trehalose at 25 °C before heat treatment and at different
temperatures during heating to 98 °C are shown. Spectra were background subtracted and normalized by the highest peak between 1600-1700 cm\(^{-1}\) at each temperature. Particular wavenumbers can be associated with protein secondary structures, such as α-helix at approximately 1650 cm\(^{-1}\) [Luo et al., 1994, Kong and Yu, 2007]. Intermolecular antiparallel β-sheet aggregation is indicated by a strong absorbance at approximately 1618 cm\(^{-1}\) and a weak absorbance at approximately 1683 cm\(^{-1}\) [Ismail et al., 1992; Dong et al., 2000]. As indicated by the arrows, with increasing temperature there is a decrease in α-helix, and an increase of intermolecular β-sheets. The decrease in α-helix is expected, as the protein unfolds with increasing temperature. However the increase in intermolecular β-sheets upon heating, and the fact that these peaks did not disappear upon cooling, is a signature for irreversibly aggregated structures.

![Figure 2.2: IR spectra of 30 mg/ml lysozyme in 80 wt% trehalose-D\(_2\)O taken at 25 °C before heating (initial state – red curve) and at several temperatures during heating to 98 °C (black curve).](image-url)
This occurrence of lysozyme aggregation in 80 wt % trehalose solution may be attributed to excluded volume interactions, which are predicted by basic macromolecular crowding theory to accelerate protein aggregation [Munishkina et al., 2004]. Another possible explanation can be found in previous research showing that crowding cosolutes stabilize the protein state with the smallest accessible surface area (SASA). This typically forces the protein unfolding equilibrium towards the native state, however another pathway for a protein to reduce its SASA is to form aggregates [Berg, 1990; O’Connor 2007]. Therefore, depending on the experimental conditions, environment and treatment, the addition of crowding cosolutes to a solution can promote irreversible protein aggregation. Several experimental, theoretical and simulation studies have shown that crowding interactions can play an important role in enhancing protein association. A few examples of such research studies are listed and summarized as follows: Polyethylene glycol (PEG) has been reported to stimulate the association of pyruvate dehydrogenase complex [Nichol et al., 1981]. The presence of dextran has been shown to increase the self-association of Ca$^{2+}$-ATPase [Kosk-Kosicka et. al., 1995]. Sucrose has been shown to increase the final gel strength of bovine serum albumin (BSA) solutions due to increasing attraction between BSA molecules [Baier and McClements, 2006]. Trimethylamine N-oxide (TMAO), a naturally occurring osmolyte, and glycerol accelerated the formation of Alzheimer's beta-amyloid peptide (1-40) fibrils, which is an irreversible reaction, even though TMAO and glycerol can act as stabilizers to increase the thermal stability of proteins [Gekko and Timasheff, 1981; Baskakov and Bolen, 1998].
2.2.2 Effect of Trehalose Concentration

As determined earlier, lysozyme at a high protein concentration of 30 mg/ml and high trehalose solvent concentration of 80 wt% aggregated upon heating. A systematic set of experiments using DSC for thermal studies and IR for structural information were conducted to determine if the same behavior was seen at lower trehalose concentrations.

DSC scans of 30 mg/ml in 0 wt%, 20 wt% and 40 wt% trehalose in H$_2$O are presented in Figure 2.3. DSC scans above 40 wt% trehalose were not collected due to the formation of inhomogeneous solid-liquid mixture at room temperature, resulting in a trehalose melting endotherm at the approximate temperature of protein denaturation, which made analysis difficult.

Table 2.3 lists the enthalpy change upon denaturation, $\Delta H_d$, for 30 mg/ml lysozyme in 0, 10, 20, 30 and 40 wt% trehalose between 2 subsequent heat scans, as shown in Figure 2.3. The samples were heated to 15 °C above $T_d$ (scan 1), then held for 5 minutes at 15 °C above $T_d$, cooled to 25°C and then reheated back to 15 °C above $T_d$ (scan 2).

Analysis of DSC scans shows that protein thermal denaturation temperature, $T_d$, and enthalpy, $\Delta H_d$, increased almost linearly with increasing amounts of trehalose in solution, as shown in Figure 2.4. Values are also listed in Table 2.3, where the enthalpy of denaturation, 32.6 J/g for lysozyme in water, is in very good agreement with values found in literature [Burova et al., 2000]. The increases in $T_d$ and $\Delta H_d$ with increasing trehalose content have been related to increased thermodynamic stabilization of the native state by other researchers [Kaushik and Bhat, 2003].
Figure 2.3: DSC heating scans for 30 mg/ml lysozyme in various concentrations of trehalose. Samples were heated at 20 °C/min (scan 1 - solid line), annealed at 15 °C above respective T_d for 5 minutes, cooled and reheated (scan 2 – dotted line).

Table 2.3: Aggregation Percentage for 30 mg/ml Lysozyme in Varying Trehalose Concentrations Determined by Enthalpy Changes From DSC scans in Figure 2.3.

<table>
<thead>
<tr>
<th>Trehalose weight %</th>
<th>T_d °C</th>
<th>ΔH_d Scan 1 J/g lys</th>
<th>ΔH_d Scan 2 J/g lys</th>
<th>Aggregation %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>76</td>
<td>32.6</td>
<td>10.5</td>
<td>67.8</td>
</tr>
<tr>
<td>10</td>
<td>77</td>
<td>34.1</td>
<td>11.6</td>
<td>66.0</td>
</tr>
<tr>
<td>20</td>
<td>79</td>
<td>35.0</td>
<td>13.2</td>
<td>62.3</td>
</tr>
<tr>
<td>30</td>
<td>82</td>
<td>36.4</td>
<td>15.7</td>
<td>56.8</td>
</tr>
<tr>
<td>40</td>
<td>84</td>
<td>37.5</td>
<td>18.6</td>
<td>50.4</td>
</tr>
</tbody>
</table>
Figure 2.4: Dependence of denaturation temperature, $T_d$ (filled circles), and enthalpy from the first scan, $\Delta H_d$ (open circles), on trehalose concentration for 30 mg/ml lysozyme.

The difference in $\Delta H_d$ between the 2 scans is a measure of lysozyme aggregation. For a lysozyme concentration of 30 mg/ml, increasing trehalose concentration decreases the protein aggregation percentage from 67.8 % in a no trehalose solution to 50.4% in a 40 wt% trehalose solution. This observation reveals that increasing the amount of trehalose concentration in solution increases the protection of lysozyme against aggregation and that theories of molecular crowding described in Section 2.2.1 do not apply. On the other hand, a possible explanation for the prevention of lysozyme aggregation is that the addition of trehalose increases the free energy level difference between the native and denatured state, in essence increasing the stability of the protein’s native state [Kaushik and Bhat, 2003]. The change in free energy is a result of
preferential exclusion of carbohydrate from the vicinity of the protein [Lee and Timasheff, 1981; Arakawa et al., 1990], which effectively reduces the concentration of partially denatured state that are prone to aggregation [Kendrick et al., 1998; Kerwin et al., 1998; Chi et al., 2003].

Quantitative analysis of IR spectra for 30 mg/ml lysozyme in 0, 20 and 40 wt% trehalose-D\textsubscript{2}O at 25 °C before any heat treatment was conducted and summarized in Table 2.4.

Table 2.4: Secondary Structure Content for Lysozyme in Various Trehalose-Water Concentrations at 25 °C.

<table>
<thead>
<tr>
<th></th>
<th>Literature Values</th>
<th>0 wt% Tre</th>
<th>20 wt% Tre</th>
<th>40 wt% Tre</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Helix</td>
<td>40</td>
<td>40</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>β-Sheet</td>
<td>19</td>
<td>18</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>turn</td>
<td>27</td>
<td>28</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>random</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

As indicated by the percentages of secondary structure content, the secondary structures of lysozyme are identical in the absence and presence of trehalose within experimental errors. This observation was also found in previous studies, where lysozyme native structure is not perturbed in the presence of polyols such as glycerol and sorbitol [Haque 2005]. The secondary structure values observed are also consistent with the literature values [Dong et al., 1990, Luo et al., 1994]
Fourier deconvoluted and second derivative IR spectra of 30 mg/ml lysozyme in 0 wt% trehalose and 40 wt% trehalose taken at 25 °C before heat treatment, 90°C or 100 °C after 5 minutes annealing for 0 and 40 wt % trehalose respectively, and at 25 °C after heat treatment are presented in Figure 2.5. Second derivative spectra provided the location and wavenumber of distinct peaks detected that are associated with particular secondary structures.

Comparison of spectral changes in the absence and presence of trehalose show similar trends. IR Fourier-deconvoluted spectra taken at 25 °C before heat treatment are centered and exhibit a large distinct second derivative peak at ~1650 cm⁻¹, which is associated with α-helical structures, the dominant structure for folded lysozyme. As the protein was heated above its unfolding temperature, the spectral peak shifted to a lower wavenumber and also flattened to a less distinct spectrum. These shifts and changes are also evident in the lack of strong peaks found in the respective secondary derivative spectra.

On the other hand, after cooling, two new peaks appeared in the spectra taken after heat treatment. These two new peaks are situated at ~1680 cm⁻¹ and ~ 1620 cm⁻¹. These peaks, as mentioned in Section 2.2.1, are typically associated with intermolecular β-sheets and irreversible protein aggregation. The only noticeable difference in Figure 2.5 between the spectra in the presence and absence of trehalose is that the peaks associated with the aggregated structures in the absence of trehalose are slightly more intense and could mark the increased aggregation percentage observed earlier in the DSC results.
Figure 2.5: Stacked plots of second derivative (A,B) and deconvoluted amide I’ spectra (C,D) of 30 mg/ml lysozyme in 0 wt% Tre-D$_2$O and 40 wt% Tre-D$_2$O. Red curves are spectra of lysozyme at 25 °C (Initial) prior to heating. Green curves represent thermally unfolded lysozyme (Heated) with spectra taken at 90°C and 100 °C for 0 wt% Tre and 40 wt% Tre, respectively. Blue curves are spectra of lysozyme taken at 25 °C (Final) after samples were heated at 20 °C/min, annealed at 90 °C or 100 °C for 5 minutes and then cooled back to 25°C.

Qualitative analysis of the IR spectra was conducted as proper quantitative analysis of band positions and band intensities is challenged by the chemical exchange between hydrogen initially situated in the buried hydrophobic amino acids and the D$_2$O from the solvent [Byler and Susi, 1986, Baezinger and Methot, 1995]. Since deuterium is heavier than hydrogen, characteristic frequencies are slightly shifted lower as bond
stretching and bending motions are slowed. This shift is approximately 3-5 cm\(^{-1}\) wavenumbers [Pevsner and Diem, 2001; Kong and Yu, 2007]. The amide I band is notated with a prime, such as amide I’ band, for samples in D\(_2\)O. The hydrogen exchange is indicated by the change in intensity of the amide II band, which is associated with the bending vibration of the amine bond on the peptide backbone. IR spectra of protein in H\(_2\)O have an amide II peak centered at 1550 cm\(^{-1}\). When a protein is in D\(_2\)O, and deuteration of the protein occurs, the amide II’ band has an intensity at a lower wavenumber of 1450 cm\(^{-1}\). H/D exchanges during protein unfolding can be simultaneously monitored by the decrease in intensity of the band at 1550 cm\(^{-1}\) and the increase in intensity at 1450 cm\(^{-1}\), as shown in Figure 2.6.

Figure 2.6: IR spectra of amide I’, amide II and amide II’ regions for 30 mg/ml lysozyme in D\(_2\)O. Lysozyme in its native state before heating at 25 \(^\circ\)C (solid line) and after heating to its denatured state (dotted line) at 90\(^\circ\)C.
Figure 2.7 presents the normalized IR spectral changes in intensity at a particular wavelength, chosen to be ~1650 cm\(^{-1}\), which is associated with \(\alpha\)-helix structure in D\(_2\)O solvent. Notice the difference in intensity changes for 30 mg/ml lysozyme in 0, 20 and 40 wt\% trehalose.

For IR spectra, \(T_d\) is generally defined as the temperature corresponding to 50\% change in structure or intensity. \(T_d\) for 0, 20 and 40 wt\% trehalose is determined to be 75.5 \(^\circ\)C, 78.1 \(^\circ\)C, and 83.2 \(^\circ\)C respectively, which is on average slightly lower than \(T_d\) found in the DSC. This may indicate that some structural changes occur before the energy changes determined by the DSC. Both DSC and IR data were collected by using the same heating and cooling rate of 20 \(^\circ\)C/min. Also seen in Figure 2.7 is the fact that lysozyme in a higher trehalose concentration refolds upon cooling at a higher temperature than lysozyme at lower trehalose concentrations. This shift in temperature is also seen in DSC cooling scans for lysozyme samples studied in Figure 2.3 as well as in DSC scans of RNase A, presented in Chapter 4, Figure 4.11. This may be attributed to the preferential exclusion of trehalose from lysozyme [Kaushik and Bhat, 2003]. With more trehalose in solution, the difference in free energy between the native and denatured state of the protein is increased, thus requiring higher temperatures to unfold. The stabilized native state could be attributed to the earlier refolding, equivalent to higher temperatures seen in the cooling curves, as lysozyme has a higher preference to be in the native state in increasing concentrations of trehalose solutions compared to lysozyme in only water.
Figure 2.7: Ratio of the difference between absorbance intensity at a temperature (Abs$_T$) and the absorbance intensity of the denatured state (Abs$_D$) to the difference in absorbance intensity of the native state of lysozyme (Abs$_N$) and Abs$_D$ for the 1652 cm$^{-1}$ band for 30 mg/ml lysozyme in various concentrations of trehalose. Samples were heated at 20 $^\circ$C/min (filled symbols), annealed at 15 $^\circ$C above their respective $T_d$ for 5 minutes, and then cooled at 20 $^\circ$C/min (open symbols).

2.2.3 Effect of Protein Concentration

Further experiments were conducted to determine the effect of protein concentration on protein aggregation. Figure 2.8 plots the percent of lysozyme aggregation for 30 mg/ml and 200 mg/ml in various amounts of trehalose as determined by DSC. Figure 2.8 shows that increasing the concentration of protein leads to more aggregation; however, increasing the trehalose concentration further prevents protein aggregation. In the absence of trehalose, 67.8% and 85.3% of the lysozyme in the 30 mg/ml sample and 200 mg/ml aggregated, respectively, while in the presence of 40 wt% trehalose, 50.4% and 59.1% of the lysozyme in the 30 mg/ml sample and 200 mg/ml
aggregated, respectively. The aggregation percentage of protein between varying low trehalose concentrations exhibits a non-linear trend. Increasing the protein concentration from 30 mg/ml to 200 mg/ml decreased $T_d$ by $\sim 1.5 \, ^\circ C$ as determined by DSC experiments with a scan rate of $20 \, ^\circ C/min$.

![Graph](image.png)

Figure 2.8: Aggregation percentages determined from DSC of 30 mg/ml lysozyme (filled triangles) and 200 mg/ml (open triangles) in various concentrations of trehalose. Samples were heated at $20 \, ^\circ C/min$ and annealed at $15 \, ^\circ C$ above respective $T_d$ for 5 minutes. Error bars correspond to variability over 3 repeated samples.

Ultraviolet spectroscopic and activity assays were conducted to determine if this phenomenon is also observed at lower protein concentrations. Figure 2.9 plots the aggregation percentages determined by UV spectroscopy for protein concentrations of 0.1, 0.5, 1 and 2 mg/ml in 0, 20 and 40 wt% trehalose-H$_2$O solutions. As observed previously, increasing protein concentration increases the aggregation percentage. At low protein concentrations, the addition of trehalose prevents protein aggregation and follows similar trends to those found in literature [Ueda et al., 2001]. An interesting note
is that even at different initial low protein concentrations, the aggregation percentage does not vary as much with trehalose concentration as in the absence of trehalose. It can be seen that there is an aggregation percentage difference of more than 15% for lysozyme in water (0.1 vs. 2 mg/ml) compared to less than 5% difference for lysozyme in 40 wt% trehalose.

Figure 2.9: Effect of initial protein concentration on aggregation percentage for lysozyme heated to 15 °C above its respective T_d for 10 minutes in the absence (circles) or presence of 20 wt% trehalose (triangles) or 40 wt% trehalose (squares). Percentages were determined by the change in protein concentration using UV. Error bars show variability over 3 repeated samples.

A proposed mechanism by other researchers regarding osmolytes that have been found to prevent aggregation is the idea that trehalose is preferentially excluded from the protein favoring the native state as the protein would prefer to decrease its hydrophobic surface area [Kendrick et al., 1998; Kerwin et al., 1998; Chi et al., 2003; Kaushik and Bhat, 2003]. Circular dichroism analysis performed by other researchers has shown that
trehalose induces α-helical conformations and some tertiary structures [Ueda et al., 2001], which could reduce the concentration of partially denatured states that are prone to aggregation [Kendrick et al., 1998; Kerwin et al., 1998; Chi et al., 2003]. With the low protein concentration, the probability of proteins to “find each other” and aggregate is decreased and inhibited by the trehalose.

Enzymatic activity assays were also conducted as another measure of protein aggregation at low protein concentrations in various trehalose amounts. Figure 2.10 shows the residual activity after heat treatment for all samples previously presented in Figure 2.9 for the UV experiments. As one can see, similar effects are observed as in the UV experiments. Increasing trehalose prevented loss of activity and increasing initial protein concentration decreased activity recovery.

![Graph Figure 2.10: Effect of initial protein concentration on irreversible enzymatic inactivation for lysozyme heated to 15 °C above its respective Td for 10 minutes in the absence (circles) or presence of 20 wt% trehalose (triangles) or 40 wt % trehalose (squares). Percentages were determined by change in activity of lysozyme before and after heat treatment. Error bars show variability over 3 samples.](image-url)
The results reveal that the addition of trehalose, up to 40 wt%, prevents protein aggregation and inactivation, where increasing the trehalose concentration decreases the aggregation and inactivation percentages. However, as reported earlier, the addition of 80 wt% trehalose in fact promotes protein aggregation. Therefore, the overall protective ability of trehalose against protein aggregation upon heating has a non-monotonic behavior. There is also a much greater sensitivity of aggregation at lower trehalose concentrations as the percent of residual activity drops faster at 0 wt% trehalose.

However, upon comparison of percentage of unaggregated protein to percent residual activity, the latter was on average lower, as shown in Table 2.5. This could indicate that there are soluble aggregates of lysozyme that seem to have folded but are still inactive in the supernatant taken after centrifugation. This has also been observed in previous studies [Shiraki et al., 2004].

Table 2.5: Comparison of Percent Unaggregated Protein to Percent of Recovered Activity

<table>
<thead>
<tr>
<th>Trehalose weight %</th>
<th>Lysozyme Concentration mg/ml</th>
<th>% Residual Activity</th>
<th>% Unaggregated Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>41.1 (1.48)</td>
<td>49.6 (1.33)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>48.2 (1.53)</td>
<td>54.7 (1.52)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>53.5 (1.82)</td>
<td>62.2 (1.70)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>61.2 (1.90)</td>
<td>65.9 (1.94)</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>61.3 (2.25)</td>
<td>65.0 (2.31)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>62.4 (2.32)</td>
<td>66.2 (2.41)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>64.2 (2.32)</td>
<td>67.7 (2.44)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>66.9 (2.43)</td>
<td>70.3 (2.58)</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
<td>70.8 (2.65)</td>
<td>73.2 (2.45)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>72.1 (2.67)</td>
<td>74.8 (2.42)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>73.4 (2.71)</td>
<td>75.8 (2.56)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>74.9 (2.74)</td>
<td>76.6 (2.60)</td>
</tr>
</tbody>
</table>

* Values in parentheses are standard deviation of three measurements for each sample.
2.2.4 Kinetics

Aggregation is a kinetic process and as shown previously, highly dependent on initial protein concentration and solvent matrix. Aggregated proteins are said to undergo both structural changes and assembly processes when transforming from natively folded monomers and such pathways have been analyzed using the Lumry-Eyring framework, shown in Scheme 1 [Lumry and Eyring, 1954; Kendrick et al., 1998; Chi et al., 2003]. In Scheme 1, N represents the native protein, I represents an intermediate state before the formation of an aggregated intermediate A$_i$, A$_m$ represents aggregated protein consisting of $m$ intermediate units (I) and A$_{m+1}$ represents one more intermediate added to the aggregated protein.

\[
N \leftrightarrow I \rightarrow A_i \tag{a}
\]

\[
A_i + A_m \rightarrow A_{m+1} \tag{b}
\]

Scheme 1: Lumry-Eyring framework of protein aggregation
[modified from Chi et al., 2003]

Scheme 1 involves a reversible reaction to form an intermediate or transition state (a) before irreversible reactions takes place leading to aggregation (b). The overall, effective order of an aggregation reaction is dependent on the rate-limiting step.

Kinetic studies were conducted to further elucidate the effects of trehalose on lysozyme aggregation. Figure 2.11 presents the aggregation percentages for 30 mg/ml and 0.1 mg/ml lysozyme in various trehalose-water concentrations in relation to anneal time at 15 °C above $T_d$. Aggregation was determined by centrifugation and UV for lysozyme concentrations of 30 mg/ml (Figure 2.11 A), and 0.1 mg/ml (Figure 2.11 B).
Figure 2.11: Lysozyme aggregation of 30 mg/ml [A] and 0.1 mg/ml [B] lysozyme concentration samples after different thermal anneal times at 15°C above $T_d$ and at trehalose solvent concentrations of 0 wt% (circles), 20 wt% (triangles) and 40 wt% (squares). Dotted lines are guides for the eye. Error bars show variability over 3 repeated samples.
Kinetic enzymatic assays were also conducted for both 30 mg/ml and 0.1 mg/ml and plotted in Figure 2.12. Similar trends are observed between the aggregation percentage studies and the enzyme inactivation studies. Lysozyme at 30 mg/ml is fully aggregated within 20 minutes of annealing at 15 °C above \( T_d \) in 40 wt% trehalose. At 0.1 mg/ml, full aggregation was not reached within 30 minutes of annealing at 15 °C above \( T_d \) both in the absence and presence of trehalose.

The rate of aggregation and inactivation for 0.1, 0.5, 1, 2, and 30 mg/ml lysozyme in 0, 20 and 40 wt% trehalose solutions was determined by fitting the percent of unaggregated protein and the percent of residual activity to heating time. In literature, a number of proteins have been found to follow first order aggregation kinetics [Lumry and Eyring, 1954; Kendrick et. al., 1998; Hirano et. al., 2007]. Sample fits for 0.1 and 30 mg/ml in varying amounts of trehalose shown in Figure 2.13 indicate that lysozyme aggregation indeed follows first order kinetics as a semilogarithmic plot of the percent of unaggregated soluble protein after heating and annealing for specified times (A) over the percent of unaggregated protein after heating but before any annealing at the high temperature (\( A_0 \)) versus time in minutes yields linear regression \( R^2 \) values close to 1 (> 0.97). Observation of first-order kinetics suggests that the rate-limiting step is due to a conformational change in the protein, and not on interactions with other units [Chi et. al., 2003]. This appears to be the case for our findings for lysozyme aggregation in trehalose. Table 2.6 lists the aggregation and inactivation rate constants for various lysozyme concentrations in varying trehalose solutions.
Figure 2.12: Lysozyme thermal enzymatic inactivation dependant on thermal anneal time and trehalose solvent concentrations of 0 wt% (circles), 20 wt % (triangles) and 40 wt% (squares) for 30 mg/ml [A] and 0.1 mg/ml [B] lysozyme concentration. Dotted lines are guides for the eye. Error bars show variability over 3 repeated samples.
Figure 2.13: Linear kinetics fits to lysozyme aggregation upon annealing above $T_d$ for 30 mg/ml (filled symbols) and 0.1 mg/ml (open symbols) lysozyme concentration samples at trehalose solvent concentrations of 0 wt% (circles), 20 wt % (triangles) and 40 wt% (squares). $A$ is the percent of unaggregated soluble protein after heating and annealing for the specified times. $A_o$ is the percent of unaggregated protein after heating but before annealing. Error bars are smaller than the symbols.

The rate of inactivation and aggregation is found to be dependent on initial protein concentration and on trehalose concentration. Lysozyme at 30 mg/ml has higher aggregation rate constants compared to 0.1 mg/ml at varying trehalose concentrations, confirming that increasing protein concentration increases the aggregation rate. The rate of protein aggregation and inactivation is depressed by trehalose. This could be the result of trehalose increasing the free energy between native and intermediate structure, therefore thermodynamically favoring the native state and leading to this rate-limiting step, which in turn would determine the overall aggregation process to be first order.
Table 2.6: Aggregation and Inactivation Rate Constants for Lysozyme in Various Trehalose Concentrations.

<table>
<thead>
<tr>
<th>Trehalose weight %</th>
<th>Lysozyme Concentration mg/ml</th>
<th>Inactivation Rate Constant min$^{-1}$</th>
<th>Aggregation Rate Constant min$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
<td>0.193 (R$^2 = 0.983$)</td>
<td>0.179 (R$^2 = 0.979$)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.069 (R$^2 = 0.978$)</td>
<td>0.059 (R$^2 = 0.982$)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.065 (R$^2 = 0.984$)</td>
<td>0.053 (R$^2 = 0.977$)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.063 (R$^2 = 0.983$)</td>
<td>0.050 (R$^2 = 0.984$)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.060 (R$^2 = 0.979$)</td>
<td>0.049 (R$^2 = 0.988$)</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>0.157 (R$^2 = 0.981$)</td>
<td>0.143 (R$^2 = 0.986$)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.045 (R$^2 = 0.983$)</td>
<td>0.037 (R$^2 = 0.984$)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.039 (R$^2 = 0.990$)</td>
<td>0.034 (R$^2 = 0.987$)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.036 (R$^2 = 0.986$)</td>
<td>0.031 (R$^2 = 0.989$)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.035 (R$^2 = 0.978$)</td>
<td>0.030 (R$^2 = 0.984$)</td>
</tr>
<tr>
<td>40</td>
<td>30</td>
<td>0.123 (R$^2 = 0.993$)</td>
<td>0.119 (R$^2 = 0.991$)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.030 (R$^2 = 0.987$)</td>
<td>0.025 (R$^2 = 0.982$)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.026 (R$^2 = 0.985$)</td>
<td>0.024 (R$^2 = 0.986$)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.024 (R$^2 = 0.987$)</td>
<td>0.023 (R$^2 = 0.987$)</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.024 (R$^2 = 0.980$)</td>
<td>0.021 (R$^2 = 0.989$)</td>
</tr>
</tbody>
</table>

Lysozyme inactivation is found to be up to 1.09 times faster than aggregation, or in other words, the inactivation rate constant is 9% higher than that of the aggregation rate constant. (i.e. 30 mg/ml lysozyme in 0 wt% trehalose has an inactivation that is approximately 1.08 times faster than aggregation: (0.193-0.179)/0.179 +1). Removal of all inactivated molecules may not have been completed by our centrifugation protocol, which could be the cause of this small, but consistent, difference between the inactivation and aggregation rate constants. This possibility was also noted in Section 2.2.3 where the percent of unaggregated protein is generally higher than the percent of active protein after the same thermal treatment.
Experiments with protein concentrations below 0.1 mg/ml were also conducted by UV and enzymatic assays to determine if there was a protein concentration low enough where no aggregation was observed after heating above $T_d$ and annealing at various times. Below 0.01 mg/ml of lysozyme concentration, there was no evidence of aggregation in the absence or presence of trehalose even after annealing above $T_d$ for 2 hours.

### 2.3 Conclusion

Systematic experiments using differential scanning calorimetry, ultraviolet spectroscopy, infrared spectroscopy and enzymatic assays were conducted to study the effect of trehalose on lysozyme aggregation in order to find a suitable solution for quench-and-refold experiments. A high concentration of lysozyme (200 mg/ml) in a high concentration of trehalose (80 wt%) promotes aggregation upon thermal denaturation. Molecular crowding is one possible reason for this finding. Adding lower amounts of trehalose to the solvent in fact prevents aggregation, which is more in line with experiments and observations seen in previous literature [Ueda et al., 2001]. Studies conducted with both high and low protein concentrations in varying trehalose solutions ranging from 0 to 40 wt% reveal that the more trehalose is added, the less aggregation and activity loss is observed. Aggregation in the absence and presence of trehalose is also found to be dependent on initial protein concentration. Increasing protein concentration increases the rate of aggregation and decreases residual enzyme activity recovery. Results also reveal that both rates of aggregation and inactivation of lysozyme in trehalose due to thermal treatment follow first-order kinetics. Possible explanations for
the ability of trehalose to prevent aggregation are (1) that the preferential exclusion of trehalose from the protein increases the free energy between the native and denatured states, thus favoring the native state (2) and/or that trehalose helps to maintain and induce secondary structure, therefore providing a lower number of partially folded states for aggregation.
2.4 References


Burova TV, Grindberg NV, Grindberg VY, Rariy RV, Klibanov AM. 2000. Calorimetric
evidence for a native-like conformation of hen egg-white lysozyme dissolved in

Byler DM, Susi H. 1986. Examination of the secondary structure of proteins by
deconvolved FTIR spectra. Biopolymers. 25:469-487.

Carrasquillo KG, Sanchez C, Griebenow K. 2000. Relationship between conformational
stability and lyophilization-induced structural changes in chymotrypsin.

Carpenter JF, Kendrick BS, Chang BS, Manning MC, Randolph TW. 1999. Inhibition of
255.

Chi EY, Krishnan S, Randolph TW, Carpenter JF. 2003. Physical stability of protein in
aqueous solution: mechanism and driving forces in nonnative protein aggregation.
Pharm. Res. 20:1325-1336.

6:3591-3595.


Dobson CM. 2003. Protein folding and disease: a view from the first Horizon

Dong A, Huang P, Caughey WS. 1990. Protein secondary structures in water from the
second derivative amide I infrared spectra. 29:3303-3308.


Shugar D. 1952. The measurement of lysozyme activity and the ultraviolet inactivation of lysozyme. Biochimica et Biophysica Acta. 8:302-309


coefficient studies of cosolvent-induced protein self-interaction. Biophys. J.
89:4211-4218.


Wang A, Robertson AD, Bolen DW. 1995. Effects of a naturally occurring compatible
osmolyte on the internal dynamics of ribonuclease-a. Biochemistry 34:15096–
15104.

Towards a molecular level understanding of protein stabilization: the interaction
Chapter 3

Structure-Energy Relations in Hen Egg White Lysozyme Observed During Refolding From a Quenched Unfolded State

The folding of proteins, the process in which their characteristic and functional three-dimensional structure forms from a random coil, is one of nature’s most ubiquitous phenomena. Understanding the mechanisms underlying protein folding is important for the rational modification and design of novel proteins, the understanding of human degenerative diseases that are tied to protein misfolding and/or aggregation, [Dobson, 2003; Chiti and Dobson, 2006] and the development of improved formulations to prevent protein degradation during storage [Cleland et al., 1993].

The fundamental objectives in protein folding research include the characterization of intermediate or partially structured states in terms of their structural, kinetic and thermodynamic properties; and the determination of their implications for
folding pathways. In this chapter, we focus on capturing the structures of protein intermediates using infrared spectroscopy.

One of the biggest challenges in studying protein folding mechanisms is the experimental difficulty to detect and capture these intermediates continuously during folding of a protein from its unfolded state. Current optical methods, such as infrared spectroscopy or circular dichroism, utilize laser temperature-jump (T-jump) or photophysical dissociation experiments with sub-µs resolution to study protein folding mechanisms. The T-jump technique uses a burst of energy to quickly heat a sample solution (e.g., a few nanoseconds) to a particular temperature. This sudden increase in temperature induces a population-redistribution among the conformational ensembles that are initially at equilibrium. The time course in which the nonequilibrium state evolves back to its equilibrium state contains information regarding protein folding and unfolding kinetics, and is dependent on the final temperature. In a simple two-state scenario, the relaxation response to the T-jump follows first-order kinetics, however if the kinetics are not-monoexponential, the folding mechanism is said to involve intermediates [Wang et al., 2005]. Photophysical dissociation experiments use photochemical triggers to change the equilibrium state and the experimental interpretation is analogous to T-jump experiments. However, these techniques do not necessarily probe the structural continuous changes starting from a fully denatured state. This is due to the limitations on the control of perturbation initiators. For example, a typical experimental setup using absorption of laser energy for heating can be successfully controlled to a maximum temperature jump of 10 °C. However, proteins thermally unfold at different temperatures depending on the solution matrix [Santoro et
al., 2002] and their denaturation temperature could be greater than the temperature jump of 10 °C above room temperature or the initial state temperature. Therefore, with a controlled heat input, relaxation response data may be collected from an intermediate structure, but not necessarily from the fully denatured state [Eaton et al., 2000; Dobson, 2004; Lapidus et al., 2007].

Standard stopped-flow techniques or rapid mixing methods, on the other hand, are able to prompt folding from fully denatured samples. In these experiments, protein folding is initiated by rapidly mixing a chemically denatured protein solution with a buffer to dilute the denaturant [Radford et al., 1992]. Common denaturants used are urea and guanidinium chloride. The kinetics of folding is then monitored with one of several possible experimental methods, such as small angle x-ray scattering (SAXS), circular dichroism, and high-resolution proton spin nuclear magnetic resonance (NMR). Mixing techniques to dilute the denaturant, however, require time resolutions on the order of 1 ms [Eaton et al., 2000; Lapidus et al., 2007]. Consequently, important events in the refolding from fully denatured states that occur on the µs time scale are not observable, which creates an important gap in fundamental understanding. Recent developments in microfluidics have allowed denaturant dilution in less than 20 µs, offering the prospect of vastly improved time resolution using stopped-flow approaches [Lapidus et al., 2007; Eaton et al., 2000]. Stopped-flow and rapid-mixing methods, however, require changes in solution conditions to follow protein folding, which could affect the protein folding pathway as previous studies have shown that refolding rate is dependent on denaturant concentrations [Kliger et al., 2004].
Here we present an experimental method that is capable of detecting the changes in intermediate structure during the folding process on arbitrarily slow (arrested) time scales starting from the fully thermally-denatured state, without the need to change solution conditions during folding. To this end, a recently introduced quench-and-refold approach [Angell and Wang, 2003] is adapted to spectroscopic scrutiny. The general approach consists of the following steps and is detailed further in the chapter:

1. Thermally unfold and denature the protein
2. Trap the unfolded protein by very rapid cooling (quench) into a low temperature glassy matrix
3. Refold the protein by slowly heating the sample across the matrix glass transition temperature.

At this stage, Fourier Transform infrared spectroscopy (IR) is used to collect spectra of the protein at various temperatures and times. Fourier transform IR spectroscopy is very useful for protein secondary structure detection and quantification [Byler and Susi, 1986; Dong et al., 1990; Arraondo et al., 1993; Pevsner and Diem, 2001; Kong and Yu 2007]. The best characterized and most-studied region is the amide I’ band [Carpenter et al., 1998] (1600–1700 cm\(^{-1}\) region), which has a high sensitivity to small variations in molecular geometry and hydrogen bonding patterns and arises primarily from the peptide carbonyl stretch vibration [Kong and Yu, 2007]. Since the carbonyl group forms hydrogen bonds to peptide NH groups in a variety of different secondary structural environments [Surewicz et al., 1993], the amide I’ band involves multiple single peaks having different frequencies, each peak being related to a specific secondary structural element (e.g. \(\alpha\)-helix, \(\beta\)-sheets). Qualitative monitoring of structural changes
can be done by following changes in absorption at a specified frequency. Quantitative analysis involves curve fitting with a superposition of Gaussian or Lorentzian peaks. Peak areas and positions are then used to provide a quantitative estimate of relative weights assigned to specific secondary structural elements [Jung, 2000]. Accordingly, protein folding intermediates can be observed and captured by taking advantage of the very slow kinetics in the viscous liquid state near the glass transition temperature (T_g) [Debenedetti and Stillinger, 2001].

Angell and Wang first presented the quench-and-refold approach and provided proof-of-concept using differential scanning calorimetry (DSC) [Angell and Wang, 2003]. In conventional DSC studies, a protein is considered fully denatured upon the completion of energy absorption, which is indicated by an endothermic peak. The temperature at the maximum of this endothermic peak is considered the denaturation temperature (T_d). The integrated area under the endothermic peak quantifies the enthalpy required to unfold the protein. Refolding of a protein is associated with an enthalpy release, calculated as the area of the exotherm found in subsequent cooling curves [Privalov and Khechinashvili, 1974; Privalov, 1979]. Figure 3.1, collected by Angell and Wang, compares DSC upscans using the quench-and-refold approach for a protein that was quenched from the folded state to that obtained following quenching from the unfolded state. DSC upscan of the quenched folded protein reveals an endothermic peak around 70 °C, whereas the DSC upscan of the quenched unfolded protein reveals an exothermic area starting from ~ 0 °C as well as the endothermic peak around 70 °C [Angell and Wang, 2003].
Angell and Wang introduced the idea that the refolding enthalpy is indicated by the observed enthalpy release starting from $\sim 0 \, ^\circ\text{C}$ [Angell and Wang, 2003]. The refolding enthalpy is then calculated as the integral of the exotherm starting at $\sim 0 \, ^\circ\text{C}$ until the folded protein baseline (extension shown with a dotted line), as shown in Figure 3.2 and represented by the dark grey area. The light grey area in Figure 3.2 is associated with the energy for protein unfolding. Figure 3.2 is a DSC quench-and-refold scan of 20 wt% lysozyme, 32 wt% EAN, 21 wt% H$_2$O and 27 wt% sucrose conducted on our own. From Figure 3.2, the integrated areas of the unfolding and refolding enthalpy are similar, providing DSC evidence of the ability of liquid nitrogen quenching to successfully trap unfolded lysozyme. Also notice that the exothermic deviation in the quenched unfolded DSC upscan is not a smooth peak as seen in the unfolding endothermic peak; this could
imply that intermediates exist upon refolding and that protein refolding is not a single-step process. Combining the quench-and-refold approach with DSC, detection of stages during refolding is possible [Angell and Wang, 2003]. Therefore, building on this proof-of-concept, we here take the idea one step further by applying the quench-and-refold approach with infrared spectroscopy to observe protein intermediate structures and the refolding process.

![DSC scan of 20 wt% lysozyme, 32 wt% EAN, 21 wt% H₂O and 27 wt% sucrose](image)

**Figure 3.2:** A DSC scan of 20 wt% lysozyme, 32 wt% EAN, 21 wt% H₂O and 27 wt% sucrose, illustrating integrated areas used to determine unfolding (light grey) and refolding (dark grey) energies of a quench-and-refold heat scan.

In this chapter, we describe the development of a new experimental technique consisting of applying the quench-and-refold approach together with IR in order to study protein folding mechanisms utilizing the glassy state. We also show that combining DSC and IR offers new insight into energetic and structural signatures of the refolding process,
and their non-trivial relationship to each other. This chapter is organized as follows: Section 3.1 lists the materials and experimental methods used in the development of this technique. In Section 3.2 the results and discussion are presented, including a detailed approach formulation and key development considerations. Then in Section 3.3 we summarize the main findings.

3.1 Experimental Methods

3.1.1 Materials

Hen egg white lysozyme (L-6876) in lyophilized form was purchased from Sigma Chemical Co. (St. Louis, USA) and was used without further purification. Sucrose (S-8501) was also purchased from Sigma Chemical Co. Ethylammonium nitrate (EAN) was synthesized by adding nitric acid (87920) to an approximately equimolar (slight excess of amine to ensure complete reaction) amount of ethylamine (A15022) in aqueous solution, both purchased from Alfa Aesar. Since the reaction was very exothermic, the drop wise addition of the acid to the amine was carried out while cooling the amine solution to -78 °C using an acetone/dry-ice bath. The mixture was stirred at room temperature for several hours. The slight excess of amine left over after the reaction was removed along with the water by heating at 80 °C in vacuum using a Büchi 461 rotary evaporator. The product was then dried at 80 °C for two days in a Fisher Isotemp Vacuum Oven Model 281 containing P₂O₅ to remove any excess water. The rotary evaporator and vacuum oven were used courtesy of Professor Prudhomme’s lab. The buffers, salts and all materials required for activity assays for lysozyme were also obtained from Sigma Chemical Co.
3.1.2 Differential Scanning Calorimetry

DSC experiments were conducted to determine the reversibility of lysozyme folding – unfolding in various solution concentrations and mixtures of EAN, sucrose and H$_2$O or D$_2$O as part of the development of the quench-and-refold approach. A Perkin-Elmer Pyris 1 differential scanning calorimeter equipped with an Intracooler 2P cooling accessory was employed. To calibrate the enthalpy of the DSC instrument, indium was used, while both indium and cyclohexane were used to calibrate the temperature. Solutions for these particular DSC experiments were prepared by adding weighed amounts of water and/or sugar to EAN. The protein in its lyophilized form was dissolved mass appropriately into 1 ml of the hydrated liquid solution. Solution densities at 25 °C were used to determine the correct weight for 1 ml of the solution. If 200 mg of lysozyme was added to 1 ml of solution, the protein concentration was labeled as “200 mg/ml”. EAN-water solutions labeled with “mol %” were measured gravimetrically using a molecular weight of 108.1 g/mol for EAN. Solution density values utilized in the following experiments are listed in Table 3.1 [Allen et al., 1985; Mathlouthi and Reiser, 1995].

An empty sealed stainless steel pan was used as a reference. Approximately 20 µl of a sample solution was placed in a stainless steel pan. DSC thermal scans started at 25 °C and involved heating to temperatures above the sample’s denaturation temperature, T$_{d}$, which was as high as 98 °C. The T$_{d}$ of the protein was found to be dependent on the solution matrix (details in Section 3.2.1.2). This was considered as “scan 1”. The samples were then cooled back to room temperature, and reheated (“scan 2”). Heating and cooling scans were performed at scan rates of 20 °C/min unless otherwise noted. The
enthalpy of unfolding, $\Delta H_u$, of lysozyme across $T_d$ was calculated from the integral of the denaturation peak and is listed in what follows with units of Joules per gram of lysozyme ($J/g$ lys). The change in enthalpy between scans 1 and scans 2 provided a quantitative measure of protein aggregation.

Table 3.1: Measured Density of Solutions Utilized in This Research$^a$.

<table>
<thead>
<tr>
<th>Solution Mixture (wt%)</th>
<th>Density (g/cm$^3$)</th>
<th>EAN mol %</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAN</td>
<td>H$_2$O</td>
<td>D$_2$O</td>
</tr>
<tr>
<td>60</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>40</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>22</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td>22</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Allen et al., 1985; Mathlouthi and Reiser, 1995

Quench-and-refold DSC experiments were executed in the lab with the availability of a new LN$_2$ cooling accessory for the Perkin-Elmer Pyris 1 differential scanning calorimeter. Aluminum DSC pans were used in these studies and a blank aluminum pan was used as a reference. Approximately 10 µl of a sample solution were placed in aluminum pans. Quench-and-refold experiments were conducted by first heating the sample in the aluminum pan to a temperature ~ 15 °C above its respective $T_d$ using the DSC. The heated pans were then immediately quenched in a cup of LN$_2$. The DSC was equilibrated to -135 °C. After equilibration, the sample was placed back into
the DSC. Heat scans at 20 °C/min were taken from -135 °C up to 100 °C to observe the
glass transition, the exothermic relaxation of the protein folding, and then the protein
unfolding energy peak. For solvent-only studies, comparison between LN2 quenched and
DSC cooled sample studies were conducted. The fastest possible cooling from the DSC
instrument was 500 °C/min, compared to a LN2 quench rate of 132 °C/sec. To isolate the
energy changes of the protein during folding and unfolding, DSC scans containing
protein were background-subtracted by DSC scans of the same mass of the solvent only
placed under the same conditions of heating, cooling, quenching and scanning. The
refolding enthalpy was calculated as the integral of the deviation in energy starting at 0
°C until the folded protein baseline, as shown in Figure 3.2. All DSC experiments were
repeated 4 times. DSC scan results of the heating rate were normalized by the protein
mass and are presented in what follows as W/g versus temperature. Data was analyzed

3.1.3 Enzymatic Activity Assay

Lysozyme activity was measured to determine whether the protein was still in
functional form in the presence of EAN and after thermal treatment to 15 °C above \( T_d \)
and annealing the protein for 10 minutes. Experiments were repeated 3 times. Lysozyme
was assayed at 25 °C using a procedure provided by Sigma [Shugar, 1952]. The activity
assay for lysozyme was based on the ability of the enzyme to lyse \( Micrococcus 
lysodeikticus \) cells. Protein solutions of 6 µg/ml were prepared. Then 0.1 ml of the
enzyme solution was diluted with 2.5 ml of 0.015% (w/v) \( Micrococcus 
lysodeikticus \) cells suspended in 66 mM potassium phosphate buffer (pH 6.24) and placed in a 1 cm
path quartz cuvette. The decrease in absorbance at 450 nm was measured in an ultraviolet spectrometer (Spectronic Genesys2 in Professor Benziger’s lab) for approximately 5 minutes. The initial linear slope of the absorbance vs. time curve was used to determine the activity of lysozyme.

### 3.1.4 Infrared Spectroscopy

IR spectra of lysozyme in EAN, sucrose and water were recorded at various temperatures using a Nicolet 6700 FTIR spectrometer equipped with a broad-band liquid nitrogen-cooled mercury cadmium telluride (MCT) detector and a Linkam FTIR600 thermal vertical cold stage with liquid nitrogen (LN$_2$) cooling and CI94 temperature control units. This Linkam temperature controlled cell has a temperature range from -196 °C to 600 °C with scan rate from 0.1 °C to 100 °C/min and a temperature stability of <0.1°C. FTIR instruments in Professor Aksays’s lab and at International Specialty Products, Inc in Wayne, NJ were used. The spectrometers were also under a continuous dry air purge. Each spectrum was generated by a co-addition of 64 interferograms collected at a 4 cm$^{-1}$ resolution. Samples for the IR studies were prepared by adding weighed amounts of protein, D$_2$O, EAN and sucrose. A 2 wt% protein, 26 wt% D$_2$O, 39 wt% EAN and 33 wt% sucrose composition was used for most experiments. Samples of approximately 6 µl were sandwiched between two AgCl$_2$ windows. For quench studies, the sandwiched windows were placed in a fitted screw and nut set-up (to prevent evaporation) and then placed in an oven (VWR Scientific Sheldon 1330G) at 90 °C for 10 minutes to unfold lysozyme. Then the sample was quickly quenched in a cup of LN$_2$. After quench, the sample was placed in the temperature-controlled cell at -135 °C and
subsequently heated at 2 °C/min with 2 minute annealing steps every 10 °C to collect spectra. To account for possible changes in spectra due to thermal effects on vibration frequencies in the solvent, reference spectra under identical scan conditions with only the solvent present were also recorded in the same cell. Protein IR spectra of the liquid samples were corrected by subtracting the contribution of the solvent at each temperature. IR spectra were collected for 4 separate samples of the same composition.

The subtracted spectra were then baseline-corrected and normalized by area in the amide I’ region [Dong et al., 1995; Carpenter et al., 1998]. Quantitative spectral analysis was performed in two steps. First, second derivative spectra were obtained using a 2nd degree, 11-point Savitsky-Golay derivative function in Grams software. This data provided the number of peaks and peak positions needed to fit the original spectra with Gaussian peaks [Susi and Byler, 1983; Susi and Byler, 1986]. In the second step, the peak number and positions were fixed, while peak intensities and widths within limits (specified later) were allowed to vary to obtain the best fit. Peak areas and positions were then used to provide a quantitative estimate of relative weights assigned to specific secondary structural elements. A more detailed description of the quantitative approach used for the quench-and-refold studies is discussed in Section 3.2.3.

3.1.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was used to elucidate the enthalpy changes of the interaction between the following components used in our samples: lysozyme, EAN and water. A MicroCal MCS-ITC was used at room temperature, 25 °C, and the cell solution consisted of either water or lysozyme and water. EAN was used as the
diluting and binding agent injected into the 2.4 mL holding cell. ITC runs consisted of 30 injections of 6 µl of a 22.3 mM EAN in water solution at 180 sec intervals (except for the first injection which was only 2 µl, and was not considered in the analysis). The lysozyme in water concentration used in the holding cell was 1.36 mM. Experiments were conducted 3 times. Analysis was performed using Microcal Origin v2.9 software.

3.2 Results and Discussion

3.2.1 Approach Formulation

In order to successfully capture the structure of the protein intermediates during the folding process, careful consideration and attention to the protein and solvent matrix utilized was needed in development of this new experimental technique. This section summarizes the protein selection and the experiments completed to establish the optimal solvent composition that was then used for quench-and refold studies with infrared spectroscopy. The following is a list of the steps utilized in the quench-and-refold approach, where respective development considerations are presented in further detail.

A. Thermally unfold and denature the protein.

To prepare the sample, protein was first dissolved in a suitable solvent matrix (which is discussed in further detail in Section 3.2.1.2). The sample was then heated in an oven at a temperature ~15 °C greater than the protein denaturation temperature, T_d, for approximately 10 minutes. T_d was determined by differential scanning calorimetry and was found to be dependent on the solvent matrix composition of the test system, which was taken into consideration for the oven heating temperature. The 10 minute anneal time
allowed the protein to thermally denature and also for hydrogen-deuterium exchange (the importance of this is explained in Section 3.2.1.2).

**B. Trap the unfolded protein into a glassy matrix by quenching.**

The sample was immediately immersed in LN$_2$ after being heated. This quench transformed the solvent into a glassy matrix, trapping the protein in its unfolded state and structure. To ensure successful trapping of the unfolded state of the protein using LN$_2$, the speed of the protein folding process was taken into consideration, which is discussed further in Section 3.2.1.1. Solvent matrix formulation, discussed in Section 3.2.1.2 dictated whether a glassy matrix was obtained at temperatures above that of LN$_2$: -196 °C. Care was also taken in not letting the sample cool before quenching, as this allowed the protein to refold before trapping of the unfolded protein.

**C Refold the protein by slowly heating the sample across the glass transition temperature.**

The quenched sample was then placed in a cold stage capable of controlled temperature scans in the Infrared spectrometer. The cold stage was already cooled to a temperature at least 20 °C below the solvent matrix’s glass transition temperature, $T_g$ before heating the protein sample in the oven. This ensured that the IR cold cell is ready after quenching of the heated sample. The sample was then slowly heated back to room temperature. At chosen temperatures and times, infrared spectroscopic scans of the structure were taken. Quantitative and qualitative data was then collected to determine the structural trend of protein folding.
3.2.1.1 Protein Selection

The model protein must have the ability to refold completely when subjected to thermal treatment, such that the full continuous range of structural changes from the denatured state to the native state is observable. Quenching and trapping of the unfolded protein into a glassy state was accomplished by dropping the sample into LN2. To ensure that the protein was indeed trapped in its unfolded state by this immersion method, the cooling rate was compared to the protein refolding rate.

Protein samples for IR experiments were smeared and sandwiched into a thin film (6 µm thickness) between 2 AgCl windows (0.5 mm thickness). AgCl windows were used, as they are the only windows that could withstand the thermal shock of quenching in LN2. The entire system (windows and sandwiched sample) was immersed into LN2 for quenching. Since AgCl windows are significantly thicker than the protein sample, cooling rates for the AgCl window (henceforth referred to as the sample) were calculated. If the cooling environment is rate limiting (as opposed to heat conduction within the solid or sample), there are no thermal gradients in the sample. Under these conditions the initial cooling rate for a wafer cooling from both sides in the quench bath was calculated using equation (1) [Bird et al., 2002].

\[
\frac{dT}{dt} = \frac{h_q(T_s - T_q)}{\rho_s C_p s L_s}
\]

where \(T_s\) is the temperature of the sample before quench (90 °C), \(T_q\) is the temperature of the quench bath (-196 °C), \(h_q\) is the heat transfer coefficient of the quench bath (460 W/m² K), \(\rho_s\) is the density of the sample (5590 kg/m³), \(C_p s\) is the specific heat of the sample (355 J/kg K), and \(L_s\) is half that sample thickness. Values in parenthesis are values used
for calculation with AgCl as the sample and LN2 as the quench bath and determined in previous studies [Birnie and Kingry, 1985; Reflex, 2006]. With this system, the cooling environment is indeed considered rate-limiting as the Biot Number is less than 10: Bi = $h_q L_s/k_s = 0.2$, where $k_s$ is the thermal conductivity (1.15 m K) [Reflex, 2006; Bird et al., 2002]. The cooling rate is determined to be 132 K/sec.

With a known cooling rate, a suitable protein for our study should not fold within the cooling time limit in order to successfully suppress folding during quenching and trap the unfolded protein into a glassy state. Therefore, only proteins considered as slow folders (> 100 ms) were considered. With this current set-up and procedure, proteins that have refolding times greater than ~2 seconds would work ($= 286 K/132 K \text{ sec}^{-1}; 286 = 90-(-196)$). Note that this limit imposed is very conservative, as decreasing temperature while quenching increases solution viscosity and therefore increases protein folding time [Jas et al., 2001; Longinotti and Corti 2008]. If faster-folding proteins are of interest, decreasing sample size increases the cooling rate. Hen egg white lysozyme (described in detail in Chapter 2) was selected as the model protein for it is considered a slow folder with a total folding time in the range of 20 seconds [Kato et al., 1982; Dobson et al., 1994]. Previous research has reported that lysozyme refolds in steps as fast as on the millisecond timescale under chemical denaturation; however specific thermal folding rates of lysozyme in compositions (EAN, sucrose and water) utilized in this work have not been studied previously. Considering the high viscosity of our sample, and low temperatures employed for our experimental conditions, the protein folding rate is sufficiently decreased, allowing for the ability to capture fully unfolded lysozyme by quenching in LN2. Experimental evidence is presented later in this chapter. Lysozyme
was also chosen for it is commonly used in experiments and known to exhibit folding intermediates [Smith et al., 1993; Kulkarni et al., 1999].

For fast-folding proteins, where refolding in the µs range is reported [Munoz, 2007], an electrospray method that produces fine particles and filaments is one suitable technique to quench and trap unfolded proteins for IR studies with cooling rates in the order of $10^5$ K/sec observed [Wang et al., 2007]. However, in the present studies we are focused on slow-folding proteins using LN$_2$ quench.

### 3.2.1.2 Solvent Formulation

The solvent (no protein added) must have the following properties to properly trap the unfolded protein and capture structural intermediates:

- Be highly resistant to crystallization
- Have low water activity so ice can not form during warm-up
- Must not denature proteins
- Be fluid at moderate temperatures

The solvent should not allow ice to form upon cooling or heating, because this would cause abrupt changes in solution composition and sample temperature. Prevention of ice crystallization also allows for a clear observation of the enthalpy released during protein refolding at low temperatures in DSC studies. If ice is formed upon quenching, a large endothermic peak is observed around and overlapping the start of the exothermic peak for protein refolding from the quenched unfolded state, thus making proper integration of the refolding energy difficult. The solvent itself should not denature the proteins at room temperature and should support the thermally denatured protein at higher temperatures.
without permitting aggregation to occur. A fluid solvent at moderate temperatures is desired, meaning that the solvent should not have too high a glass transition temperature, thus allowing protein to thermally unfold. Therefore, the solvent’s melting and glass transition temperatures ($T_m$, $T_g$) are important in experimental design. The sample solution should not exhibit any aggregates at the protein unfolding temperature, thus the solvent $T_m$ cannot be too high. On the contrary, $T_g$ should not be too low, as our LN$_2$ cooled cold-stage (FTIR600 stage with a minimum temperature of -196 °C) is required to be able to keep the quenched sample well below (~20 °C) its $T_g$.

Consider Figure 3.3, a state diagram of the trehalose-water system consisting of equilibrium data (solubility and freezing curves), the glass transition curve and the liquidus curve extrapolated beyond the eutectic point. Trehalose is purchased and utilized for many experiments and applications in its dihydrate form. Trehalose dihydrate has a $T_m$ of approximately 100 °C. Upon heating trehalose dihydrate above its $T_m$, water is released to form anhydrous trehalose crystals, which have a $T_m$ of approximately 215 °C [Chen et al., 2000]. The extension of the liquidus curve beyond the eutectic point is when the solute becomes supersaturated and is in a metastable state. Along the metastable extension, a critical solute dependent concentration is reached, whereupon mobility is hindered to such an extent that the solution vitrifies. Where the extended line intersects the glass transition curve, ice formation ceases within the time-scale of the measurement and is considered the maximally freeze-concentrated $T_g$ of the frozen system [Roberts and DeBenedetti, 2002]. From a literature review on trehalose-water studies, the $T_g$ of pure trehalose is still a matter of debate, as reports range between 73 °C and 115 °C [Chen et al., 2000].
The approach is demonstrated through positions marked on the diagram as A, B, C and D. If, for example, we take a solvent composition of 75 wt % trehalose and 25 wt% water, the solvent is at an extremely high viscosity and very hard to mix and make homogenous at room temperature (A). The sample consisting of the solvent with the protein has to be heated above the solubility temperature, approximately 80 °C, in order to start with a homogenous solution (B). The sample actually has to be above 90 °C for lysozyme to completely unfold. When the sample is quenched, it must be held at a temperature lower than the solution’s glass transition temperature (C), which is approximately -40 °C. Then to gather structural intermediates and changes, the sample is slowly heated across the glass transition temperature back to room temperature (D).

Figure 3.3: State diagram for trehalose-water binary mixture modified from literature review [Chen et al., 2000].
Proteins are often stabilized in the glassy state by use of carbohydrates, trehalose being found of special effectiveness due to its high $T_g$ [Hagan et al., 2005]. Therefore, initial attempts on the quench and refold approach consisted of using an 80 wt% trehalose and 20 wt% water solvent. However, as shown in Chapter 2, DSC and IR experiments reveal that lysozyme at such high trehalose concentration aggregated upon thermal treatment (Figures 2.2 and 2.11) and is not able to retain any activity (Figure 2.12). Further experiments described in Chapter 2 also reveal that hen egg white lysozyme dissolved in water alone aggregate upon thermal treatment above its $T_d$; therefore an excipient is needed to prevent aggregation.

In literature, researchers have shown [Summers and Flowers II, 2000; Kaftzik et al., 2002] that ionic liquids support reversible protein folding. Angell and coworkers recently demonstrated the generality of this behavior in protic ionic liquids within an optimal range of proton activity that was characterized using NMR proton chemical shift measurements [Byrne et al, 2007; Byrne and Angell, 2008]. Both groups also demonstrated that ionic liquids could prevent protein aggregation [Summer and Flowers II, 2000; Angell and Wang, 2003]. Another relevant finding was that protein aggregation could be prevented even at high protein concentrations (above 200 mg/ml) when protic ionic liquids are added [Byrne and Angell, 2008]. Such high protein concentrations allow for clear, large and smooth calorimetric peaks and scans when using a Pyris 1 DSC, and do not require the use of a microcalorimeter. Protic ionic liquids are also currently under intense study for their excellent stability, low volatility and unusual solvent properties that proved successful in chemical synthesis, enzyme catalysis and green chemistry [Rogers and Seddon, 2003; Byrne and Angell, 2008]. Though many
different types of protic ionic liquids or mixtures could be used [Xu and Angell, 2003], ethylammonium nitrate (EAN), \([\text{CH}_3\text{CH}_2\text{NH}_3^+][\text{NO}_3^-]\), was chosen to be added to the solvent to prevent protein aggregation when the protein was thermally treated. EAN is believed to be the first known room temperature ionic liquid [Allen et al., 1985] and is fairly simple to synthesize. The ionic liquid is measured to have a melting temperature of 13 °C using differential scanning calorimetry.

The mechanism behind EAN’s ability to prevent aggregation upon thermal treatment is not well understood. When lysozyme is thermally denatured, the hydrophobic core of the protein is exposed, but the disulfide bonds remain intact [Khechinashvili et al., 1973; Privalov & Khechinashvili, 1974; Griko et al., 1995; Ibara Molero & Sanchez-Ruiz, 1997]. Due to the exposed hydrophobic core, most globular proteins can then exhibit intermolecular association leading to aggregates and even to precipitation [Goldberg et al., 1991]. Figure 3.4 shows one theory for the prevention of protein aggregation by EAN.

Figure 3.4: Illustration of the proposed explanation for EAN’s ability to prevent aggregation during thermal denaturation and support refolding [Summers and Flowers, 2000].
This theory for the prevention of aggregation is that the ethyl group of EAN interacts with the hydrophobic portion of the protein and protects it from intermolecular association while the charged portion of the salt stabilizes the electrostatic interactions of its secondary structure [Kohn et al., 1997].

As part of the experimental development in the use of EAN, further studies were conducted on the relationship between EAN and lysozyme, and to determine the optimal EAN concentration to prevent lysozyme aggregation for use in subsequent quench-and-refold studies. The first inquiry was to determine the concentration limits that would fully prevent protein aggregation and precipitation. Figure 3.5 shows lysozyme dissolved in various amounts of EAN in water mixtures at different temperatures. The lysozyme concentration was 200 mg/ml. As one can see, lysozyme dissolved in pure water and in a 20 mol% EAN in water solvent results in a clear solution, which is also observed at higher temperatures, indicating complete dissolution and no aggregation. Since solution clarity is seen at high protein concentrations, one can also assume that the solution will also be clear at lower protein concentrations at both high and low temperatures. In a 10 mol% EAN solution, 200 mg/ml lysozyme appears to be insoluble at 25 °C.

However, when the solution was heated to 45 °C and beyond, the solution became clear showing dissolution of lysozyme at higher temperatures. Upon cooling back to room temperature, lysozyme returns to the same turbid solution as before heating. Lysozyme placed in 5 mol% EAN at 25 °C also results in a very turbid solution, indicating the insolubility of lysozyme in such a medium. Upon heating, the solution starts to become less turbid as some of the protein is dissolved at 55 °C.
Figure 3.5: Photos of the turbidity of 200 mg/ml lysozyme dissolved in water only and EAN-water solutions at different temperatures. Photos at 25 °C were taken after 30 minutes from sample preparations.

However the solution never became completely clear, and when heated to 75 °C, it turns completely white, even more turbid than before heating and almost solid-like, resulting in precipitation. This behavior is irreversible upon cooling back to room temperature, indicating that lysozyme associates and aggregates upon heating in a
medium of 5 mol% EAN. Therefore, from visual inspection of the solutions upon thermal treatment, only a solution consisting of 20 mol% EAN allows for full solubility of lysozyme at 200 mg/ml and does not exhibit protein association upon heating.

Differential scanning calorimetry of the above samples was conducted and representative scans are presented in Figure 3.6. In Chapter 2, it was shown that lysozyme unfolding in water heated above its denaturation temperature, T_d, is irreversible. In these DSC studies, lysozyme did not aggregate in 20 mol% EAN. This follows from the fact that the energy absorbed upon unfolding, ΔH_d, in successive heating scans in EAN is identical. T_d of lysozyme in 20 mol% EAN is 57.1 ± 1.2 °C with ΔH_d of 24.7 ± 1.3 J/g lys, which is significantly lower than both T_d and ΔH_d of lysozyme in water only, 75.8 ± 1.4 °C and 32.6 ± 1.9 J/g lys respectively, found in Chapter 2. There is an observed decrease in the denaturation temperature by ~19 °C and unfolding enthalpy by ~ 8 J/g lys with an increase in EAN in solution.

In 10 mol% EAN, lysozyme denaturation is still considered reversible, as seen by the overlapping unfolding energy peaks in Figure 3.6, despite the fact that at room temperature, the sample solution is slightly turbid. This indicates that the aggregated lysozyme in this case is not permanently associated as the solution clears up at 45 °C, well before the endothermic peak indicating protein unfolding. T_d of lysozyme is 64.6 ± 1.3 °C and ΔH_d is 27.7 ± 1.6 J/g lys in 10 mol % EAN. These values are lower than lysozyme in water, but higher than in 20 mol% EAN, following the trend that increasing EAN in solution, decreases both the T_d and ΔH_d.
As shown in Figure 3.6, lysozyme in 5 mol% EAN aggregates significantly upon heating above its $T_d$ of $69.0 \pm 2 \, ^\circ C$, where $\Delta H_d$ for unfolding is greatly decreased between “scan 1” and “scan 2” from $29.5 \pm 1.6 \, J/g$ lys to $7.5 \pm 0.7 \, J/g$ lys. DSC evidence matches well with the pictures of solutions in Figure 3.5 that show significant protein aggregation with increased solution turbidity and protein precipitation at 75 $^\circ C$. From the DSC results, an EAN-water solution with concentration greater than 10 mol% ensures prevention of lysozyme aggregation. Taking into consideration the visual observations and DSC experiments, a solution concentration of 20 mol% EAN was chosen for the quench-and refold experiments.
Though EAN prevents lysozyme aggregation, there may be a temperature that is high enough to cause lysozyme to still aggregate. A 200 mg/ml lysozyme in 20 mol% EAN solution was scanned repeatedly in a DSC, increasing the temperature with each cycle by 10 °C and holding the sample for 2 minutes at each high temperature. The denaturation temperature for each cycle did not change and stays at 57.1 ± 0.5 °C as shown in Figure 3.7.

![Heat Flow vs Temperature Graph](image)

Figure 3.7: Determination of the reversible and irreversible aggregation temperature of 200 mg/mL lysozyme in a 20 mol% EAN solution upon multiple scans. DSC heating and cooling scan rates were 20 °C/min.

By comparison of the unfolding enthalpies listed in Table 3.2, lysozyme does in fact begin to irreversibly aggregate after heating to 100 °C, with a 17.3 % decrease in enthalpy of unfolding in the fourth cycle, which implies that 17.3 % of the protein has aggregated. After heating to 110 °C, in cycle 5, 38.5% of the lysozyme is considered irreversibly aggregated. This irreversible aggregation temperature is over 40 °C above the
The denaturation temperature for our samples that consisted of a high lysozyme concentration (200 mg/ml). Previously literature studies showed that the temperature range between reversible and irreversible denaturation for a lysozyme concentration of 67 µg/ml in H$_2$O is 25 °C [Jasco, 2005]. Thus, using EAN provides a much greater temperature range for studies, within which there is no aggregation. Further experiments conducted in the lab revealed that decreasing protein concentration increases the temperature range between reversible unfolding and irreversible aggregation, which follows the observation from Chapter 2, where decreasing protein is found to decrease aggregation.

Table 3.2: Melting Temperature, Enthalpy and Aggregation Percentage from Figure 3.7 DSC Heating Scans.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Tm (°C)</th>
<th>ΔH (J/g)</th>
<th>Aggregation %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57.1</td>
<td>24.5</td>
<td>--</td>
</tr>
<tr>
<td>2</td>
<td>57.1</td>
<td>24.4</td>
<td>--</td>
</tr>
<tr>
<td>3</td>
<td>57.1</td>
<td>24.45</td>
<td>--</td>
</tr>
<tr>
<td>4</td>
<td>57.1</td>
<td>20.25</td>
<td>17.3 %</td>
</tr>
<tr>
<td>5</td>
<td>57.1</td>
<td>15.05</td>
<td>38.5 %</td>
</tr>
</tbody>
</table>

The demonstrated ability of EAN and other ionic liquids to prevent protein aggregation suggests that these systems have great potential for providing long-term stabilization of high-concentration protein solutions. The next step was to determine if lysozyme was still in fact active in EAN. Therefore an enzymatic activity assay was performed with lysozyme dissolved in D$_2$O, H$_2$O, 20 mol% EAN and 80 mol% H$_2$O, and in 20 mol% EAN and 80 mol% H$_2$O that have been heated above T$_d$ and cooled back to room temperature. Figure 3.8 shows that lysozyme retains full activity in all solutions.
with a calculated average activity of 45131 (± 3.1 %) units/mg of lysozyme compared to 45540 units/mg listed on the bottle. The activity assay results also reveal that neither EAN nor D₂O affects enzyme functionality. This is another experimental indication that lysozyme behaves in a fully reversible way upon heat treatment in 20 mol% EAN and 80 mol% H₂O.

![Figure 3.8](image)

**Figure 3.8:** Active units per mg of lysozyme dissolved in D₂O, H₂O, and 20 mol% EAN-80 mol% H₂O. The last bar shows the active units of lysozyme in 20 mol% EAN-80 mol% H₂O that was heated to 80 °C, held for 10 minutes at 80 °C, then cooled back to room temperature for assay. Solutions consisted of 200 mg/ml lysozyme.

However, EAN with water as a solvent matrix alone did not satisfy all the requirements needed for the quench-and-refold experiments, as ice crystallization is not suppressed during cooling. The presence of a considerable water content is usually considered necessary for the stabilization of folded protein structures as well as allowing for possible refolding upon denaturation [Angell and Wang, 2003]. Therefore, sucrose
was added to the mixture to prevent ice crystallization and so permit vitrification [Carpenter and Crowe, 1988; Crowe et al., 1998]. Addition of sugars, such as sucrose and glucose, also completely repressed the protein cold-unfolding phenomenon [Byrne and Angell, 2008]. If protein cold unfolding is observed, attempts at capturing protein folding intermediates by the proposed quench-and-refold approach would be invalidated. Previous experiments in literature have shown that a large denaturation heat capacity increase accompanied protein cold unfolding [Franks et al., 1988; Hatley and Franks, 1989]. In Figure 3.9, the DSC heating scans of lysozyme dissolved in EAN-water and in the presence of sucrose are shown. In the presence of sucrose, the DSC scan shows no evidence of an endothermic peak around 0 °C, typically indicative of ice crystals melting.

![Figure 3.9: DSC scans of 200 mg/ml lysozyme in 60 wt% EAN-water and 40 wt% EAN, 27 wt% H$_2$O and 33 wt% sucrose, showing ice crystallization only in the absence of sucrose.](image-url)
The addition of sucrose also did not affect protein aggregation as shown in Figure 3.10 when EAN was included in the solution matrix. The optimum solvent make-up was then determined to be a solvent composition (no protein) of 40 wt% EAN, 27 wt% H\textsubscript{2}O and 33 wt% sucrose, and the composition was the base for subsequent quench-and-refold studies in IR experiments. The pH of the chosen solvent is 6.7, and was determined using an Accumet AP71 pH detector and Digi-sense 5938 pH detector.

![DSC Heating Curves](image)

**Fig. 3.10:** Successive DSC heating curves of 200 mg/ml lysozyme in 40 wt% sucrose and 60 wt% H\textsubscript{2}O (no EAN) and in a solution of 40 wt% EAN, 27 wt% H\textsubscript{2}O and 33 wt% sucrose (EAN) showing the reversibility of lysozyme (unfolding–refolding) and prevention of aggregation using EAN in this formulation with sucrose [Cho et al., 2009].

The DSC scans of 200 mg/ml of lysozyme in 40 wt% EAN, 27 wt% H\textsubscript{2}O and 33 wt% sucrose solution reveals similar enthalpy peaks between 3 scan cycles of the same sample and lysozyme unfolds at 68.5 ± 1.1°C. The curves labeled “no EAN” in Figure
3.10 have a $T_m$ of 83.8 $\pm$ 0.9°C and provide further proof that EAN is needed to prevent aggregation, as the endotherm peak decreases in subsequent runs for lysozyme in 40 wt% sucrose and 60 wt% H$_2$O.

Previous DSC-based quench-and-refold studies used H$_2$O [Angell and Wang, 2003], however in our IR experiments, we used D$_2$O, and this is due to H$_2$O’s overlapping absorbance with the amide I’ region [Solomon and Lever, 1999]. As shown in Figure 3.11, there is a larger percent transmission peak region at the area of interest between 1600 to 1700 cm$^{-1}$ for H$_2$O compared to D$_2$O. The absorbance measurement seen in our IR studies is the logarithm of the inverse of the percent transmission. The H$_2$O absorbance greatly exceeded that of the protein sample, making accurate and useful measurements of the protein signal difficult. Therefore the low absorbance of D$_2$O is desired when subtracting solvent effects for IR analysis.

A DSC test was conducted in our lab to check if the exchange of D$_2$O and H$_2$O would affect protein reversibility when subjected to thermal treatment. The DSC upscans of lysozyme in EAN, sucrose and D$_2$O have similar enthalpy peaks between 3 scan cycles of the same sample, thus revealing that lysozyme is still fully reversibly in D$_2$O. $T_d$ also increases slightly to 70.7 °C, however the enthalpy remains equivalent at 28.15 $\pm$ 1.8 J/g. In IR experiments though, careful consideration of hydrogen-deuterium (H/D) exchange when the protein is unfolded must be taken into account when analyzing the data, as secondary structure assigned to characteristic wavenumbers is shifted to lower wavenumbers as described earlier in Chapter 2 [Mcbride-Warren and Mueller, 1972; Jung, 2000; Kong and Yu, 2007].

125
Figure 3.11: IR spectra of the percent transmission with respect to wavenumber of D$_2$O and H$_2$O. Area of interest is between 1600 – 1700 cm$^{-1}$, where a large peak is evident for H$_2$O. Percent transmission is the inverse of absorbance [Solomon and Lever, 1999].

In consideration of the IR absorbance limits and to avoid saturation of the sample spectra, a composition matrix for the IR quench-and-refold studies of 2 wt% lysozyme, 26 wt% D$_2$O, 39 wt% EAN and 33 wt% sucrose was utilized. In this solvent, lysozyme unfolds at ~ 70 °C and the solvent matrix T$_g$ was ~ -100 °C (Figure 3.1). Although the solvent composition is not exactly the same in the DSC and IR experiments, previous calorimetric measurements showed that T$_g$ varies by at most 10 °C in the presence or absence of EAN [Byrne and Angell, unpublished]. It is known that the native structure of a folded protein is dependent on its amino acid sequence and the solution environment to which it is dissolved [Anfinsen, 1973; Creighton, 1993]. Therefore, the secondary structure of lysozyme in our solution medium was compared to literature.

Figure 3.12 is a curve fit of the IR spectra for lysozyme in our solvent mixture at room temperature (25 °C). The peak positions in wavenumber identified by the second derivative analysis are 1680, 1674, 1666, 1652, 1640, 1627, 1620 and 1610 cm$^{-1}$. 

126
Lysozyme is predominantly $\alpha$–helical, as evident by the large peak at 1653 cm$^{-1}$. The higher wavenumbers at and above 1652 cm$^{-1}$ except 1674 cm$^{-1}$ are assigned to $\beta$–turns, while the peak at 1640 cm$^{-1}$ is related to random structures. 1674 cm$^{-1}$ and the lower wavenumbers at 1627 cm$^{-1}$ and below are associated with $\beta$–sheet structures [Luo et al., 1994, Kong and Yu, 2007]. As detailed in Table 3.3, comparison to literature with lysozyme in a neutral buffer reveals very similar $\alpha$-helix, $\beta$-sheet, turn, and random coil secondary structure contents [Luo et al., 1994] indicating that lysozyme in our solution is in a native state.

Figure 3.12: IR spectrum of deuterated lysozyme at 25 °C (2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O), after background subtraction and baseline correction between 1600 and 1700 cm$^{-1}$. Also shown are the individual Gaussian peaks used for quantitative analysis (dotted lines), and the resulting curve fit spectrum, which is indistinguishable from the original spectrum (solid line).
Table 3.3: Comparison of Protein Secondary Structures Determined by Fourier Transform Infrared Spectroscopy (FTIR) to Previous Studies.

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>X-ray(^1) Content %</th>
<th>FTIR(^2) Content %</th>
<th>FTIR(^3) Content %</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Helix</td>
<td>45</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>β-Sheet</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>turn</td>
<td>23</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>random</td>
<td>13</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

\(^1\)Levitt and Greer, 1977
\(^2\)Dong et al., 1990, Luo et al., 1994
\(^3\)This work

We note that some amino acid side-chains absorb in the amide I’ region. However, such side-chain modes are not directly sensitive to protein backbone conformation, and normally give rise to a very weak signal compared to the backbone carbonyl mode (Amide I). Many of these modes, furthermore, are shifted out of the 1600 - 1700 cm\(^{-1}\) region in D\(_2\)O [Venyaminov and Kalnun, 1990; Barth, 2000]. For example, the band positions (cm\(^{-1}\)) for ASN, GLN, ARG, and LYS in D\(_2\)O are: ASN [1648 (CO stretch)], GLN [1635-1654 (CO stretch); 1163 (NH\(_2\) in-plane bending); 1409 (CN stretch)]; ARG [1608 (CN\(_3\)H\(_5\)\(^+\) asymmetric stretch); 1586 (CN\(_3\)H\(_5\)\(^+\) symmetric stretch)]; LYS [1201 (NH\(_3\)\(^+\) asymmetric in-plane bending); 1170 (NH\(_3\)\(^+\) symmetric in-plane bending)].

3.2.2 IR Qualitative Analysis

Figure 3.13 represents a typical set of raw IR spectra from 1500 cm\(^{-1}\) to 1800 cm\(^{-1}\) of 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D\(_2\)O after heating and quenching the unfolded protein. The sample was first heated to 90 °C in an oven for 10 minutes to allow the protein to unfold and complete all H/D exchange. In protein IR
studies, focus and presentation is mainly placed on the amide I’ region from 1600 cm\(^{-1}\) to 1700 cm\(^{-1}\) as it is extensively characterized and the optimal region to determine protein secondary structure [Carpenter et al., 1998; Kong and Yu, 2007], however other regions can provide some useful general information.

IR spectra of protein in water-based buffers have an amide II peak centered at 1550 cm\(^{-1}\) [Dong et al., 1990; Jung 2000]. This amide II peak is associated with the bending vibration of the amine bond on the peptide backbone. However, since deuterium was present in solution, H/D exchange will occur and the amide II’ peak appears at \(~1450\) cm\(^{-1}\) as seen in Figure 2.6 [Jung 2000; Kong and Yu, 2007].

Figure 3.13: Original background unsubtracted IR spectra of a 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D\(_2\)O sample starting at \(-135\) °C following heating and quenching with liquid nitrogen and slowly heating at 2°C/min back to 25 °C with spectra taken every 10 °C.
In the raw spectra in Figure 3.13, there is no visible peak around 1550 cm\(^{-1}\). This confirms that full deuturation of the unfolded protein is complete. Incomplete deuturation would affect the lysozyme spectrum when a background spectrum using D\(_2\)O, EAN and sucrose (no water) is subtracted from the raw spectrum. The peak centered at 1760 cm\(^{-1}\) corresponds to EAN absorbance, as determined by IR spectra of pure EAN. Therefore to ensure full subtraction of the background leaving only the protein spectra, subtraction factors were determined and used when the peak at 1760 cm\(^{-1}\) was eliminated, revealing a flat baseline between 1725 cm\(^{-1}\) to 1800 cm\(^{-1}\). There is a small sucrose and EAN absorbance region within 1600-1700 cm\(^{-1}\), however this is also subtracted out.

Figure 3.14 shows raw spectra baselined from 1600 cm\(^{-1}\) to 1700 cm\(^{-1}\) of the solvent mixture itself, revealing small changes at different temperatures; therefore, careful background subtraction was executed using respective temperature spectral scans.
Figure 3.15 is a control experiment to ensure that spectral changes observed during the refolding of lysozyme at low temperatures were in fact due to the protein alone. This was done by taking the spectra of the lysozyme sample at room temperature and comparing it with folded protein that was slowly cooled at 100 K/min using the Linkam cooling unit in the IR, and with folded protein that was quenched by liquid nitrogen.

![Graph showing absorbance vs. wavenumber](image)

**Figure 3.15**: Spectra of folded lysozyme (2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D₂O) at 25 °C before any thermal treatment, and at -135 °C after slow-cooling (20°C/min), and quenching in liquid nitrogen. As can be seen, the IR spectra are very similar.

The spectra presented in Figure 3.15 and henceforth have already been background-subtracted by their respective solvent only temperature spectra, leading to the conclusion that any change in the spectra is due the protein. This also showed that cold unfolding of lysozyme does not occur in the presence of EAN and sucrose.
Figure 3.16 presents IR spectra that support the previously observed [Angell and Wang, 2003] ability to trap unfolded protein by quenching using LN2: note, in particular, that spectra right before quench (90 °C; thermally unfolded) and right after quench (-135 °C quench) are very similar.

Figure. 3.16: IR spectra of lysozyme (2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D2O) sample at 25 °C, then heated in the oven at 90 °C for 10 minutes, and then immediately quenched with LN2. The sample is then heated from -135 °C to 25 °C at 20 °C/min with 3 minute annealing intervals every 5 °C to collect spectra [Cho et al., 2009].

Once the unfolded protein was trapped in a glassy matrix, it was slowly heated at 20 °C/min and stopped every 5 °C to collect an IR spectrum. IR scans were collected to observe protein structural changes and intermediates as the protein refolded to the native state. Spectra collected at both 20 °C/min and 2 °C/min heating rates after liquid nitrogen
quench show no differences. It can be seen that the protein spectra before thermal treatment (25 °C) and after the entire quench-and-refold process (25 °C end) are very similar, producing spectral evidence of the reversibility of lysozyme unfolding and subsequent refolding.

A qualitative approach to analyze and study IR spectra data is to plot the temperature dependence of absorbance intensities at chosen frequencies (Figure 3.17), since characteristic frequencies are associated with particular protein secondary structures. Absorbance peaks around 1650 cm\(^{-1}\) are generally associated with \(\alpha\)-helices, while peaks around 1630 cm\(^{-1}\) are associated with intramolecular \(\beta\)-sheets [Surewicz et al., 1993].

![Absorbance vs Temperature](image)

Fig. 3.17: Absorbance level at 1652 cm\(^{-1}\) and 1627 cm\(^{-1}\), associated with \(\alpha\)-helix and \(\beta\)-sheets respectively, for initial states before heat treatment (1652 cm\(^{-1}\) green triangle; 1627 cm\(^{-1}\) blue circle) and after heating and quenching (1652 cm\(^{-1}\) blue diamonds; 1627 cm\(^{-1}\) pink squares). Error bars for 4 scans within the size of each data point [Cho et al., 2009].
Figure 3.17 also validates the fact that lysozyme unfolding is reversible in our solution medium as the intensity values before thermal treatment and after quench are the same. At ~ -100 °C, α-helices started forming. In contrast, the β-sheet content began to decrease at a higher temperature, ~ -65 °C. This observation reveals that refolding of protein in the glassy matrix occurs through intermediates, which aligns with previous studies showing multiple conformations in the lysozyme folding process [Matagne et al., 2000]. This observation of a continuous structural evolution represents a qualitatively new picture of the refolding of lysozyme following thermal denaturation. Varying onset temperature and temperature ranges where changes of intensity for representative absorbance levels occur reveal differences in the evolution of different secondary structural elements. α-Helix formation is reflected in a mild increase between -115 °C and -50 °C, whereas the onset of β-sheet change does not occur until approximately ~ -65 °C with a significant step change. Since the onset of α-helix structural changes occurs at a lower temperature than that of β-sheet, one can conclude that the α-domain of lysozyme folded earlier than the β-domain [Kiefhaber, 1995].

3.2.3 IR Quantitative Analysis

Of interest was the actual structural percentage composition of lysozyme from quench-and-refold spectra data. Quantitative analysis of IR spectra is based on the assumption that the protein could be considered a linear sum of several fundamental secondary structural elements. These fundamental secondary structural elements have been assigned to particular amide I’ band components by comparison of IR spectra with
X-ray crystal structures of proteins. Table 3.4 is a summary of the accepted and available range for assignments of secondary structure elements in the amide I’ band component for proteins in D$_2$O.

Table 3.4: Amide I’ Band Frequencies and Assignments for Protein Secondary Structure in D$_2$O$^a$

<table>
<thead>
<tr>
<th>Mean frequencies (cm$^{-1}$)</th>
<th>Assignment</th>
<th>Mean Frequencies (cm$^{-1}$)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1624 ± 4</td>
<td>β-sheet</td>
<td>1663 ± 4</td>
<td>turn</td>
</tr>
<tr>
<td>1631 ± 3</td>
<td>β-sheet</td>
<td>1671 ± 3</td>
<td>turn</td>
</tr>
<tr>
<td>1637 ± 3</td>
<td>β-sheet</td>
<td>1675 ± 5</td>
<td>β-sheet</td>
</tr>
<tr>
<td>1641 ± 2</td>
<td>random</td>
<td>1683 ± 2</td>
<td>turn</td>
</tr>
<tr>
<td>1653 ± 4</td>
<td>α-helix</td>
<td>1689 ± 2</td>
<td>turn</td>
</tr>
</tbody>
</table>

$^a$Kong and Yu, 2007

The amide I’ band contours are complex and consist of overlapping bands. To obtain quantitative results, it was necessary to separate the intrinsically overlapping bands in the amide I’ region. One way to accomplish this was by Fourier deconvolution of the original spectra [Byler and Susi, 1986]; a band narrowing process, which was achieved at the expense of spectral quality of the original band. However, due to the low signal to noise ratio (S/N) in the IR spectra of lysozyme in EAN, sucrose and D$_2$O collected from our studies (less than the required 500 S/N), another approach to obtain quantitative information by curve-fitting was implemented and is described below.

The important factors that needed to be considered in curve fitting were the number of component bands, their positions, shapes, widths and form of the baseline. The native structure of lysozyme was known from literature [Dong et al., 1990], and as shown in Section 3.2.1.2, lysozyme in EAN, sucrose and D$_2$O is in its native state, with IR spectra similar to those available in the literature. Therefore subsequent curve fittings
for the low temperature spectra used constraints resulting from the known native state curve fit and respective second derivative spectrum that was smoothed with an eleven point Savitsky-Golay function to obtain a systematic protocol. From Figure 3.12, 4 distinct and strong peak positions and widths (FWHH: full width at half height of the fitted peak) relating to different fundamental secondary structure elements from the native state fit were chosen as constraints for low temperature spectra. Table 3.5 summarizes the constraints posed.

Table 3.5: Constraints for Subsequent Quantitative Curve Fitting of Quench-and-Refold Spectra Taken From Native State Curve Fitting.

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>Peak location</th>
<th>FWHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>turns</td>
<td>1666 cm(^{-1})</td>
<td>15.4 cm(^{-1})</td>
</tr>
<tr>
<td>(\alpha)-helix</td>
<td>1652 cm(^{-1})</td>
<td>21.9 cm(^{-1})</td>
</tr>
<tr>
<td>random coil</td>
<td>1638 cm(^{-1})</td>
<td>16.7 cm(^{-1})</td>
</tr>
<tr>
<td>(\beta)-sheet</td>
<td>1627 cm(^{-1})</td>
<td>20 cm(^{-1})</td>
</tr>
</tbody>
</table>

FWHH was limited in the curve-fitting software, Grams, such that the fitted peaks for the unfolded protein or its intermediates did not have FWHH greater than that of the native state at the chosen fixed peak positions; this allowed for only the peak intensities to vary and resulted in the best match to the original spectra. Aside from the 4 restricted peaks, curve-fitting iterations began with the same peak locations and FWHH as those found in the native state. However, the intensity of these peaks was allowed to vary for the best fit. As the protein structure deviated from the native state, it was necessary to add peaks to obtain the best match. The peak additions were done using information from their respective second-derivative spectra. The addition of peaks for fitting posed a limitation to the quantitative analysis, as will be explained below. However, this
approach provided a systematic quantitative analysis of the structural changes that occurred during lysozyme refolding in regards to the 4 specific peaks used as constraints.

Figure 3.18 is an example of the second derivative and curve fitting of IR spectra for lysozyme at 25 °C, and for lysozyme that was thermally denatured, quenched, and slowly heated to -65 °C. It can be seen that there are significant visual differences in peak intensities between the native state and the structural intermediate at -65°C.

Figure 3.18: Comparison of the second derivative spectrum (top) and the curve fitting of spectra (bottom) for 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D_{2}O at 25 °C (folded) and at -65 °C after thermal denaturation, quench, and slow heating across T_{g}.
Two additional peaks were needed for curve fitting at -65 °C, which were determined by examining the second derivative spectrum, shown above the original spectra in Figure 3.18. With such constraints, curve fitting of subsequent spectra was accomplished with confidence for spectra closer to the native state (-75°C to 25°C). Quantitative analysis below -75°C was difficult given the 4 peak constraints imposed. One reason for the breakdown of the procedure is that the peak locations used for curve-fitting with the 4 peak constraints imposed did not match peak locations indicated by the second derivative spectrum. This is evident in Figure 3.19, which compares the curve-fitting peaks for the IR spectra of lysozyme at -135°C (after thermal denaturation and quench but before any subsequent heating for refolding) with and without the 4 constraints imposed on its second derivative curve. Spectra between -135°C and -75°C could still be fitted without the constraints (Figure 3.19); however, it is difficult to validate which peak correlated to a particular structure, as the fit is not unique.

Table 3.6 lists the secondary structure content percentages determined using the curve fitting protocol proposed for all 4 peaks of interest. The percent of secondary structure assignment calculated at 1666 cm⁻¹ for the turn structure increases upon folding, as did the α-helix assignment at 1652 cm⁻¹. On the other hand, the amount of random coil at 1638 cm⁻¹ decreases as the protein refolded. β-sheets are also seen to decrease upon folding.
Figure 3.19: Example of quantitative analysis procedure breakdown seen at very low temperatures due to very different peak areas fitted to the same spectra. Curve fitting of spectra of 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, 26 wt% D$_2$O at -135 °C after heat and-quench.

Table 3.6: Sample Secondary Structure Percentages Determined by Curve Fitting of Quench-and-Refold Spectra.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>1666 cm$^{-1}$</th>
<th>1652 cm$^{-1}$</th>
<th>1638 cm$^{-1}$</th>
<th>1627 cm$^{-1}$</th>
<th>% Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 °C initial</td>
<td>8.1</td>
<td>42.6</td>
<td>12.7</td>
<td>11.4</td>
<td>25.2</td>
</tr>
<tr>
<td>25 °C final</td>
<td>8.3</td>
<td>41.9</td>
<td>13.1</td>
<td>10.8</td>
<td>25.9</td>
</tr>
<tr>
<td>-15 °C</td>
<td>16.7</td>
<td>33.6</td>
<td>20.7</td>
<td>12.1</td>
<td>16.9</td>
</tr>
<tr>
<td>-65 °C</td>
<td>14.6</td>
<td>29.8</td>
<td>24.6</td>
<td>12.1</td>
<td>18.9</td>
</tr>
<tr>
<td>-75 °C</td>
<td>15.1</td>
<td>28.8</td>
<td>25.4</td>
<td>17.1</td>
<td>13.6</td>
</tr>
</tbody>
</table>

Error is within ± 10% for 4 different samples.
Figure 3.20 plots the quantitative data from the intermediate state at -75°C back to the folded state at 25°C; α-helix and β-sheet follow the same qualitative trend of Figure 3.17, where α-helix increases and β-sheet decreases, though subtle differences are observed. The percentage of α-helix content is shown to increase almost linearly between -30 °C and 25 °C and a change in the trend for β-sheet percentage is observed at -65 °C in both the qualitative and quantitative analyses. On the other hand, qualitative analysis reveals a further decrease in β-sheet percentage around 0 °C, where quantitative analysis does not.

Figure 3.20: Secondary structure content percentage at 1652 cm⁻¹ (α-helix) and 1627 cm⁻¹ (β-sheets). Error bars for 4 scans are shown. Quantitative analysis trend for the folding of lysozyme is similar to that seen qualitatively in Figure 3.16.
Differences in the evolution of specific secondary structures provided through qualitative and quantitative analysis of quench-and-refold of lysozyme are observed. For example, a decrease in both absorbance (Figure 3.17) and percentage of secondary structure (Figure 3.20) between -75 °C to -65 °C is seen for β-sheet formation whereas α-helix content does not change over the same temperature range. The absence of changes that are all or nothing implies that protein intermediates exist upon refolding, suggesting a complexity or richness of secondary structural formations that needs to be further investigated. This subtle evolution of specific secondary structural change over a continuum from low-temperature spectra analysis has not been widely observed before as most other techniques simply report the percent of protein that refolds to its native state over time [Dobson et al., 1994].

3.2.4 Comparison of DSC and IR

Comparison of DSC and IR data using the quench-and-refold approach reveals an interesting difference in onset temperature of observable changes during refolding. DSC reheating curves do not show any enthalpic relaxation (refolding) until 0 °C (Figure 3.1) whereas spectral shifts and qualitative changes in secondary structure content are observed at much lower temperatures starting as low as -110°C (Figure 3.17). By contrast, IR spectra of 2 wt% lysozyme, 39 wt% EAN, 33 wt% sucrose, and 25 wt% D2O heated through the unfolding temperature range complement the calorimetric findings. Results from both techniques insinuate that the unfolding of protein by thermal denaturation is a cooperative process. In Figure 3.21, qualitative changes in secondary structure of the α-helix signature at 1652 cm⁻¹ and enthalpy changes are observed over
the same temperature range. The temperature associated with the DSC peak maximum and 50% change in absorbance level is approximately 70 °C. This implies that the secondary structural changes during lysozyme unfolding upon thermal treatment are accompanied by an enthalpy change.

![Graph showing absorbance and heat flow changes](image)

Figure 3.21: Changes in absorbance level at 1652 cm\(^{-1}\) (α-helix) for thermal unfolding of lysozyme in 40 wt% EAN, 33 wt% sucrose and 27 wt% D\(_2\)O (diamonds). Note that the unfolding temperature range seen in the IR is similar to that found calorimetrically (dotted line). IR and DSC heating scans were both at 20 °C/min.

Since DSC and IR measurements reveal an important gap in the temperature at which refolding-related signals are manifested upon heating, perhaps the quench-and-refold approach using IR is able to capture secondary structural changes during refolding.
that are energy-neutral. Further experimental studies were completed to elucidate this difference between DSC and IR measurements. These studies included a more detailed look at the protein concentration, solvent composition and properties, and kinetics using lysozyme as the model protein.

3.2.4.1 Protein Concentration

Since the protein concentrations differ by an order of magnitude (200 mg/ml in DSC, 20 mg/ml in IR), it was decided to test whether lower biomolecular concentrations, with concomitant decrease in protein–protein interactions, would shift the DSC onset of refolding towards lower temperatures. Figure 3.22 shows DSC heating curves of quenched unfolded and quenched folded protein for three different lysozyme concentrations.

The DSC heating curve of the quenched unfolded protein shows an extended enthalpic relaxation with no distinct peak roughly between 0 and 60 °C, providing evidence of non-cooperative folding. Areas for the enthalpic relaxation and the unfolding peak are similar at 28.5 ± 1.8 J/g, thus indicating that the quenching of lysozyme using LN₂ is able to successfully trap a fully unfolded protein at various protein concentrations. Note that calorimetric signatures of refolding of the quenched-unfolded sample all began at ~0 °C. Thus, changing the protein concentration does not cause a systematic shift in the calorimetric onset of refolding and establishes that the difference in refolding behavior is not due to the difference in protein concentration.
3.2.4.2 Solvent Composition

In the DSC quench-and-refold studies presented thus far, the onset temperature for the protein refolding exotherm is observed at 0 °C. Byrne and Angell by private communication and collaboration show in Figure 3.23 that even changing the type of ionic liquid or sugar does not significantly change the onset temperature for the refolding exotherm.
Figure 3.23: DSC quench-and-refold scans of 25 wt% lysozyme in solvent mixtures of 25 wt% water and the following components: [A] 25 wt% Triethylamine sulfonate (TEAMS) and 25 wt% sucrose [B] 25 wt% Hydroxyethylammonium nitrate (HOEAN) and 25 wt% sucrose [C] 25 wt% ethylmethylimidazolium nitrate and 25 wt% sucrose [D] 25 wt% 1:1 mix of ethylammonium perchlorate and triethyl-amonium triflouroacetate and 25 wt% glucose. Heating scans were run at 20 °C/min [Byrne and Angell,, unpublished data].

The ionic liquids chosen has a measured protic activity within a particular range that allows for full reversibility of protein [Byrne et al., 2008]. Changing the type of ionic liquid influences the $T_d$ and the energy required to unfold the protein.

The ionic liquids used in panels [A], [B] and [D] of Figure 3.23 are protic, whereas the one used in panel [C] is aprotic. One major difference seen in [C] compared to the other scans is that the onset of refolding is less sudden. This could be an indication of the role played by hydrogen bonding in the refolding process and energetics, as aprotic
ionic liquids are unable to hydrogen-bond with the unfolded protein. Though sugars have been found to hydrogen-bond to and stabilize proteins at low hydrations [Carpenter and Crowe, 1989; Prestrelski et al., 1995; Crowe et al., 1998; Carpenter et al., 1999], sugars are thought to be preferentially excluded from the protein in solution, which would result in minimal hydrogen bonding between sugar and protein upon folding [Arakawa and Timasheff, 1982; Lee and Timasheff, 1981; Xie and Timasheff, 1997]. In [D], glucose was used instead of sucrose, which does not lead to a significant change in onset temperature. Thus, the energy relaxation starting at 0 °C in DSC quench-and-refold experiments seems to be a general observation.

3.2.4.3 Solvent Evaluation

The studies reported in Section 3.2.4.1 and 3.2.4.2 focused on confirming if the onset of enthalpy changes seen in the DSC with quench-and-refold runs always occurred at 0 °C by way of changing protein concentration, ionic liquid or sugar. In this section, a closer evaluation of the solvent properties itself was conducted to determine whether the solvent could be responsible for inducing the change seen at 0 °C in DSC scans or any structural changes seen in IR spectra at low temperatures.

One consideration in the development of the quench-and-refold technique was to take advantage of the high viscosity of fluids in the vicinity of the glassy state to trap intermediate structures, as well as to quench the unfolded protein in a glassy state. Low molecular mobility of protein in glassy matrix has been found in literature [Franks, 1993]. Again, DSC quench-and-refold heat scans reveal a change in enthalpy at 0 °C, and one thing to consider was whether the viscosity of the solvent significantly changes at
0°C, causing the deviation. Figure 3.24, modified from Byrne et al. 2007, plots the viscosity of various solvent combinations including EAN, sugars and water [Byrne et al, 2007; Longinotti and Corti 2008].

From Figure 3.24, the viscosities of only sugar-EAN-water solutions (no proteins) are only about one order of magnitude greater than that of only water. This difference is sufficiently low such that the refolding of protein in sugar-EAN-water solutions is not significantly affected by viscous retardation of molecular processes at temperatures above 0 °C. The viscosity changes notably in a solvent with no EAN (64 wt% sucrose-H₂O) at temperatures much lower than 0 °C; however, a significant change or jump in

Figure 3.24: Comparison of the viscosity for different mixtures of water, sucrose, glucose and EAN [modified from Byrne et al, 2007; Longinotti and Corti 2008].
viscosity at 0 °C is not observed. Therefore, viscosity changes near 0 °C is not considered the reason for the sudden release in enthalpy at 0 °C observed in DSC quench-and-refold lysozyme experimental scans.

As a further test to ascertain if the deviation at 0 °C is related to the solvent, isothermal titration calorimetry (ITC) was performed to determine whether the observed onset of enthalpy change upon refolding is related to EAN-water interaction. EAN at 22.3 mM was titrated into pure water at 25 °C. Injection of water into water as an instrumental control revealed negligible heat due to mixing. Enthalpy release from titrating EAN into water reveal a constant value per injectant at 13 cal/mol. EAN at 22.3 mM was then titrated into a 1.36 mM lysozyme in water solution to observe whether lysozyme and EAN interacted. Figure 3.25 shows the excess heat released due to mixing of lysozyme and EAN in water.

Excess heat was determined by subtracting the constant enthalpy from injecting EAN into pure water from the enthalpy detected from injecting EAN into a solution containing lysozyme and water. The enthalpy of interaction between lysozyme and EAN is seen to decrease with subsequent injection of EAN with a starting enthalpy release of approximately 0.38 kcal/mol. If we continue with more injections, nearing infinite dilution of lysozyme in the cell, the excess heat approaches 0 kcal/mol. Since the partial molar heat of dilution in Figure 3.25 [B] reveals only a small decrease with each injection, this implies that there is a small interaction between lysozyme and EAN at room temperature. To relate to the concentrations used in DSC quench-and-refold experiments (20 wt% lysozyme, 32 wt% EAN, 21 wt% water and 27 wt% sucrose), we have approximately 215 moles of EAN to 1 mole of lysozyme, which is well beyond the
scope of this experiment due to volume limitations in the ITC. Therefore, this lysozyme and EAN interaction is considered insignificant and not a reason for the sudden exotherm starting at 0 °C observed in DSC quench-and-refold experiments.

Figure 3.25: Raw ITC [A] and normalized [B] excess heat release of lysozyme and EAN binding from titrating 22.3 mM EAN into 1.36 mM lysozyme in water.
It can be seen from the DSC-cooled and LN$_2$-quenched solvents composed of 40 wt% EAN, 27 wt% H$_2$O and 33 wt% sucrose, and 60 wt% sucrose and 40 wt% H$_2$O (no EAN), with no protein present, (Figure 3.26) that neither scans show an energy relaxation at 0 °C, generally seen in DSC quench-and-refold heat scans in samples containing lysozyme.

Figure 3.26: DSC heat scans after DSC-quench (dotted line) and LN$_2$ quench (solid line) of [A] 40 wt% EAN, 27 wt% H$_2$O and 33 wt% sucrose and [B] 60 wt% sucrose and 40 wt% H$_2$O. Note that when LN$_2$ quenched, an endothermic peak exists around -50°C, whereas the DSC quenched heat scans do not show such an endothermic peak.
This seems to indicate that the solvent does not promote the exotherm at 0 °C. On the other hand, a small endothermic peak can be seen around -50 °C for the LN2-quenched upscans, whereas the DSC-cooled heat scans do not show such an endothermic peak, neither in 40 wt% EAN, 27 wt% H2O and 33 wt% sucrose, nor in a 60 wt% sucrose-water solvent. This small endothermic peak at -50 °C could indicate that the faster cooling of the solvent using LN2, as opposed to cooling in DSC, traps a non-equilibrium structure that melts at -50 °C. These results suggest that DSC quenching was slow enough to form a homogenous glass, whereas LN2 quenching forms a polyamorph.

Figure 3.27 [A] reveals that the endotherm at -50 °C is also observed in raw (not solvent background subtracted) DSC up scans of lysozyme in 40 wt% EAN, 27 wt% H2O and 33 wt% sucrose. The endothermic peak can then be successfully subtracted out revealing only the protein energetic changes that manifested themselves at 0 °C. Upon decrease of the water content to a solvent composition of 47 wt% EAN, 19 wt% H2O and 34 wt% sucrose, an endothermic peak around -50 °C is not visible in neither the LN2-quenched nor DSC-cooled scans. Therefore it seems that the endothermic peak at -50 °C is affected by the water content. In either case, the solvent background could be successfully subtracted out from the protein DSC scans and the solvent itself does not exhibit exotherms at 0 °C.

Of particular interest in connection with this endotherm peak found in the DSC was to determine if it affected and promoted the changes seen in the IR before and around -50 °C. As seen in Figure 3.17, the majority of α-helix is formed before -50°C and no changes occur across -50 °C, while there is a definite change in β-sheet below -50 °C, and minor changes occur across -50 °C.
Figure 3.27: DSC raw quench-and-refold scans of [A] 20 wt % lysozyme, 32 wt% EAN, 21 wt% H₂O, and 27 wt% sucrose (solid line) and 40 wt% EAN, 27 wt% H₂O and 33 wt% sucrose (dotted line), both LN₂ quenched [B] DSC scans comparing DSC quenched and LN₂ quenched solvent mixture of 47 wt% EAN, 19 wt% H₂O and 34 wt% sucrose, revealing a straight baseline, and no endothermic peak as seen in [A] at -50 °C.

Qualitative analysis of secondary structural changes in IR spectra for the quench-and-refold of lysozyme in the lower-water-content solvent is compared to that in 40 wt% EAN, 27 wt% H₂O and 33 wt% sucrose in Figure 3.28. As one can see, the qualitative
trends of structural changes are the same, indicating that the endothermic peak at ~ -50 °C does not induce IR spectral changes, especially since α-helix formation across -50 °C is still consistent.

Therefore, the difference in both DSC versus IR-determined onset temperature for folding is still observed and is not affected by whether the solvent exhibits an endothermic peak. The protein energetic and structural changes are captured by the successful subtraction of their respective solvents. This suggests that IR quench-and-refold experiments are able to capture energy-neutral structural changes.

Figure 3.28: Visual comparison of the qualitative secondary structural changes for 2 wt% lysozyme and 98 wt % solvent of 2 different compositions: 40 wt% EAN, 27 wt% D₂O, 33 wt% sucrose (filled) and 47 wt% EAN, 19 wt% D₂O, 34 wt% sucrose (empty). Red filled shapes indicate the initial state before quench-and-refold.
3.2.4.4 Kinetics

Kinetic studies were conducted in an attempt to further understand the differences seen in the IR and DSC, since the DSC quench-and-refold upscans were heated at a rate of 20 °C/min and most IR samples were heated at 2 °C/min. Folded and thermally denatured quenched lysozyme samples in EAN, sucrose and H₂O composition were heated at 20 °C/min to different temperatures and then annealed for 1 hour (the approximate time difference in the upscans between IR and DSC in the temperature range of -135 °C and 0 °C). After annealing, samples were quenched again from the annealing temperature and heat scans were then recorded. Figure 3.29 shows the resulting DSC upscans for a sample consisting of 20 wt% lysozyme, 38 wt% EAN, 15 wt% H₂O and 27 wt% sucrose. Unfolding and refolding enthalpies for each resulting DSC upscan are listed in Table 3.7.

<table>
<thead>
<tr>
<th>Anneal Temperature (°C)</th>
<th>Unfolding Enthalpy (J/g lys)</th>
<th>Refolding Enthalpy (J/g lys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 °C</td>
<td>29.0</td>
<td>0</td>
</tr>
<tr>
<td>-5 °C</td>
<td>29.2</td>
<td>7.9</td>
</tr>
<tr>
<td>-15 °C</td>
<td>29.3</td>
<td>16.8</td>
</tr>
<tr>
<td>-25 °C</td>
<td>29.1</td>
<td>29.2</td>
</tr>
<tr>
<td>-55 °C</td>
<td>28.9</td>
<td>28.8</td>
</tr>
<tr>
<td>-135 °C</td>
<td>29.3</td>
<td>29.0</td>
</tr>
</tbody>
</table>
Figure 3.29: Effect of annealing quenched samples of 20 wt % lysozyme, 38 wt% EAN, 27 wt% sucrose and 15 wt% water. Solid lines represent separate samples for each anneal temperature shown that were heated for thermal denaturation and then quenched with LN₂. After quenching, samples were heated at 20 °C/min to the anneal temperature, and then held at this temperature for 1 hour. After annealing at each temperature, samples were rapidly quenched again before above DSC heat scans were collected. Dotted lines represent DSC-cooled samples of folded protein.

Annealing for 1 hour at temperatures between -135 °C and -25 °C does not reveal any refolding or unfolding energetic changes in the DSC upscans as the integral of the exothermic deviation between 0 and 60 °C is found to be equal (within experimental error) to the endothermic peak centered at 62.3 ± 1.4 °C. This confirms the idea that the exothermic deviation between 0 and 60 °C is the refolding energy of lysozyme after thermal denaturation and quench. These low temperature annealed scans also reveal a slight curvature between ~ -90 °C and 0 °C. The large exotherm is still observed when
lysozyme was annealed at -25 °C, with the integration of the unfolding and folding energy still the same within experimental error. However, when lysozyme was annealed at higher temperatures above -25°C, there is noticeable loss in the exotherm between 0 °C and ~ 60 °C. Annealing above -25 °C released ‘trapped’ enthalpy hidden when samples were scanned at that faster rate of 20 °C/min. Therefore, protein structural changes at low temperatures below -25 °C could be associated with energy-neutral changes that occur at very fast times in the initial folding phase.

Thermal anneals were conducted in the IR for 2 wt % lysozyme, 39 wt% EAN, 33 wt% sucrose and 26 wt% D2O samples to determine whether intermediates captured by quench-and-refold are considered “equilibrium states” under our experimental protocol. Samples were heated at 20 °C/min to the anneal temperature and then held for 3 hours (maximum time to control the sample at a given temperature using the Linkam sample holder due to the size of LN2 Dewar system). Spectra were collected every 15 minutes. IR spectra and secondary structural percentages of lysozyme at respective temperatures upon heating at 20 °C/min and 2°C/min after LN2 quench are found to be similar. Upon heating of protein after quenching, IR spectral changes are observed at temperatures as low as -115 °C. However, when the sample was held at -115 °C, subsequent changes in the IR spectra are not observed within the experimental time limit of 3 hours. Therefore at -115 °C, the intermediate structure could be said to be in equilibrium.

Anneals at higher temperatures were conducted, and only at a temperature at or above -15 °C are changes in IR spectra observed. This implied that the viscosity of the solution below -25 °C is sufficient to keep the intermediate structure unchanged within the experimental time limit of 3 hours. Qualitative analysis of the α-helix structure in
Figure 3.30 indicates that protein completely folded within the experimental time limit for an anneal at 5 °C, and that in 3 hours, the protein annealed at -5 °C, which is already mostly folded (>90 %), almost returns back to its native state.

![Graph showing absorbance ratio vs. anneal time]

Figure 3.30: Qualitative changes for the refolding of α-helical secondary structure for 2 wt % lysozyme, 39 wt% EAN, 33 wt% sucrose and 26 wt% D₂O after thermal anneals. Lysozyme was thermally denatured, and then quenched in LN₂. IR samples were then heated at 20 °C/min to -25 °C, -15 °C, -5 °C or 5 °C and held for 3 hours. An IR spectrum was collected every 15 minutes. Changes were determined by the ratio of the absorbance at a particular anneal temperature (Absorbance_AT) and the absorbance of lysozyme in its native state (Absorbance_N).

Comparing the annealed samples of IR spectra and DSC scans at temperature above -25 °C reveals that the structural changes that occurred during annealing can be correlated with the loss of exotherm found above 0 °C. Although, the reason behind the sudden start of enthalpy relaxation at 0 °C is still in question, the anneal experiments provide further evidence for the observation of energy-neutral structural changes.
The quench-and-refold approach applied to both DSC and IR appears to be able to capture the formation of secondary structural elements before energetic native contacts of the hydrophobic core fall into place. Such energy-neutral rearrangements have not been observed previously in IR. Structural changes such as the fast hydrophobic collapse [Lapidus et al., 2007] can occur at very short time scales and in most cases during other experimental technique’s ‘dead time’. In our studies, these very short time scales appear to be associated with phenomena occurring at low temperatures. Temperature-jump kinetic studies of hydrophobic collapse revealed a lack of activation energy at such short time scales [Sadqi, et al., 2003]. The lack of corresponding calorimetric signal suggests that the refolding events seen in the IR at low temperatures are likewise energy-neutral.

3.3 Conclusion

In summary, we have developed a sensitive method to track the protein intermediate structures upon folding from the fully unfolded state back to the native state. Lysozyme was used as our model protein in a solvent medium of EAN, sucrose and H₂O/D₂O. In such a medium, lysozyme is still a functional and active enzyme. The main advantage of this technique is the opportunity to observe subtle structural changes as they evolve over a continuum in time, due to the slow kinetics found in the viscous liquid state near T_g.

An interesting observation is seen when comparing quench-and-refold IR and DSC results, where the temperature of detection for protein refolding varies by almost 100 °C. Further studies were completed on the system to understand this difference. Solvent viscosity and protein concentration are determined not to have an effect on the
consistently observed start of enthalpy release at 0°C in quench-and-refold DSC results. Though EAN and lysozyme are found to weakly interact and bind, systematic studies of the solvent itself reveal that background subtractions for DSC and IR data analyses are sufficient in isolating the major changes in protein structure and energy relaxation. DSC quench-and-refold of the solvent itself does not exhibit energy relaxation at 0 °C.

Thermal annealing experiments of lysozyme in EAN, sucrose and H2O/D2O at varying low temperatures reveal that intermediates held at temperatures below -15 °C are in equilibrium within the experimental time frame, where anneals conducted at higher temperatures result in protein refolding. Anneals performed at temperatures below -25 °C reveal that any structural changes observed in the IR are in fact energy neutral, due to the lack of energy loss in the refolding peak, similarly indicating that structural changes occurring above -25 °C have an associated energy component.

The combined use of IR and DSC reveals the ability of IR to capture early structural changes that are energy-neutral using the quench-and-refold approach. Such energy-neutral rearrangements have not been observed previously in IR.
3.4 References


Byrne N, Angell CA. unpublished data.


     Annu. Rev. Physiol. 60:73-103.
Demetrius L. 2002. Thermodynamics and kinetics of protein folding : an evolutionary
Dobson CM, Evans, PA, Radford, S. 1994. Understanding how protein fold: the
     Methods. 34:4-14.
Dong A, Huang P, Caughey WS. 1990. Protein secondary structures in water from the
     second derivative amide I infrared spectra. 29:3303-3308.
Dong AC, Prestrelski SJ, Allison SD, Carpenter JF. 1995. Infrared spectroscopic studies
     of lyophilization-induced and temperature-induced protein aggregation. J. Pharm.
     Sci. 84:415-424.
     29:327-359.
Franks, F., Hatley, R. H. & Friedman, H. L. 1988. The thermodynamics of protein

renaturation and aggregation during the refolding of denatured-reduced egg white

Griko YV, Freire E, Privalov G, Dael HV, Privalov PL. 1995. The unfolding
thermodynamics of C-type lysozyme—A calorimetric study of the heat


Hatley, R. H. & Franks, F. 1989. The cold-induced denaturation of lactate dehydrogenase
at sub-zero temperatures in the absence of perturbants. FEBS Lett.

Ibara-Molero B, Sanchez-Ruiz JM. 1997. Are there equilibrium intermediate states in the
urea-induced unfolding of hen egg-white lysozyme? Biochemistry 36:9616 –


Jasco Inc. 2005. Thermal denaturation of hen-egg lysozyme with concurrent CD and
Fluorescence Detection. CD Application Note 01-03.

Jung C. 2000. Insight into protein structure and protein-ligand recognition by Fourier


Chapter 4

Structure-Energy Relations in Bovine Pancreatic Ribonuclease A Observed During Refolding From a Quenched Unfolded State

Understanding and elucidating protein-folding mechanisms is an important research area. In the previous chapter, a novel experimental technique to study protein folding intermediate structures is introduced and described in detail. This quench-and-refold experimental approach was first applied to differential scanning calorimetry (DSC), and then extended to infrared spectroscopy (IR), which reveals the ability to trap and observe different intermediate structures of lysozyme. Energy-neutral structural changes are observed in the early stages of folding, which is equivalent to lower temperatures in the quench-and-refold experiments. This is indicated by the difference in onset temperatures at which structural and energy changes were detected in the IR and the DSC, respectively. In order to determine whether this phenomenon and difference between IR and DSC is also observed in other protein, another slow-folding protein was
chosen for further experimental studies, under the same DSC and IR experimental conditions and procedures as presented in Chapter 3.

Bovine pancreatic ribonuclease A (RNase A) is one of the classic model systems of protein science. It is an endonuclease that cleaves single-stranded RNA, and contains 124 residues. RNase A has 4 disulfide bonds and a molecular weight of 13.7 kDa. Unlike lysozyme, which is mostly $\alpha$-helical, RNase A consists mostly of $\beta$-sheets: $\sim$15% $\alpha$-helix and $\sim$40% $\beta$-sheets [Dong et al., 1990]. RNase A can be reversibly denatured, refolding when returned to conditions that favor the native structure. Previous RNase A protein folding studies have determined that folding intermediates exist and that complete folding is completed in 400 seconds [Adler and Scheraga, 1988; Schmid and Blascheck, 1981]. In fact, RNase A thermal unfolding was found to deviate from 2-state behavior [Stelea et al., 2001].

In Chapter 3, it is shown that the optimal solvent for lysozyme studies consists of a mixture of ethylammonium nitrate (EAN), sucrose and water, which prevents aggregation and allows for vitrification in the quench-and-refold samples. Solvent composition is shown to have a great effect on the reversibility of protein folding, and extensive experiments have been done to study and understand the mechanism behind protein stability or instability. Protein solution denaturants, such as urea and guanidine HCl, are found to increase protein aggregation, decrease protein activity and decrease protein denaturation temperature. These excipients preferentially bind to proteins, destabilizing their native conformation [Wang 2005]. On the other hand, protein solution stabilizers, such as sorbitol [Xie and Timasheff, 1997], sucrose [Lee and Timasheff, 1981], trehalose [Xie and Timasheff, 1997], and arginine [Arakawa et al., 2007] to name
a few, prevent aggregation, increase denaturation temperature, increase the enthalpy change of unfolding, and allow the protein to retain most of its activity. Ionic liquids have also been found to prevent protein aggregation and retain activity [Lange et al., 2005; Fujita et al., 2006; Byrne and Angell, 2008; Byrne et al., 2009]. Structural studies using spectroscopic methods such as circular dichroism or infrared spectroscopy in the presence of denaturants or stabilizers mostly focus on the spectral changes from the native state [Johnson 1988; Dunbar et al., 1997], and on protein aggregation kinetics. Kinetics and structural deviation from the native state are determined by observing changes in the absorbance level of a spectral peak normally associated with intermolecular β-sheet formation, an indication of aggregated protein [Kato et al., 1981; Dong et al. 1994; Chi et al., 2003; Yan et al., 2006]. Researchers have also studied solvent effects on protein folding kinetics, mostly by chemical denaturation [Itzhaki and Evans, 1996; Went et al., 2004; Maxwell et al., 2005], however, the effect of the solvent on the mechanism of secondary structure formation during protein folding after thermal denaturation has not, to our knowledge, been investigated. With the use of the quench-and-refold approach, one objective of this RNase A work is to study how a solvent can affect a continuous evolution of structure. The results described in Chapter 3 and 4 could provide insight into protein aggregation or misfolding.

This chapter is organized as follows: additional materials and experimental methods not present in Chapter 3, are described in Section 4.1. Results are presented and discussed in Section 4.2. The main findings are summarized in Section 4.3.
4.1 Experimental Methods

Ribonuclease A from bovine pancreas (R5500) was purchased from Sigma-Aldrich and was used without further purification. DSC experiments were first conducted to determine the reversibility of RNase A folding and unfolding in various solvent concentrations and mixtures of EAN, sucrose and H₂O. The DSC quench-and-refold approach was then applied to RNase A. DSC thermal scans were conducted as described in Chapter 3.1.2.

IR was utilized to determine first whether RNase A is still in its native state in various solvent compositions of EAN, sucrose and D₂O and to capture protein secondary structural intermediates during folding. IR quench-and-refold experimental procedures and qualitative and quantitative analysis for RNase A closely followed the protocol detailed in Chapter 3.1.4 and 3.2.3.

4.2 Results and Discussion

4.2.1 RNase A in EAN, Sucrose and H₂O/D₂O

DSC and IR spectral studies were performed on RNase A under the same experimental conditions and procedures described in Chapter 3. Experimental samples consisted of 20 wt% (DSC) or 2 wt% (IR) RNase A and 80 wt% (DSC) or 98 wt% (IR) of a solvent mixture consisting of 27 wt% water (DSC:H₂O, IR:D₂O) 40 wt% EAN and 33 wt% sucrose. Figure 4.1 shows that RNase A is fully reversible upon thermal denaturation as the denaturation peaks are the same over multiple successive scans on the same sample. This is seen in Chapter 3 with lysozyme in the same mixtures. RNase A in
water has a denaturation temperature of 64 °C [Liu and Sturtevant, 1996], and in our particular EAN, sucrose and water matrix, the $T_d$ is $55.2 \pm 1.0$ °C. A similar magnitude of the decrease in $T_d$ due to the solvent is also observed in Chapter 3 when using lysozyme. The enthalpy of unfolding is determined to be $27.8 \pm 1.3$ J/g RNase A, which is a slight decrease from the energy change for RNase A in water at ~ $30$ J/g RNase A [Liu and Sturtevant, 1996]. Again, similar decrease in enthalpy of unfolding is also observed with lysozyme.

![DSC scans of 20 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose. Sample heated to 75 °C and scanned at 20 °C/min. Reversibility is shown.](image)

Figure 4.1: DSC scans of 20 wt% RNase A in 27 wt% H2O, 40 wt% EAN and 33 wt% sucrose. Sample heated to 75 °C and scanned at 20 °C/min. Reversibility is shown.

Figure 4.2 shows DSC upscans of folded and unfolded RNase A in EAN, sucrose and water after LN2 quench. Another proof of full protein reversibility is the fact that the unfolding peaks of the folded quench and unfolded quench scans can be seen to superimpose. As seen in previous DSC lysozyme studies in Chapter 3, there exists a
sudden energy deviation starting at 0 °C and continuing until the denaturation temperature, 55.2 ± 0.7 °C, is observed. This confirms that this observation in the DSC is not only true for lysozyme. The area of the exothermic refolding energy, 27.8 ± 1.1 J/g RNase A, is also the same as that of the endothermic unfolding peak, 27.5 ± 0.9 J/g RNase A, again implying full reversibility of RNase A. As observed by Angell and Wang, the glass transition temperature (T_g) of quenched samples with unfolded protein is up to approximately 5 °C higher than for quenched samples with folded protein as seen by the shift to the right, higher temperatures, of the solid line in the DSC scan of Figure 4.2 [Angell and Wang, 2003]. This shift in T_g between quenched folded and unfolded protein is also observed for lysozyme in Chapter 3.

Figure 4.2: DSC solvent background subtracted heat scans for 20 wt% RNase A in 40 wt% EAN, 27 wt% H_2O and 33 wt% sucrose; LN_2 quench of unfolded/ thermally denatured protein (solid line) and LN_2 quench of folded protein (dotted line). Heat scans were conducted at 20 °C/min.
The secondary structure components of RNase A in EAN, sucrose and water at 25℃ prior to any heat treatment are determined by IR, and the analysis is presented in Figure 4.3. From second derivative analysis, as described in Chapter 3, peak positions that describe the secondary structure are identified at 1690, 1679, 1664, 1652, 1641, 1632, 1620 and 1610 cm⁻¹.

![Figure 4.3: Second derivative (top) and IR spectrum (bottom) of native state RNase A: 2 wt% RNase A in 40 wt% EAN, 27 wt% D₂O and 33 wt% sucrose. Solid line is the IR spectrum and dotted lines represent the fitted Gaussian peaks.](image)
For RNase A, 1652 cm\(^{-1}\) is the peak associated with \(\alpha\)-helices, while 1641 cm\(^{-1}\) corresponds to random coils. RNase A contains predominantly \(\beta\)-sheets, which is evident by the large peak at 1632 cm\(^{-1}\). The peaks at 1620 and 1610 cm\(^{-1}\) also measure \(\beta\)-sheet content. Turn structures are determined at 1690, 1679 and 1664 cm\(^{-1}\) [Kong and Yu, 2007].

A quantitative analysis using Gaussian curve fitting of the RNase A IR spectrum reveals that RNase A is indeed in its native state, as secondary structural composition is comparable to that of RNase A in neutral buffer found in IR literature [Levitt and Greer, 1977; Dong et al., 1990]. Secondary structural content percentages from this work and from literature are listed in Table 4.1.

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>X-ray(^1)</th>
<th>FTIR(^2)</th>
<th>FTIR(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-Helix</td>
<td>23</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>(\beta)-Sheet</td>
<td>46</td>
<td>40</td>
<td>39</td>
</tr>
<tr>
<td>turn</td>
<td>21</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>random</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

\(^1\) Levitt and Greer, 1977  
\(^2\) Dong et al., 1990  
\(^3\) This Work

Though the values are closely correlated between the two IR experiments, the X-ray values are a little different, which could arise from the analysis of the IR data, as baseline location is very important. IR quantitative analysis is seen to be very useful for many different proteins, such as lysozyme, trypsin and myoglobin [Dong et al., 1990].
IR spectra were collected for the as-supplied RNase A at 25 °C, after LN$_2$ quenching from 25 °C, and after slow cooling at 20 °C/min to -135 °C from 25 °C. As one can see in Figure 4.4, the spectra superimpose onto each other, confirming that RNase A does not cold-unfold in our solvent matrix and that any subsequent changes observed in IR spectra of quench-and-refold studies corresponds to structural changes during folding.

![IR spectra graph](image)

Figure 4.4: IR spectra of 2 wt% RNase A in 40 wt% EAN, 27 wt% D$_2$O and 33 wt% sucrose before any quench, after fast quenching with liquid nitrogen and slow quenching at 20 °C/min.

The changes in IR spectra upon LN$_2$ quench-and-refold of RNase A are presented in Figure 4.5. RNase A samples were placed in an oven at 75 °C for 10 minutes to allow for full hydrogen-deuterium exchange. A lower oven temperature was used to unfold the RNase A relative to the temperature used in Chapter 3 for lysozyme. This is because $T_d$
of RNase A in this particular solution mixture is approximately 15 °C lower than that of lysozyme in the same solution matrix. After quenching, the sample was heated at 2 °C/min with 3 minute annealing intervals every 5 °C. IR spectra was collected at each anneal temperature.

![Image](image.png)

Figure 4.5: Successive quench-and-refold IR spectra of 2 wt% RNase A in 40 wt% EAN, 27 wt% D₂O and 33 wt% sucrose. Sample was first heated in the oven at 75°C for 10 minutes and then immediately quenched with LN₂. The sample was then heated from -135 °C to 25 °C at 2 °C/min with 2 minutes annealing intervals every 10 °C for spectra collection.

Note that the spectra taken before any thermal treatment (25 °C initial) and after quench-and-refold (25 °C end) are similar, thereby showing that the protein structurally returns to its native state. The ability to quench is also revealed by the shape similarity of the IR spectra for thermally denatured RNase A at 75 °C heat and -135 °C quench.
Visually, the spectrum of unfolded RNase A shows a shift in maximum peak position from $\sim1630\ \text{cm}^{-1}$ to $\sim1640\ \text{cm}^{-1}$, revealing that after thermal denaturation, $\beta$-sheet is decreased while random coils are increased and become the dominant structure. The spectra also reveals that the unfolded protein has higher turn content, while $\alpha$-helix content decreases. These changes in secondary structural characteristics between the folded and unfolded state of RNase A by thermal denaturation have also been observed in the literature [Torrent et al., 2001; Yan et al., 2006].

A qualitative measure of changes in secondary structure from IR data is to plot the temperature dependence of absorbance intensities at chosen frequencies. These chosen frequencies are associated with particular protein secondary structures. For RNase A, and absorbance intensity at $1632\ \text{cm}^{-1}$ signifies $\beta$-sheet structures and absorbance intensity at $1652\ \text{cm}^{-1}$ is associated with $\alpha$-helical structures.

To conduct the quantitative analysis, the same protocol was implemented as described in Chapter 3.2.3. For RNase A, the 4 distinct and strong peaks chosen from Figure 4.3 as constraints for peak fitting are summarized in Table 4.2. Again, under the proposed quantitative analysis protocol, suitable Gaussian curve fits were only possible for IR spectra obtained at -75 °C and above, where the structure is closer to the native state.

<table>
<thead>
<tr>
<th>Secondary Structure</th>
<th>Peak Location</th>
<th>FWHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>turns</td>
<td>1664 cm$^{-1}$</td>
<td>17.6 cm$^{-1}$</td>
</tr>
<tr>
<td>$\alpha$-helix</td>
<td>1652 cm$^{-1}$</td>
<td>15.2 cm$^{-1}$</td>
</tr>
<tr>
<td>random coil</td>
<td>1641 cm$^{-1}$</td>
<td>16.4 cm$^{-1}$</td>
</tr>
<tr>
<td>$\beta$-sheet</td>
<td>1632 cm$^{-1}$</td>
<td>21.1 cm$^{-1}$</td>
</tr>
</tbody>
</table>
Figure 4.6 plots the qualitative (absorbance level) and quantitative (Gaussian curve fitting) changes for α-helix (1652 cm\(^{-1}\)) and β-sheet (1632 cm\(^{-1}\)) signals, observed during folding at low temperatures, which again reveal the continuous evolution of structure.

Figure 4.6: Qualitative [A] and quantitative [B] IR analyses for α-helical and β-sheet secondary structural changes of 2 wt% RNase A in 40 wt% EAN, 27 wt% D\(_2\)O and 33 wt% sucrose for initial states before treatment (green triangle and blue circle) and upon slow reheating after initial heating to 75 °C and LN\(_2\) quenching (blue diamonds and pink squares). Error bars for 4 scans are within the size of each data point.
By comparing the α-helix and β-sheet formation trend, different rates and temperatures where secondary structure changes during protein folding are seen. Intermediates have also been observed during folding in both chemical and thermal denaturation in research presented in previous literature [Schmid and Blaschek, 1981; Adler and Scheraga, 1988; Udgoankar and Baldwin, 1990; Houry and Scheraga, 1996].

Our technique reveals that many intermediates exist at low temperatures that are not necessarily native-like, especially at very low temperatures. From the qualitative trend, α-helical structure starts forming at approximately -115 ºC, and remains approximately constant between -75 ºC and -15 ºC. When heated towards 25 ºC, RNase A refolds back to the native state and α-helical formation is completed. β-sheets, on the other hand, start forming at a higher temperature, with a significant change in intensity between -75 ºC and -65 ºC. This change is more sudden than the initial change found for α-helix and could possibly be relevant to an intermediate observed in the literature, where β-sheets are formed cooperatively [Udgaonker and Baldwin, 1990]. β-sheets continued forming almost linearly until -15 ºC where formation then slows until the native state is reached. From the comparison of the two qualitative trends, α-helical structures form earlier, at lower temperatures, than β-sheets. This is also observed with lysozyme in Chapter 3 and is consistent with other reports found in literature [Schmid and Blascheck, 1981; Adler and Scheraga, 1988; Gruebele 1999].

Comparison of quantitative and qualitative analyses reveal overall consistent trends, however there are also some differences, which require further experiments with other structural probes in order to reach a satisfactory interpretation. The similarities seen
are that α-helical structure does not change between -75 °C and -15 °C, whereas β-sheet structure continues to evolve after an initial large change at temperatures lower than -65 °C. On the other hand, the sharpness of these evolutionary changes differ; for example, the evolutionary change of β-sheet structure between -50 °C and 0 °C seems to exhibit more linear behavior in the qualitative trend than in the quantitative trend.

Upon comparison of the IR (Figure 4.6) and DSC (Figure 4.2) data for RNase A, the temperatures corresponding to the onset of structural and energetic changes differ, as was found to be the case for lysozyme (Chapter 3). Enthalpy changes consistently start at 0 °C, while secondary structural changes occur at a temperature as low as -115 °C. For both RNase A and lysozyme, folding to the native state is completed by 25 °C. This suggests that the initial changes seen in the IR are plausibly related to energy-neutral structural changes.

4.2.2 Effect of EAN Concentration

In this section, results are presented on the effect of EAN concentration on the folding behavior and overall reversibility of RNase A quench-and-refold. To ensure a homogenous glass matrix during quench-and-refold, the sugar content with respect to water was kept at a similar weight percentage of approximately 60 wt% sucrose and 40 wt% water whenever EAN is removed from the solution. DSC solution samples consisted of 20 wt% RNase A, and 80 wt% of mixed solvent with H2O, while IR solution samples consisted of 2 wt% RNase A and 98 wt% of solvent with D2O. In this section, the solvent matrixes consisting of EAN, water and sucrose are referred to by their corresponding mole percentage of EAN.
These shortened names in bold print (t.ex. **20 mol EAN** refers to a solvent composition with 20 mole percent of EAN) and their respective solvent weight compositions are shown below:

**20 mol EAN**: 40 wt% EAN, 27 wt% H₂O or D₂O, and 33 wt% sucrose  
**10 mol EAN**: 22 wt% EAN, 32 wt% H₂O or D₂O, and 46 wt% sucrose  
**no EAN**: 60 wt% sucrose and 40 wt% H₂O or D₂O  

The effect of EAN concentration on the folding behavior and overall reversibility of RNase A quench-and-refold is possible due to the ability of RNase A to be fully reversible even in the absence of EAN when subjected to thermal denaturation. This is shown in Figure 4.7 where DSC scans of fully reversible RNase A in 10 mole EAN and no EAN solution matrices are presented.

![Figure 4.7](image-url)

**Figure 4.7**: DSC heat scans for 20 wt% RNase A in 10 mol EAN and no EAN matrix compositions. Scans conducted at 20 ºC/min and heated to 20 ºC above T_d: first scan (solid line) second scan (dotted line).
Reversibility of RNase A in 20 mol EAN is shown previously in Figure 4.1. Table 4.3 lists the denaturation temperature ($T_d$) of RNase A in the various solutions and its change in enthalpy ($\Delta H_d$) determined by integrating the calorimetric peak at $T_d$. $T_g$, which was determined from quench-and-refold experiments presented later in this section, is also listed for comparison. As shown earlier for lysozyme and shown here for RNase A, increasing sucrose concentration [Liu and Sturtevant, 1996] and decreasing EAN concentration causes an increase in both the $T_d$ and $\Delta H_d$ during protein unfolding.

Table 4.3: Comparison of Thermal Properties Obtained From DSC Scans for RNase A in Varying Solvent Compositions.

<table>
<thead>
<tr>
<th>Solvent Matrix</th>
<th>$T_d$ (°C)</th>
<th>$\Delta H_d$ (J/g RNase A)</th>
<th>$T_g$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 mol EAN</td>
<td>55.2</td>
<td>27.8</td>
<td>-107</td>
</tr>
<tr>
<td>10 mol EAN</td>
<td>64.4</td>
<td>33.6</td>
<td>-94</td>
</tr>
<tr>
<td>no EAN</td>
<td>75.1</td>
<td>40.1</td>
<td>-75</td>
</tr>
</tbody>
</table>

Error determined from 4 different samples:
$^1$error ± 2%
$^2$error ± 5%

DSC heat upscans of LN$_2$ quenched folded and unfolded LN$_2$ quenched RNase A in 10 mol EAN and no EAN solvents in Figure 4.8 show that the exotherm onset is not observed until 0 °C. The exothermic refolding enthalpies, 34.1 J/g RNase A and 39.8 J/g RNase A for 10 mole EAN and no EAN respectively are similar to the endothermic unfolding enthalpies shown in Table 4.3, thus showing that LN$_2$ quenching is able to trap the unfolded protein.
Figure 4.8: DSC solvent background subtracted heat scans for 20 wt% RNase A in 10 mol EAN and no EAN matrix compositions; LN₂ quench of unfolded/ thermally denatured protein (solid line) and LN₂ quench of folded protein (dotted line).

RNase A is considered reversible upon thermal treatment as observed by the superimposed thermal unfolding peaks of quench-folded and quench-unfolded protein. Comparing the exothermic trends, the onset of quench-unfolded changes upon reheating is less sudden with no EAN. However, in the absence of EAN, most of the enthalpy flux
occurs closer to 0 °C, unlike the rather continuous enthalpy flux continuing towards $T_d$ in the presence of EAN. The exothermic deviation pattern starting at 0 °C in the absence of EAN is similar to that seen in the DSC scan of lysozyme in an aprotic ionic liquid (Figure 3.23 C). The different nature of the exotherm with no EAN suggests that EAN plays an important role in the unfolding and refolding process; however the exact nature and details are not understood at present. This is also consistent with observed changes in endotherms, $\Delta H_d$, listed in Table 4.3.

The corresponding native state IR spectra are presented in Figure 4.9. One noticeable trend in the spectra is that removing EAN and increasing sugar and water sharpened (decreased width and increased height of the peak) the IR spectrum peak. This slight change of the peak, however, does not affect the quantity of secondary structure since the frequency of the amide I' band is not shifted and analysis is done with Gaussian peak area percentages. This observation has been seen previously when comparing IR spectra of hydrated and dehydrated protein samples, where hydrated protein spectra have sharper peaks [Pevsner and Diem, 2001]. Quantitative analysis reveals that the secondary structural percentages of the native structure are still consistent with literature (RNase A in D$_2$O only) between the various solvents, as listed in Table 4.4.

Table 4.4: Secondary Structure Composition of RNase A in Various Solvent Matrices at 25 °C.

<table>
<thead>
<tr>
<th></th>
<th>D$_2$O$^1$</th>
<th>20 mol EAN</th>
<th>10 mol EAN</th>
<th>no EAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-Helix</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>$\beta$-Sheet</td>
<td>40</td>
<td>39</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>turn</td>
<td>36</td>
<td>34</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>random</td>
<td>9</td>
<td>11</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

$^1$Dong et al., 1990
Figure 4.9: Native state IR spectra for RNase A in no EAN and in 10 mol EAN taken at 25 °C. Solid lines are the raw spectra and dotted lines represent the fitted Gaussian peaks.

Comparing the qualitative (Figure 4.10) and quantitative (Figure 4.11) trends for each respective quench-and-refold of RNase A in the different solvents, for both α-helical and β-sheet formation, reveals significant spectral changes occurring at the same
Figure 4.10 Qualitative IR analyses for $\alpha$-helix (1652 cm$^{-1}$) and $\beta$-sheet (1632 cm$^{-1}$) secondary structural changes of RNase A in varying compositions of EAN, sucrose and D$_2$O. Error bars for 4 scans are within the size of the data point.
Figure 4.11 Quantitative IR analyses for β-sheet -1632 cm⁻¹ [A] and α-helical -1652 cm⁻¹ [B] secondary structural changes for RNase A in various compositions of EAN, sucrose and D₂O. Error bars for 4 scans are within the size of the data point.
temperatures. α-helical formation starts at lower temperatures compared to β-sheet, and exhibits a two-step trend, whereas, β-sheet formation exhibits just one large continuous change.

Both qualitative and quantitative trends reveal that changing the solvent does in fact have an effect on the mechanism for particular secondary structural changes. This is shown by the differences in observed absorbance intensity and secondary structural percentages for RNase A in the different solvent matrices at the same temperature during refolding, particularly higher temperatures, and indicates a shift in the evolution of structure. Table 4.5 lists secondary structural content determined by the four chosen and constrained peaks introduced in the quantitative analysis protocol for RNase A refolding in various solvent mixtures. In sample solutions with increasing amounts of EAN, the formation of secondary structures occurs at higher temperatures.

Previous research has shown that if a protein is in an environment that stabilizes the native protein globule, the solution stabilizes other globular structures and intermediates as well [Finkelstein and Galzskitskaya, 2004]. This stabilization of the native state is argued to increase the folding rate, since the folding nucleus is stabilized and favored whereas the misfolded structures are unstable. The reason for the instability of the misfolded state is the energy gap between the misfolded state and the native fold [Scalley and Baker, 1997; Finkelstein and Galzskitskaya, 2004]. Therefore, one can presume that increasing the energy gap between the native and the nonnative state, which further stabilizes the native state, would lead to faster secondary structure formation upon refolding. This explanation is consistent with results obtained by DSC and IR for RNase A in various solutions.
Table 4.5: Summary of Quantitative Secondary Structural Content for RNase A in Various Solvent Compositions During Quench-and-Refold.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Secondary Structure Peak</th>
<th>20 mol EAN</th>
<th>10 mol EAN</th>
<th>no EAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 °C initial</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>15.7</td>
<td>15.8</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>29.9</td>
<td>29.3</td>
<td>31.9</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>14.2</td>
<td>13.8</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>11.1</td>
<td>10.3</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>29.1</td>
<td>30.8</td>
<td>29.5</td>
</tr>
<tr>
<td>-75 °C</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>7.1</td>
<td>7.1</td>
<td>6.84</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>11.9</td>
<td>13.2</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>22.1</td>
<td>21.4</td>
<td>21.8</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>17.7</td>
<td>16.5</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>41.2</td>
<td>41.8</td>
<td>42.16</td>
</tr>
<tr>
<td>-55 °C</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>7.1</td>
<td>7.9</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>20.3</td>
<td>20.51</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>21.9</td>
<td>20.7</td>
<td>20.7</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>14.6</td>
<td>13.4</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>36.1</td>
<td>37.49</td>
<td>37.3</td>
</tr>
<tr>
<td>-35 °C</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>7.8</td>
<td>9.5</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>20.6</td>
<td>20.3</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>21.3</td>
<td>19.3</td>
<td>19.1</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>14.5</td>
<td>13.4</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>35.8</td>
<td>37.5</td>
<td>37.5</td>
</tr>
<tr>
<td>-15 °C</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>10.9</td>
<td>12.6</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>23.9</td>
<td>26.4</td>
<td>30.3</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>18.5</td>
<td>16.6</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>13.32</td>
<td>11.3</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>33.38</td>
<td>33.1</td>
<td>30.7</td>
</tr>
<tr>
<td>5 °C</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>12.7</td>
<td>14.2</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>27.8</td>
<td>27.8</td>
<td>30.9</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>16.9</td>
<td>15.2</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>11.8</td>
<td>10.8</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>30.8</td>
<td>32.0</td>
<td>30.6</td>
</tr>
<tr>
<td>25 °C end</td>
<td>1652 cm⁻¹ – α-helix</td>
<td>15.6</td>
<td>16.1</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>1632 cm⁻¹ – β-sheet</td>
<td>29.3</td>
<td>28.9</td>
<td>31.3</td>
</tr>
<tr>
<td></td>
<td>1668 cm⁻¹ – turns</td>
<td>14.7</td>
<td>13.1</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>1641 cm⁻¹ – random</td>
<td>10.9</td>
<td>10.7</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>unassigned</td>
<td>29.5</td>
<td>31.2</td>
<td>29.2</td>
</tr>
</tbody>
</table>
Figure 4.12 shows DSC heating and cooling scans of RNase A in various amounts of EAN, water and sucrose compositions. As presented previously in Table 4.3 and also shown in Figure 4.12, $T_d$ decreases with increasing EAN concentration. Table 4.3 reveals that the $\Delta H_d$ decreases with increasing EAN concentration as well. These DSC results indicate EAN as a destabilizer of the native state, though not strong enough to unfold the protein. Figure 4.12 also reveals that the differences between the unfolding and refolding temperature of RNase A are greater at higher EAN amounts; with the difference between $T_d$ and the refolding temperature from the cooling scan, $T_F$, equal to 7 °C, 15 °C and 25 °C in no EAN, 10 mol EAN and 20 mol EAN solution mixtures, respectively. The smaller difference between $T_d$ and $T_F$ for RNase A in the no EAN solution is an indication of RNase A’s ability to fold faster or earlier, equivalent to higher temperatures, when the native state is more stabilized.

Figure 4.12: DSC heating (solid lines, ramp up left to right) and cooling (dotted lines, ram down right to left) scans of 20 wt % RNase A in solvent compositions of 20 mol EAN, 10 mol EAN and no EAN. Samples were run at 20 °C/min.
The change in evolution of secondary structural changes observed in IR for RNase A in solutions of no EAN, 10 mol EAN and 20 mole EAN also reveal faster structural formations for RNase A, with higher native state stability in EAN-free solution. Both absorbance intensities and structural percent changes occur earlier, which in this experimental set-up corresponds to lower temperatures, than changes in 10 mol EAN and 20 mol EAN solvent matrices.

The difference in the evolution of secondary structural changes observed from the IR spectra could possibly be linked to the difference in shape of the exothermic refolding area found in the DSC quench-and-refold scans (Figure 4.8). Samples in no EAN solution matrix have a higher $T_g$ resulting in onset changes occurring at higher temperatures, which could be the reason for the slower, less sudden start at 0 °C in the DSC. On the other hand, the changes in secondary structure occur at lower temperatures in no EAN, which could be associated with the deeper energy deviation closer to 0 °C.

Again, upon comparison of DSC and IR data for RNase A, even in different solvent mixtures, the difference in the onset of structural and energetic changes is still observable. Figure 4.13 presents DSC and IR spectra of RNase A unfolding in various mixtures of EAN, D$_2$O and sucrose showing cooperative unfolding at their respective protein denaturation temperatures. As observed for lysozyme, exchanging H$_2$O to D$_2$O increases the $T_d$ by approximately 2.7 °C. Since RNase A consists of mostly $\beta$-sheets, the changes in absorbance level at 1632 cm$^{-1}$ are utilized to determine the thermal unfolding of RNase A. DSC and IR trends are very similar to those in Figure 3.21 for lysozyme in the 20 mol EAN solvent matrix, where the same transition temperatures between DSC and IR are observed at scan rates of 20 °C/min.
Figure 4.13: Changes in absorbance level (diamonds) at 1633 cm\(^{-1}\), β-sheets, and DSC scans (solid lines) for thermal unfolding of RNase A in various solvent matrices: 20 mol EAN, 10 mol EAN, and no EAN. Both IR and DSC heating scans were conducted at 20 °C/min and all the solution matrices consisted of D\(_2\)O.

Though there exist studies of how excipients and solvent composition can affect the unfolding properties and aggregation of proteins [Arakawa and Timasheff, 1982; Carpenter et al., 1999; Byrne et al., 2009; Chi et al., 2003; Lee and Timasheff., 1981], the quench-and-refold technique can provide new insights into solvent effects on protein refolding by observing changes in secondary structure formation using IR.
4.3 Conclusions

Extending the studies in Chapter 3, RNase A was subject to LN_2 quench-and-refold experiments in various solvent compositions of EAN, sucrose and water and studied by DSC and IR. The results with RNase A confirms this new quench-and-refold approach with IR as a useful tool to capture and study the secondary structure of protein intermediates during folding after thermal denaturation. Comparison of DSC and IR observations on quenched unfolded RNase A shows that the onset temperature for the folding energy relaxation is consistently at 0 °C, whereas the onset of structural changes is consistently at much lower temperatures. As shown in Chapter 3 for lysozyme, the formation of secondary structural elements for RNase A at early refolding stages, equivalent to very low temperatures in our experiments, appear likely to be energy neutral. The unfolding of RNase A is fully reversible in solutions in the presence and absence of EAN. Though the qualitative and quantitative trends of folding RNase A are similar between varying amounts of EAN in solution, the evolution of secondary structural changes varies with temperature. In increasing amounts of EAN, secondary structure formations for RNase A occur at comparatively higher temperatures, thereby implying that RNase A has slower folding rates in the presence of EAN. Since changes in the solvent composition affect the nature of the evolution of secondary structure, this could possibly imply a change in protein folding mechanism.
4.4 References


Byrne N, Belieres JN, Angell CA. 2009. The ‘refoldability’ of selected proteins in ionic


Liu Y, Sturtevant JM. 1996. The observed change in heat capacity accompanying the thermal unfolding of protein depends on the composition of the solution and the method employed to change the temperature of unfolding. Biochemistry. 35:3059-3062.


Chapter 5

Concluding Remarks and Suggestion for Future Work

This dissertation is divided into two parts. The first part focuses on an observational study on protein reversibility when subject to thermal denaturation in the absence and presence of sugar (trehalose), determined by a number of experimental techniques. The second part is the introduction of a new experimental technique to study protein folding mechanisms by capturing intermediate structures during protein refolding with infrared spectroscopy using a quench-and-refold approach and includes the development of the protocol. By comparison of differential scanning calorimetry (DSC) and infrared spectroscopy (IR) results, insights into energy effects and structure formation relations are obtained.

In Chapter 2, a systematic set of experiments using DSC, IR, ultraviolet spectroscopy (UV), and enzymatic assays was conducted to study the effect of trehalose on lysozyme aggregation. This was initially completed to determine a suitable solvent for use in quench-and-refold experiments. DSC detects irreversible protein aggregation by the decrease in denaturation enthalpy change observed between an initial heating scan
and a second heating scan of the same sample. Using IR, protein aggregation was detected by the appearance of two spectral peaks associated with intermolecular β-sheet upon heating. When cooling the sample, the two spectral peaks did not disappear. UV spectroscopy was used to measure both protein concentration and the number of active units in solution. Thermally treated samples were centrifuged and the supernatant was taken for further analysis. Protein irreversibility was determined by the decrease in protein concentration and number of active units remaining in solution after thermal treatment.

IR spectra of a high concentration of lysozyme (200 mg/ml) in a high concentration of trehalose (80 wt%) solution reveal that the high concentrations of trehalose promoted protein aggregation. Therefore, the initially proposed idea of increasing viscosity to prevent aggregation [Munishkina et al., 2004; Tavarti et al., 2007] is not suitable with high concentrations of trehalose. Instead, molecular crowding is one possible reason for the promotion of protein aggregation. Adding lower amounts of trehalose to the solvent in fact prevents aggregation, which is more in line with experiments and observations seen in previous literature [Ueda et al., 2001]. Studies conducted with both high and low protein concentrations in varying trehalose solutions ranging from 0 to 40 wt% reveal that the more trehalose is added, the less aggregation and activity loss is observed. Aggregation in the absence and presence of trehalose is also found to be dependent on initial protein concentration. For both high and low protein concentrations, increasing protein concentration increases the rate of aggregation and decreases residual enzyme activity recovery. Results also reveal that both rates of aggregation and inactivation of lysozyme in trehalose due to thermal treatment follow
first-order kinetics. Possible explanations for the ability of trehalose to prevent aggregation are (1) that the preferential exclusion of trehalose from the protein increases the free energy difference between native and denatured states of a protein, thus favoring the native state (2) and/or that trehalose helps to maintain and induce secondary structure, therefore providing a lower number of partially folded states to aggregate.

Chapter 2 focuses on only one protein, lysozyme, and one excipient, trehalose. It would be of interest to extend this experimental study to other proteins to study whether the different aggregation mechanisms in high and low trehalose concentration are also observed. Previous studies have shown that protein stability when subjected to thermal treatment can vary depending on the type of protein, both in solution and solid-state [Lee et al., 2006]. It would be useful to systematically investigate a set of proteins that cover a wide range of molecular weights, amino acid sequence, and structural motifs. Such a study could provide further understanding into protein structure and function and how excipients and proteins interact.

Another extension of this study could be to change the excipient to other carbohydrates. Protein interaction with an excipient can greatly affect stability. For example, urea [Rudolph and Lilie, 1996] interacts with the protein whereas carbohydrates are preferentially excluded from the surface [Lee and Timasheff, 1981; Arakawa et al., 1990]. Dextran would be an interesting starting choice, as it has been found that dextran is unable to retain protein activity and structure after freeze-drying protein samples, while other carbohydrates, such as trehalose have the ability to do so. [Lee et al, 2006]. Another recent experimental study by Allison et al., shows that solid formulations containing a mixture of two different carbohydrates (e.g. sucrose and dextran) have better protein
stability compared to having only one carbohydrate [Allison et al., 2000]. A protein aggregation study with mixtures of excipients, which would better simulate protein stability in cells, could be of interest. Other excipients that may be of interest and known as protein stabilizers in studies with low concentration are arginine [Arakawa and Tsumoto, 2003], arginine ethylester [Shiraki et al., 2004 ] or sorbitol [Peterson et al., 2004]. There are currently few studies that explore high protein and high excipient concentrations in solution.

Another aspect to consider is how thermal treatments affect protein aggregation, such as heating a protein above its denaturation temperature. It may be of interest to compare isothermal and thermal unfolding of protein in different excipients. Thermodynamic data can be obtained using a DSC that varies with the temperature and an isothermal titration calorimetry (ITC) using different compositions. At some point there should be a “crossing range” between ITC and DSC data. This finding in combination structural information from IR or circular dichroism studies could provide further insight into thermal effects on protein stability and protein-excipient relations and binding.

In Chapter 3, we describe the development and usefulness of a sensitive method to track the protein intermediate structures upon folding from the fully unfolded state back to the native state. A quench-and-refold approach utilizing the glassy state is introduced using infrared spectroscopy to analyze the structural state of the protein. The main advantage of this technique is the opportunity to observe subtle structural changes as they evolve over a continuum in time, due to the slow kinetics found in the viscous liquid state near $T_g$. Careful selection of protein and solvent compositions were necessary
for the development of this new experimental technique. Protein should fold slowly enough to quench using LN₂, and the solvent should not crystallize and denature or aggregate the protein. Lysozyme was used as our model protein as it is considered a slow folding protein [Kato et al., 1982; Dobson et al., 1994] and is well characterized in structure [Dong et al., 1990; Levitt and Greer, 1977; Manavalan and Johnson, 1987]. As determined in Chapter 2, lysozyme in water and trehalose aggregated upon heating above the unfolding transition denaturation temperature, Tₐ. Therefore another suitable solvent was needed. Ethylammonium nitrate (EAN) was found to prevent aggregation of lysozyme for protein concentrations up to 200 mg/ml in solvent mixtures containing more than 10 mol% EAN-water. Sucrose was added to the solvent to permit vitrification and to prevent ice crystallization during quench-and-refold. An optimal solvent composition of EAN, sucrose and H₂O/D₂O was determined. In such a medium, lysozyme is still a functional and active enzyme. Both IR and DSC results provide evidence of successful quenching and trapping of unfolded lysozyme to the glassy state.

An interesting observation is seen when comparing quench-and-refold IR and DSC results of lysozyme in EAN, sucrose and H₂O/D₂O, where the temperature of detection for protein refolding by the two methods differs by almost 100 °C. Quench-and-refold DSC scans reveal a refolding exotherm starting at 0 °C, which continues up to the endothermic protein unfolding peak centered at ~68 °C for lysozyme in 40 wt% EAN, 33 wt% sucrose and 27 wt% H₂O. Qualitative analysis of quench-and-refold IR spectra indicate that structural changes (particularly that of α-helix formation) are observed at temperatures as low as -115 °C. Further studies were completed on the system to understand this temperature difference between IR and DSC. Protein concentration does
not affect the consistently observed onset of energy release at 0 °C. DSC quench-and-refold of the solvent itself does not exhibit energy deviation at 0 °C. A systematic study of the solvent itself in DSC and IR reveal that background subtractions for data analysis are sufficient for isolating the major changes in protein structure and energy relaxation. Also, solvent viscosity measured at different temperatures above and below 0 °C does not show an unusual change right at 0 °C to warrant the start of energy release. DSC quench-and-refold of lysozyme in different mixtures of ionic liquids and sugars still result in an enthalpy release starting at 0 °C. However, noticeable differences in the shape of the exotherm are observed, and could possibly be attributed to the interaction between solvent and protein. Through isothermal titration calorimetry experiments (ITC), EAN and lysozyme are found to weakly interact and bind together. Because these ITC experiments were conducted only at room temperature, it would be of great interest to determine if there are thermal contribution from the differential binding of EAN to unfolded (higher temperatures) and folded (room temperature) protein. A systematic set of experiments using the ITC with different solvents known to interact differently with protein would also be of interest. This could further elucidate solvent-protein interactions. The driving force behind the consistent enthalpy relaxation starting at 0 °C in DSC heating scans of quenched unfolded protein still remains an open question.

Thermal annealing experiments of lysozyme in EAN, sucrose and H_2O/D_2O at varying low temperatures reveal that intermediates held at temperatures below -15 °C are in equilibrium within the experimental time frame. Anneals conducted at higher temperatures result in protein refolding. Anneals performed at temperatures below -25 °C reveal that any structural changes observed in the IR are in fact energy neutral, due to
the lack of energy loss in the refolding peak, similarly indicating that structural changes occurring above -25 °C have an associated energy component. The combined use of IR and DSC reveals the ability of IR to capture early structural changes that can be considered energy-neutral using the proposed quench-and-refold approach. Such energy-neutral rearrangements have not been observed previously in IR.

In Chapter 4, a different slow-folding protein, RNase A, was studied using the new experimental technique. RNase A results, in the same solvent composition of EAN, sucrose and H₂O/D₂O as Chapter 3 experiments with lysozyme, further establishes the quench-and-refold approach with IR as a useful tool to capture and study the secondary structure of protein intermediates during folding after thermal denaturation. Comparison of DSC and IR observation on quenched unfolded RNase A shows that the onset temperature for folding enthalpic relaxation is still 0 °C, whereas the onset of structural changes revealed by IR is much lower. As shown in Chapter 3 for lysozyme, the formation of secondary structural elements for RNase A at early refolding stages, equivalent to very low temperatures in our experiments, appear likely to be energy-neutral. However, to further confirm such findings, kinetic experiments should also be conducted for RNase A. It would also be of interest to run activity assays of RNase A in EAN, sucrose and water solution mixtures to determine if this particular protein is still active.

Thermal denaturation of RNase A is found to be fully reversible in solutions in the presence and absence of EAN. Extending the studies in Chapter 3, RNase A was subject to DSC and IR quench-and-refold experiments in various solvent compositions, and conducted to study the effect of EAN on the folding mechanism of specific
secondary structures. Though the qualitative and quantitative structural trends of folding RNase A are similar between varying amounts of EAN in solution, the evolution of secondary structural changes varies with temperature. In increasing amounts of EAN in solution, secondary structure formation from quench-and-refold data occur at higher temperatures, and thus considered to have a slower folding rate. Since changes in the solvent composition affect the nature of the evolution of secondary structure, this could possibly imply a change in protein folding mechanism.

This quench-and-refold approach provides the opportunity to observe protein folding mechanisms using IR, by following subtle structural changes as they evolve over a continuum in time from a denatured protein state back to its native state. Utilizing both the quench-and-refold approach with DSC and IR reveals thermal and structural formation relations. This new technique can be applied to study protein folding mechanisms of other proteins and in different solvent systems, though careful consideration of the LN$_2$ immersion cooling rate and reversibility of the protein is necessary. One method to increase the cooling rate of the sample using LN$_2$ immersion for IR studies is to decrease the thickness of the AgCl window. However, protein folding studies on faster folding proteins can also be achieved by using a faster quenching method than LN$_2$ immersion. One possible method is to use an electrospray that can quench and trap unfolded protein with cooling rates four orders of magnitude faster than with LN$_2$ [Wang et al., 2007]. Since quenching can be applied at any stage in unfolding, perhaps a study where quench-and-refold of protein at different stages of its unfolding (basically quench from different temperatures across its denaturation temperature) could provide further insight into structural-energy relations.
With increasing attention and interest in ionic liquids for protein applications with their ability to prevent protein aggregation [Summers and Flowers II, 2000; Byrne and Angell, 2008], a set of experiments using the quench-and-refold approach with IR for lysozyme in different ionic liquids would be of interest. Angell and coworkers have recently demonstrated the generality of protein reversibility in protic ionic liquids. Characterizing different ionic liquids’ proton activity using NMR proton chemical shift measurement, an optimal range of proton activity for ionic liquid mixtures that could prevent protein aggregation was determined [Byrne et al., 2007; Byrne and Angell, 2008]. Of particular interest is that DSC scans of lysozyme in ionic liquids and mixtures at the edges of the “high refoldability zone” were characterized by a 3-state unfolding (two distinct denaturation peaks) instead of the usual 2-state unfolding (one denaturation). Initial interest would be to determine the secondary structure of lysozyme when it exhibits 3-state unfolding and to check if there is a different intermediate structure between the two unfolding peaks. Quench-and-refold studies using both DSC and IR would be of interest, and could help provide information on whether the refolding mechanism (2-state or 3-state) is affected or dependent on protein unfolding characteristics. This is just one interesting example where the quench-and-refold approach with DSC and IR can be potentially applied and useful.
5.1 References


effects of sugar and polyols on the protein structure and function: role of