

STABILITY, UNFOLDING, AND AGGREGATION OF THE GAMMA D AND
GAMMA S HUMAN EYE LENS CRYSTALLINS

by

Ishara Amenti Rakem Mills-Henry

B.S./M.S. Biology
Clark Atlanta University, 2000

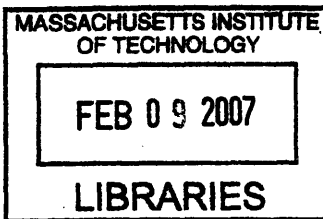
Submitted to the Department of Biology in partial fulfillment
of the requirement for the degree of

DOCTOR OF PHILOSOPHY


at the
Massachusetts Institute of Technology

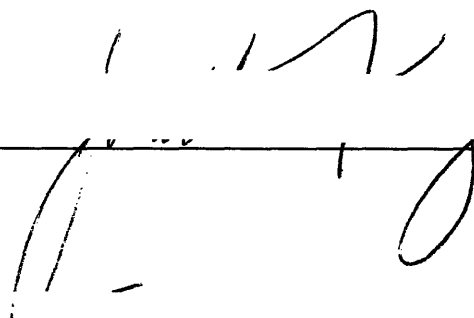
February 2007


© 2007 Ishara Amenti Rakem Mills-Henry
All rights reserved



The author hereby grants MIT permission to reproduce and to distribute publicly paper
and electronic copies of this thesis document in whole or in part.

Signature of Author  Department of Biology
February 2007

Certified by  Jonathan King
Thesis Supervisor

Accepted by  Stephen P. Bell
Chairman, Department Committee on Graduate Students

ARCHIVES

STABILITY, UNFOLDING, AND AGGREGATION OF THE GAMMA D AND GAMMA S HUMAN EYE LENS CRYSTALLINS

Submitted to the department of Biology at the Massachusetts Institute of Technology on
January 23, 2007 in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in Biochemistry

ABSTRACT

The transparency of the human eye lens depends on the properties of the α -crystallin and $\beta\gamma$ -crystallin families of proteins, which accumulate to very high concentrations in mature lens fiber cells. The β - and γ -crystallins are thought to be primarily structural proteins while α -crystallin possess an additional chaperone activity. Aggregation of partially unfolded or covalently damaged forms of these proteins results in cataract, which is the leading cause of blindness in the world. The biochemical basis of the very high crystallin stability, and the nature of the misfolded, modified, or aggregated states, are thus of considerable importance in understanding the etiology of loss of lens transparency.

All vertebrate β - and γ -crystallins are homologous and contain two highly symmetrical domains with a hydrophobic interface connected by an interdomain linker. The overall sequences, fold topology, and domain interfaces of vertebrate $\beta\gamma$ -crystallins are highly conserved. The two domain β - and γ -crystallins are believed to have evolved by gene duplication and fusion from an ancestral single domain $\beta\gamma$ -crystallin. This thesis is focused on stability and aggregation properties of human γ D crystallin (γ D_{WT}) and human γ S crystallin (γ S_{WT}), two of the most abundant proteins in the human lens. Terminally differentiated fiber cells in the central nucleus of the lens are enucleated and devoid of organelles. The γ D crystallins synthesized *in utero* must remain stable and soluble throughout life. The γ S protein is more prevalent in the cortical regions of the lens, where protein degradation and synthesis do occur.

Given the importance of long-term solubility for the crystallins, it seems likely that selection for the two-domain form is related to the need for very long-term stability. Comparison of the stabilities of the isolated domains and the intact protein indicated that the domain interface contributes a ΔG_{H_2O} of ~ 4.2 kcal \cdot mol⁻¹ to the stability of the complete γ D_{WT} two-domain protein. The differential stability observed for the γ D isolated domains was not as distinct for the γ S isolated domains. These results support the idea that selection for increased thermal stability was one of the factors leading to the evolution of two domain crystallins.

A distinct hysteresis occurs during equilibrium unfolding and refolding, due to a kinetic barrier in the unfolding pathway. By extrapolating kinetic unfolding results from denaturing GuHCl concentrations to buffer, I show that the $t_{1/2}$ for the initial unfolding step is ~ 19 years. The value extrapolated for the γ S_{WT} is not as long, though still significant. This supports the earlier conclusion that the domain interface is an important source of stability.

Previous studies had shown that upon dilution from denaturant partially folded intermediates of γ D crystallin formed highly ordered fibrous aggregates that were not amyloid in nature. This aggregation reaction of γ D_{WT} polypeptide chains competing with

productive refolding provided a model for cataractogenesis *in vitro*. The structurally homologous $\gamma_{\text{S}_{\text{WT}}}$ crystallin did not exhibit an off-pathway aggregation under the same conditions as $\gamma_{\text{D}_{\text{WT}}}$. This suggested that the pathway of aggregation involved specific amino acids or sequences essential for association and was not a general feature of the γ -crystallins. To investigate this disparity between two structurally similar crystallins, chimera proteins were created in an attempt to narrow down regions of the protein that promoted aggregation in $\gamma_{\text{D}_{\text{WT}}}$ or regions in $\gamma_{\text{S}_{\text{WT}}}$ that inhibited aggregation. The aggregation behavior upon refolding was analyzed for the chimeras and isolated single domains. Partitioning of refolding chains into the aggregation pathway was strongest for the full-length proteins that retained the γ_{D} interface. This result is consistent with a domain swapping mechanism for the off-pathway aggregation of the crystallins. This aggregation reaction may be coupled to the increased stability of the γ_{D} -protein, as a kind of evolutionary cost of the extremely stable and long-lived native state conformation.

Thesis Supervisor: Jonathan King, Professor of Biology

ACKNOWLEDGEMENTS

As I end another phase of my life, I must pause and give thanks to those who have made completing it possible. Completing this degree has been probably the most difficult thing in my life thus far, but the rewards that I have gained along the way will always be cherished. I left this section the last thing that I wrote since I knew it would be the hardest to write. I really do not know how to thank everyone who has assisted me along the way. It reminds me of one of my favorite quotes:

“People will not remember what you said or what you did. But what they will remember is how you made them feel.”

-Maya Angelou

Thanks to all who have touched my life in some way.

First, I would like to thank my advisor, Professor Jonathan King, who is a true mentor giving not only scientific advice but also career and personal advice. I thank him for taking me and accepting me into the lab without prior knowledge of my ability as a scientist. But most of all, I thank him for giving me faith in academia when I had lost it and for introducing me to the different types of opportunities that are available to me. I look forward to more informative discussions with you in the future.

To my committee members, Professors Robert Sauer and Amy Keating. Thank you for your support and helpful suggestions especially since time was of the essence the last two years. Also, along with Professors Thomas Schwartz and Sean Decatur, thank you for reading my thesis and for your helpful suggestions.

Dr. Shannon Flaugh (Thol). I don't even know where to begin to thank you for teaching me not only protein biochemistry, but also for teaching me so much about life. Your genuine personality and selflessness amazes me. I think we hit it off the first time we meet at Logan airport when I picked you up for the prospectus weekend. I don't know how you saw the small 3 X 5 notebook with your name on it, but I can't help to think that it was fate that led us to each other. I will treasure your friendship forever.

Dr. Melissa Kosinski-Collins. Thank you for giving me the advice to contact Jon about joining the lab. It was one of the best decisions that I could have made. From studying with you for the preliminary exams to doing the South Beach diet together, your friendship has been priceless. May we continue to be great colleagues and friends.

Team Crystallin! You are the best! Thank you for proofreading this thesis as well as listening to my several practices of my thesis defense. Jiejun Chen, your commitment to great science is motivating. I have enjoyed our conversations about life as much as our scientific ones. Keep smiling and working hard! Ligia Acosta Sampson, I will never forget the subway in Argentina and our new phrase “packed like the crystallins in the lens”. Keep on doing great things in the lab! Thanks for sharing your purification knowledge and for performing the aggregation kinetic experiments. Yongting Wang,

thank you sharing your knowledge of kinetics and for the assisting with the analytical size exclusion chromatography. You have been a great addition to the lab and your stories about writing your thesis kept me from going crazy. Kate Drahos, your cheerful personality fits perfectly in the nature of the lab, the best of luck to you! To Robin Nance and Cecile Lin, thanks for the technical support and the fun you both brought to the lab. I know you both will be valuable to the future of science. To everyone in the King lab, past and present, thank you for the supportive environment that you have provided during my tenure in the lab. To Dr. Peter Weigele, Cammie Hasse-Pettingell, and Ryan Simkovsky, thank you for the support and helpful advice. Thank you also to Welkin Pope, Professor Jacqueline Piret, Philip Campbell, Professor David Gossard, and Cindy Woolley.

I would like to thank my former advisor, Professor Ilaria Rebay, for teaching me how to think like a scientist. To my former committee members, Professors Terry Orr-Weaver and Frank Solomon, thank you for your continual support in my progress while I was in Ilaria's lab. To Professor Mousumi Mutsuddi for being my best friend and for constantly reminding me there is more to life than working in the lab. David Doroquez for his friendship and never ending positive attitude. To my good friend Dr. Mandy Tam, I will never forget the time we went looking for apartments together and the pet Iguana in one apartment scared you to death. We have been through thick and thin together. Thank you for helping me to get started with this document! To ACME (Academy of Courageous Minority Engineers), Eric, Aisha, Reggie, Lincoln, Robbin, Mark, James, and Shani. Without this organization, I don't know I could have made it through some difficult hurdles at MIT. Thank you for holding me accountable!

Last but never least, to my family, the best family a person could ever have been blessed with. If I named everyone, this thesis would be a few pages longer, but I just wanted to thank everyone for their prayers, support, and love throughout this process. To my husband, Melvin M. Henry, your love truly completes me. I appreciate your patience and complete support that you have given throughout this process. You have truly been my anchor. To my parents, Carol and Norris, you always taught me the value of education and never to give up. When people ask me where I get my tenacity, I say you have never met my parents. Thank you for your unwavering encouragement, advice, and love. I love you so much. I would like to thank my siblings, Abdullah (Abe), Felicia, CL, Tony, and Rich. Especially to my brother, Abe, thank you for the example that you live. The pursuit of your dream inspires me daily. Simone, thank you for always being there for me no matter what time it is and reminding me to take some R&R. To Aunt Ana Ruf (Elenor Tucker), you truly are my second mom, thank you for your support.

This thesis is dedicated to my grandmother, Regina Spence, although she did not have even a high school education, she taught her children and grandchildren the value and importance of education. May your legacy live on forever.

This research was supported by NIH GM17980 awarded to Jonathan King and the David and Lucille Packard Foundation Fellowship and UNCF•Merck Graduate Dissertation Fellowship awarded to Ishara Mills.

BIOGRAPHICAL NOTE
Ishara Amenti Rakem Mills-Henry

Education

- Ph.D. Massachusetts Institute of Technology, Department of Biology,
Expected 2007 Cambridge, MA.
- M.S. Clark Atlanta University, Department of Biological Sciences,
July 2000 Thesis: The over-expression of ER α in osteoblast-like cells, *in vitro*.
- B.S. Clark Atlanta University, Department of Biological Sciences,
July 2000 Major: Biology, Summa Cum Laude, Atlanta, GA.

Research and Professional Experience

- 2004 to 2007 Graduate Research Assistant in the laboratory of Professor
Jonathan King, MIT Department of Biology, Cambridge, MA.
- 2000 to 2004 Graduate Research Assistant in the laboratory of Professor Ilaria
Rebay, MIT Department of Biology, Cambridge, MA.
- 1998 to 2000 Graduate Research Assistant in the laboratory of Professor Kevin
Sean Kimbro, Clark Atlanta University Department of Biological
Sciences, Atlanta, GA.
- Summer 1997 Research Intern in the Central Nervous System/ Cardiovascular
Division, Schering-Plough Research Institute, Kenilworth, NJ.
- 1996 to 1998 Undergraduate Research Assistant in the laboratory of Professor
David Collart, Clark Atlanta University Department of Biological
Sciences, Atlanta, GA.
- Summer 1996 Research Intern in the Agricultural Division, Ceregen, Monsanto
Chemical Company, St. Louis, MO.

Publications

- Flaugh S, Mills I, King J (2006) *Glutamine Deamidation Destabilizes Human γ D-Crystallin and Lowers the Kinetic Barrier to Unfolding*. J. Biol Chem. 2006; 281(41):30782-30793.
- Tootle T, Silver S, Davies E, Newman V, Latek R, Mills I, Selengut J, Parlikar B, Rebay I (2003) *The transcription factor Eyes absent is a protein tyrosine phosphatase*. Nature 2003; 426(6964):299-302.

TABLE OF CONTENTS

Title Page	1
Abstract	2
Acknowledgements.....	4
Biographical Note	6
Table of Contents	7
List of Figures	11
List of Tables	13
Chapter One: Introduction.....	14
A. Vertebrate Eye	15
B. Human Lens Anatomy and Composition	17
C. Lens Development	17
D. Lens Transparency	18
E. Vertebrate Genes Forming the Lens Crystallins	20
F. Conformation and Properties of the Lens Crystallins	22
1. α -crystallin	24
2. β - and γ -crystallins	25
3. Stability of the $\beta\gamma$ -crystallins	29
4. Microbial Crystallins	30
G. The Crystallin Properties and Lens Evolution	32
1. Vertebrate Ubiquitous Crystallins	32
2. Taxon-specific Crystallins	33
H. B-Sheet Protein Folding	35
I. Other Proteins with Greek Key Motifs	37
J. Folding, Unfolding and Aggregation of γD_{WT}	38
K. Basis for Globular Protein Stability	42
L. Cataract as a Protein Deposition Disease	46
1. Mutations in Crystallin Genes Leading to Congenital Cataracts	47
2. Potential Causes of Mature-Onset Cataract	48
3. Role of α -crystallin	49
M. Mechanisms of Protein Aggregation Related to Human Disease	49
1. Amyloidosis	51
2. Loop sheet insertion – The Case of the Serpins	52
3. Domain Swapping – In General	55
4. Light Chain Amyloidosis and Light Chain Deposition Disorder.....	58
N. The Biological Context of this Thesis	58
Chapter Two: Folding and Stability of the Isolated Greek Key Domains of the Long-Lived Human Lens Gamma D Crystallin	60
A. Introduction	61
B. Materials and Methods	66

1.	Preparation of Constructs for Isolated Domains and γ_{SWT}	66
2.	Expression and Purification of Proteins	67
3.	Analytical Size-Exclusion Chromatography	67
4.	Circular Dichroism	68
5.	Thermal Denaturation.....	68
6.	Fluorescence Spectroscopy	69
7.	Equilibrium Unfolding and Refolding	69
8.	Productive Refolding Kinetics	70
C.	Results	70
1.	Protein Purification and Characterization	70
2.	Analytical Size Exclusion Chromatography	73
3.	Circular Dichroism (CD) Spectroscopy	75
4.	Fluorescence Spectroscopy of the Purified Proteins	76
5.	Thermal Denaturation Indicates Differential Domain Stability.....	81
6.	Equilibrium Unfolding and Refolding <i>in vitro</i>	83
7.	γ_{SWT} Equilibrium Unfolding/Refolding Analysis also Demonstrate Domain Stability	85
8.	Productive Refolding of γ_{DWT} , γ_{SWT} and their Individual Domains	90
D.	Discussion	92
1.	Native-Like Folded Conformation of the Isolated Domains.....	95
2.	Stability of the Full-Length γ -Crystallins and their Isolated Domain.....	95
3.	Kinetic Refolding of the Isolated Domains and their Respective Full-Length Proteins.....	96
4.	Biochemical Basis of Stability Differences of the Isolated Domains.....	98
5.	Comparisons with Other Crystallins	101
6.	Gene Duplication in the Crystallins	103

Chapter Three: Differential Kinetic Stability of Gamma D and Gamma S, and Their Isolated Greek Key Domains104

A.	Introduction	105
B.	Materials and Methods	107
1.	Preparation of Constructs for Wild Type and Isolated Domains	107
2.	Expression and Purification of Proteins	108
3.	Equilibrium Unfolding and Refolding	108
4.	Unfolding Kinetics	109
5.	Linear Extrapolation of Unfolding Kinetics	109
C.	Results	110
1.	Equilibrium Unfolding/Refolding Experiments at Different Temperatures and Equilibration Times in Hysteresis	110
2.	Equilibrium Unfolding/Refolding of Wild Type and Isolated Proteins	112

3. Unfolding Kinetics Analysis of the Wild Type and Isolated Proteins	112
4. Half-Chevron Plot Analysis Predicts Longer Lived Times for $\gamma_{D_{WT}}$ than for $\gamma_{S_{WT}}$	121
D. Discussion	126
1. $\gamma_{D_{WT}}$ but not $\gamma_{S_{WT}}$ Demonstrates a Kinetically Controlled Hysteresis	126
2. Kinetic Unfolding of the full-length γ -Crystallins and their Isolated Domains	127
3. Basis for Kinetic Stability in Crystallins	130
4. Evolutionary and Physiological Relevance for the Increased Kinetic Stability of γ_D Crystallins	131

Chapter Four: Domain-Exchanged Human Gamma D and S Crystallin Chimeras

Demonstrate Differential Aggregation Properties	134
A. Introduction	135
B. Materials and Methods	136
1. Expression and Purification of Proteins.....	136
2. Analytical Size Exclusion Chromatography	137
3. Circular Dichroism Spectroscopy	139
4. Fluorescence Spectroscopy	139
5. Thermal Denaturation	139
6. Equilibrium Unfolding and Refolding	140
7. Solution Turbidity Measurements	140
8. Aggregation Kinetics	141
9. Infrared (IR) Spectroscopy	141
C. Results	143
1. Preparation of Chimera Proteins	143
2. Purification of Chimera Proteins	147
3. Analytical Size Exclusion Chromatography	147
4. Circular Dichroism Spectroscopy	148
5. Fluorescence Spectroscopy	148
6. Thermal Denaturation of the Chimeras Compared to WT Proteins	153
7. Equilibrium Unfolding and Refolding of the Chimeras	156
8. Equilibrium Refolding Aggregation Properties of the Chimera Proteins	160
9. Aggregation Kinetics Demonstrate Differential Aggregation Properties	165
10. Aggregation Properties at Higher Protein Concentrations	166
11. IR Spectroscopy of the Chimeras	170
D. Discussion	172
1. Structural Aspects of the Chimera Proteins	172

2. Stability of the Chimera Proteins	173
3. Aggregation of the Chimeras	174
4. Summary of the Chimera Structural, and Aggregation Properties	176
5. A Domain-Swapping Mechanism for γ D _{WT} Aggregation Model.....	178
Chapter Five: Final Discussion	183
A. Summary of Chapters	184
Chapter Six: References	189
Chapter Seven: Appendices.....	209
A. Summary of Previous Work.....	210
B. Protein Parameters.....	212
C. Equilibrium Data Fitting Equations.....	213
1. Calculating Guanidine Hydrochloride Concentrations	213
2. Two-state Equilibrium Unfolding/Refolding.....	213
3. Three-state Equilibrium Unfolding/Refolding	214
D. Kinetic Data Fitting Equations.....	217
1. Two-state Kinetics.....	217
2. Three-state Kinetics.....	218
3. Four-state Kinetics	219

LIST OF FIGURES

Chapter One: Introduction	14
1-1 A diagram of the eye.....	16
1-2 The features of the lens.....	19
1-3 Model of cataract.....	21
1-4 Ubiquitous vertebrate crystallins.....	23
1-5 Sequence alignment of the three abundant γ -crystallins, γ_C , γ_S , γ_D	27
1-6 Phylogenetic tree of the β -and γ -crystallins and their gene structure.....	34
1-7 Greek Key topology two-dimensional schematic.....	39
1-8 Human γ_D and murine γ_S crystallin high resolution structure.....	41
1-9 Model of cataract <i>in vitro</i>	50
1-10 Serpin protease, α_1 -antitrypsin normal function and disease state.....	54
1-11 Mechanisms of domain-swapping.....	57
Chapter Two: Folding and Stability of the Isolated Greek Key Domains of the Long-Lived Human Lens Gamma D Crystallin	60
2-1 Crystal structure of human $\gamma_{D_{WT}}$ and NMR structure of murine $\gamma_{S_{WT}}$	64
2-2 Protein sequence alignment of $\gamma_{D_{WT}}$ and $\gamma_{S_{WT}}$	72
2-3 Analytical SEC of the wild type and isolated domain.....	74
2-4 Far-UV CD spectroscopy of wild type and isolated domains.....	77
2-5 Fluorescence emission spectra of wild type and isolated domains	80
2-6 Thermal denaturation of wild type and isolated domains.....	82
2-7 Equilibrium unfolding/refolding for $\gamma_{D_{WT}}$, γ_{D_N} , γ_{D_C}	84
2-8 Equilibrium unfolding/refolding for $\gamma_{S_{WT}}$, γ_{S_N} , γ_{S_C}	88
2-9 Kinetic refolding of $\gamma_{D_{WT}}$ and isolated domains	91
2-10 Kinetic refolding of $\gamma_{S_{WT}}$ and isolated domains.....	93
Chapter Three: Differential Kinetic Stability of Gamma D and Gamma S, and their Isolated Greek Key Domains	104
3-1 Equilibrium unfolding/refolding of $\gamma_{D_{WT}}$ under different conditions.....	111
3-2 Equilibrium unfolding/refolding of wild type and chimeras at 18°C.....	113
3-3 Kinetic unfolding at 5.5 M GuHCl at 18°C.....	117
3-4 Kinetic unfolding at 3.5 M GuHCl at 18°C.....	119
3-5 Linear extrapolations of kinetic unfolding rates for $\gamma_{D_{WT}}$ and isolated domains.....	123
3-6 Linear extrapolation of kinetic unfolding rates for $\gamma_{S_{WT}}$ and isolated domains.....	124
3-7 Model of $\gamma_{D_{WT}}$ and $\gamma_{S_{WT}}$ unfolding in no denaturant conditions.....	129

Chapter Four: Domain-Exchanged Human Gamma D and S Crystallin Chimeras	
Demonstrate Differential Aggregation Properties	134
4-1 Sequence alignment of γD_{WT} and γS_{WT} proteins.....	138
4-2 Homology modeling of the chimeras based on the γD_{WT} structure.....	142
4-3 Analytical SEC of the wild type and chimeras.....	149
4-4 Far-UV CD spectroscopy of wild type and chimeras.....	150
4-5 Fluorescence emission spectra of wild type and chimeras.....	152
4-6 Thermal denaturation of wild type and chimeras.....	155
4-7 Equilibrium unfolding/refolding of wild type and chimeras.....	158
4-8 Fluorescence spectroscopy of the refolding sample of wild type and chimeras	163
4-9 Solution turbidity of the refolding sample of wild type and chimeras	164
4-10 Aggregation kinetics of the wild type and chimeras	167
4-11 Equilibrium refolding aggregation at various protein concentrations.....	169
4-12 Infrared spectroscopy of the wild type and chimeras.....	171
4-13 Mechanism of γD_{WT} aggregation domain swapping model.....	181
4-14 Crystal structure of domain swapped dimer in bovine $\beta B2$ crystallin.....	182
Chapter Five: Final Discussion.....	183
5-1 Interface residues in the N-terminal and C-terminal domain of γD_{WT}	185
5-2 A model for cataractogenesis <i>in vitro</i> - the domain swapped model.....	186

LIST OF TABLES

Chapter One: Introduction.....		14
1-1	Amino acid abundance in proteins compared to γ C, γ D, γ S crystallin.....	26
 Chapter Two: Folding and Stability of the Isolated Greek Key Domains of the Long-Lived Human Lens Gamma D Crystallin.....		60
2-1	Deconvoluted CD spectra and fluorescence emission spectra maximums for γ D and γ S wild type and isolated domains.....	79
2-2	Equilibrium unfolding/refolding at 37°C and thermal parameters of γ D wild type and isolated domains	86
2-3	Equilibrium unfolding/refolding at 37°C and thermal parameters of γ S wild type and isolated domains	89
2-4	Productive refolding kinetic parameters of γ D and γ S wild type and isolated domains at 18°C.....	94
 Chapter Three: Differential Kinetic Stability of Gamma D and Gamma S, and their Isolated Greek Key Domains		104
3-1	Equilibrium unfolding and refolding parameters for γ D and γ S wild type and isolated domains at 18°C.....	114
3-2	Kinetic unfolding parameters for γ D and γ S wild type and isolated domains at 18°C.....	120
3-3	Linear extrapolated unfolding rate constants and half-lives of γ D and γ S wild type and isolated domains at 18°C	125
 Chapter Four: Domain-Exchanged Human Gamma D and S Crystallin Chimeras Demonstrate Differential Aggregation Properties		134
4-1	Composition of interface in wild type and chimera proteins.....	146
4-2	Deconvoluted CD spectra and fluorescence emission spectra maximums for γ D and γ S wild type and chimeras.....	151
4-3	Equilibrium unfolding and refolding at 37°C and thermal unfolding parameters of γ D and γ S wild type and chimeras.....	159
4-4	Summary of γ D and γ S wild type and chimera structural, stability, and refolding aggregation properties.....	177

CHAPTER ONE:

INTRODUCTION

A. VERTEBRATE EYE

The beauty and function of the eye have captured the attention of artists and scientists from Leonardo DaVinci to Charles Darwin to modern scientists. How the eye functions and how it evolved into various forms have been pressing questions addressed by generations of biologists. The ability of organisms to respond to light is common among most metazoans (Arendt 2003). However, the evolution of organs to use light to form images of their environment was a major step forward in early metazoan evolution. There are many different types of eyes ranging from simple pinhole eyes to complex terrestrial eyes.

The evolution of the vertebrate camera eye, with its image forming capacity, presumably provided selective advantage for predation in primitive organisms. In order to fulfill this purpose, the early vertebrate eye probably consisted of a simple small pinhole opening, but its light gathering capacity would have been limited (Fernald 2000). To correct for this, the spherical lens evolved to gather and refract light to a concentrated location allowing for increased visual field and acuity.

The multiple anatomical eye structures displayed by different species suggest that the eye evolved polyphyletically or in other words multiple times throughout evolution (Land 2005). However, there are complex genetic networks controlling eye development that are conserved in many organisms. These genes are called the master regulators of eye development because they can induce eye structures in other tissues (Gehring 2002). An example of this is *eyeless* or *Pax6*, a gene that is involved in the development of the brain and eye both in vertebrates and invertebrates. A loss of function mutation in the *Pax6* gene will result in a no eye phenotype; whereas a gain of function induces ectopic eye formation (Halder et al. 1995). In addition to the conserved genetic networks, pigmented photoreceptors are also conserved throughout multiple eye types (Nilsson 2004).

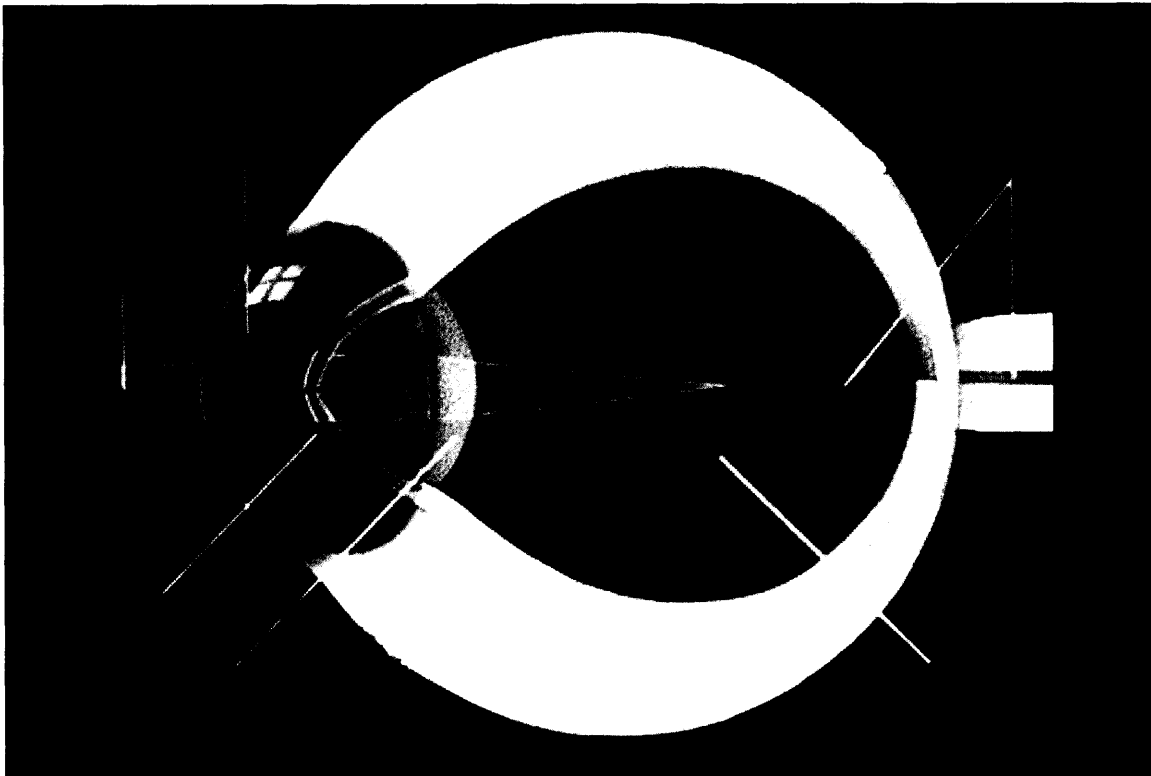


Figure 1-1. A diagram of the eye. Different tissues of the eye are noted: lens, cornea, pupil, retina, and optic nerve. (Courtesy of the National Eye Institute, National Institutes of Health)

B. HUMAN LENS ANATOMY AND COMPOSITION

The lens is a small, clear spheroid situated behind the cornea of the eye. Its main function is to fine focus light using its high refractive index properties onto the photoreceptors in the retina to allow for proper spatial vision (Fig 1-1). The structure, components, and proteins of the lens are essential for maintaining such visual acuity. The lens consists of the epithelial, cortical, and nuclear cellular layers, all derived from the embryonic surface ectoderm, also the source of the corneal epithelia. The nuclear core of the lens consists of primary fiber cells arranged in an ordered array. The secondary fiber cells can be as long as a few mm surrounding the nuclear core region of the lens.

The fiber cells are filled with proteins called the crystallins. There are three major vertebrate crystallins, α , β , and γ that comprise the major soluble components of the lens. Alpha (α)-crystallin provides a structural and chaperone function while the β and γ -crystallins are primarily structural proteins.

C. LENS DEVELOPMENT

Early in embryogenesis, the neural ectoderm forms the optic cup inducing the lens placode (fated ectoderm) to thicken and invaginate to form the lens vesicle (Henry and Grainger 1990). The posterior cells in the lens vesicle elongate and differentiate into the primary lens fiber cells while the anterior cells form the lens epithelial monolayer (Lovicu and Robinson 2004, Fig. 1-2A). The lens epithelium is a regenerative source of secondary lens fiber cells throughout life. Secondary lens fiber cells elongate and differentiate forming concentric layers of cells packed in hexagonal manner surrounding the nucleus. In the adult lens, the lens nucleus consists of primary lens fiber cells and mature (terminally differentiated) secondary fiber cells (Fig. 1-2B). The cortical region of the lens consists of gradually differentiating lens fiber cells, indicating that these outer layers remain capable of protein synthesis and regeneration.

During the development of the lens, the lens fiber cells terminally differentiate in an apoptotic-like process that degrades all organelles including the nucleus (Lovicu and McAvoy 2005). This process limits the scattering of light necessary for lens transparency and refraction. Therefore, the proteins in the central region of the lens that were expressed *in utero*, do not experience protein degradation or regeneration. The lens is continually adding cell layers onto these previously differentiated cells throughout life. Thus, the outer cortical and epithelial cell layers remain metabolically active.

The expression of crystallins is coordinated with the differentiation of the lens fiber cells. α -crystallins are expressed in epithelial cells but are upregulated upon lens fiber cell differentiation. β and γ -crystallins are lens fiber cell specific, varying in spatial and temporal developmental expression, predicted to be important for the refractive gradient (Bloemendal et al. 2004).

D. LENS TRANSPARENCY

Crystallins are extremely abundant in the human lens present at concentrations of 200-450 mg/ml (Fagerholm et al. 1981b; Siezen et al. 1988). The crystallins maintain high protein concentration, short-range order, and polydispersity in the lens contributing to its transparency and high refractive index (Delaye and Tardieu 1983; Tardieu 1988). As noted above, terminally differentiated fiber cells in the lens are enucleated and devoid of all other organelles to prevent diffraction of light and allow for proper focusing. Thus, the crystallins must remain soluble throughout life despite high concentrations of protein, continual UV exposure and oxidative stress.

The achievement of the uniform distribution of the crystallins has been studied by X-ray and visible light scattering studies. These studies have shown that light does scatter in lens protein solutions at concentrations lower than ~ 130 mg/ml. At higher concentrations, the light scatter decreased accordingly with increased protein concentration (Bettelheim and Siew 1983; Tardieu et al. 1992). This phenomenon was named short range order, resulting in a reduction of scattered light and contributing to transparency.

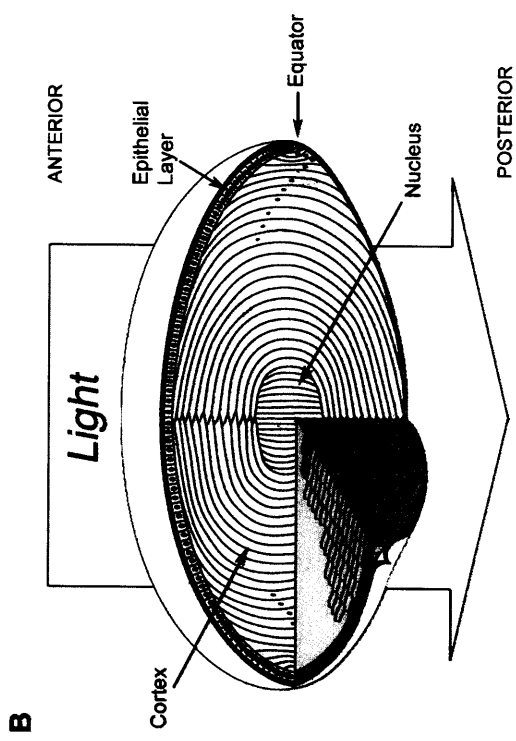
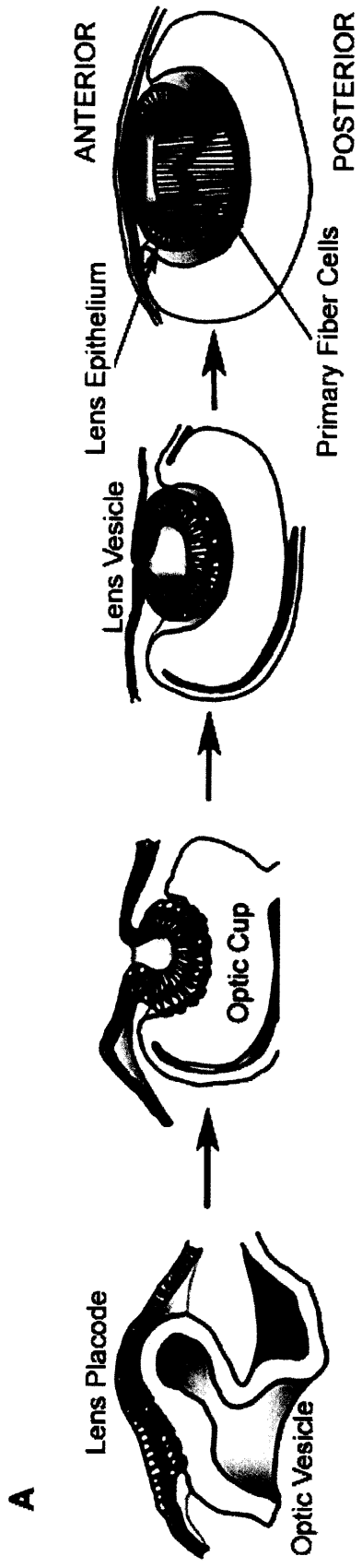


Figure 1-2. The features of the lens (A) Development and differentiation of the lens (B) Components of the lens (Modified from Lovicu and McAvoy, 2005, *Developmental Biology* 280: 1-14)

At such high protein concentrations, the polydispersity of the crystallins is proposed to prevent crystallization. The different sized and shaped crystallins and their surface properties contribute to this polydispersity. The crystallins can be thought of as building blocks fitting together to form a highly-ordered matrix (Fig. 1-3). The surface charge of the crystallins - the β -crystallins are repulsive while the γ -crystallins are attractive - is thought to be important for the short range order. Microequilibrium and surface plasmon resonance data of the different crystallins has demonstrated homologous and heterologous interactions (Biswas and Das 2004; Ponce et al. 2006). These coulombic interactions between the crystallins possibly assist in their solubility as well.

E. VERTEBRATE GENES FORMING THE LENS CRYSTALLINS

Several genes encode the vertebrate ubiquitous crystallins. There are four different groups of γ -crystallin, γM , γN , γS , and $\gamma A-F$. The γS crystallins subgroup is expressed in all vertebrates; while the $\gamma A-F$ crystallin expression is limited to mammalian systems and γM crystallin are mostly present in aquatic vertebrates (Wistow et al. 2005). γN crystallin is present in all vertebrate genomes, similar to γS . However, γN expression has only been found in some organisms with limited expression in specific tissues. In mammalian organisms, *γS crystallin* is located on a different chromosome from the *$\gamma A-F$ crystallins* which are clustered together in tandem on the same chromosome. The *$\gamma A-F$ crystallins* have the highest sequence similarity among one another.

There are two types of β -crystallins, acidic and basic. Both β -crystallins have N-terminal extensions and the basic crystallins have an additional C-terminal extension. The acidic β -crystallins include $\beta A1$, $\beta A2$, $\beta A3$, and $\beta A4$, while the basic β -crystallins include $\beta B1$, $\beta B2$, and $\beta B3$. The $\beta A1$ and $\beta A3$ crystallins are expressed by the same gene with the $\beta A3$ transcript initiating more upstream than $\beta A1$. This difference corresponds to a longer N-terminal extension. There are two separate functional α -crystallin genes, αA is expressed in a lens specific manner while αB is expressed in other tissues such as the muscle, brain, and heart.

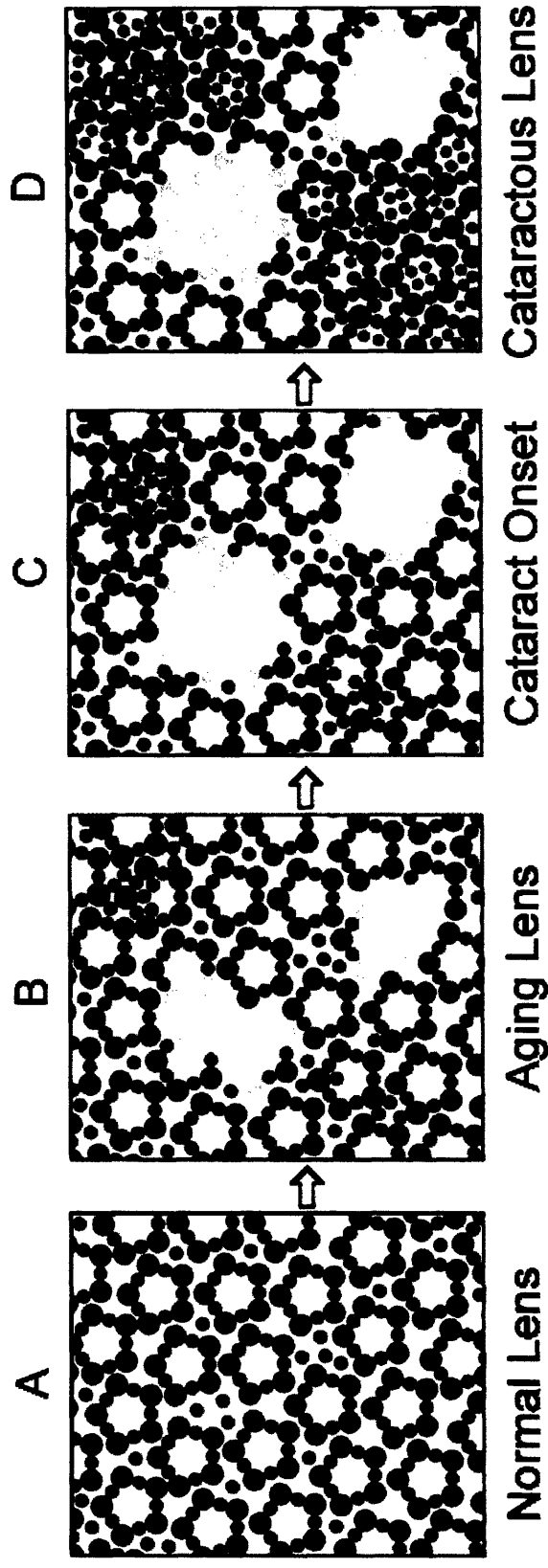


Figure 1-3. Model of cataract. (A) Schematic of crystallins polydispersity in normal lens (B) Aging lens introduce slight fluctuations (C) Beginning stages of cataract with some opacity (D) Cataractous lens opacity (Modified from Ponce et al. *Mol Vis.* (2006) 12: 879-84)

In the human genome, $\gamma A-F$ crystallin genes are in tandem repeats on chromosome 2, and γS crystallin is located on chromosome 3. Both γE and γF are pseudogenes, and γN (located on chromosome 7) transcripts have not been detected in the lens (Hejtmancik et al. 2001). Among the γ -crystallins that are expressed, γC , γS and γD are most abundant in that order whereas γA and γB are present at low levels. The β -crystallin $\beta A4$, $\beta B2$, $\beta B3$, and *pseudo* $\beta B2$ cluster is located on chromosome 22. In contrast other β -crystallins, $\beta A1$ and 3 are found on chromosome 17 while $\beta A2$ is on chromosome 2 near the $\gamma A-F$ cluster. For the α -crystallin genes, the αA gene is located on chromosome 21, and αB gene on chromosome 11 (Hejtmancik et al. 2001).

In mammalian organisms, β - and γ -crystallins are fiber-cell specific. In contrast, α -crystallins are expressed both in the epithelial layer of the lens and in the lens fiber cells. In particular, $\gamma A-F$ crystallins and $\beta B1$ crystallins are expressed *in utero* and are localized to the lens nucleus (Chambers and Russell 1991; Flaugh et al. 2006; Lampi et al. 1998). In contrast, $\beta B2$ and γS -crystallin, are postnatally expressed in the secondary fiber cells and are expressed throughout life (Peek et al. 1992b; Ueda et al. 2002; Wistow et al. 2002).

F. CONFORMATION AND PROPERTIES OF THE LENS CRYSTALLINS

The ubiquitous vertebrate crystallins, α -, β -, and γ -crystallins are small globular proteins. α -crystallins and γ -crystallins are approximately 20 kDa and β -crystallins are slightly larger, ranging from 23kDa - 28 kDa in size. The crystallins lack metals or prosthetic groups as well as in most cases disulfide bonds. Structurally, the crystallins can be divided into two separate families, the α -crystallins and the $\beta\gamma$ -crystallins. The structure of α -crystallin differs from the β - and γ -crystallins in that it is thought to have one β -sandwich domain while the β - and γ -crystallins have two β -sandwiches consisting of double Greek Key motifs. The stabilities of the crystallins vary with the monomeric γ -crystallin more stable than the oligomeric β - and α -crystallins (Fig. 1-4).

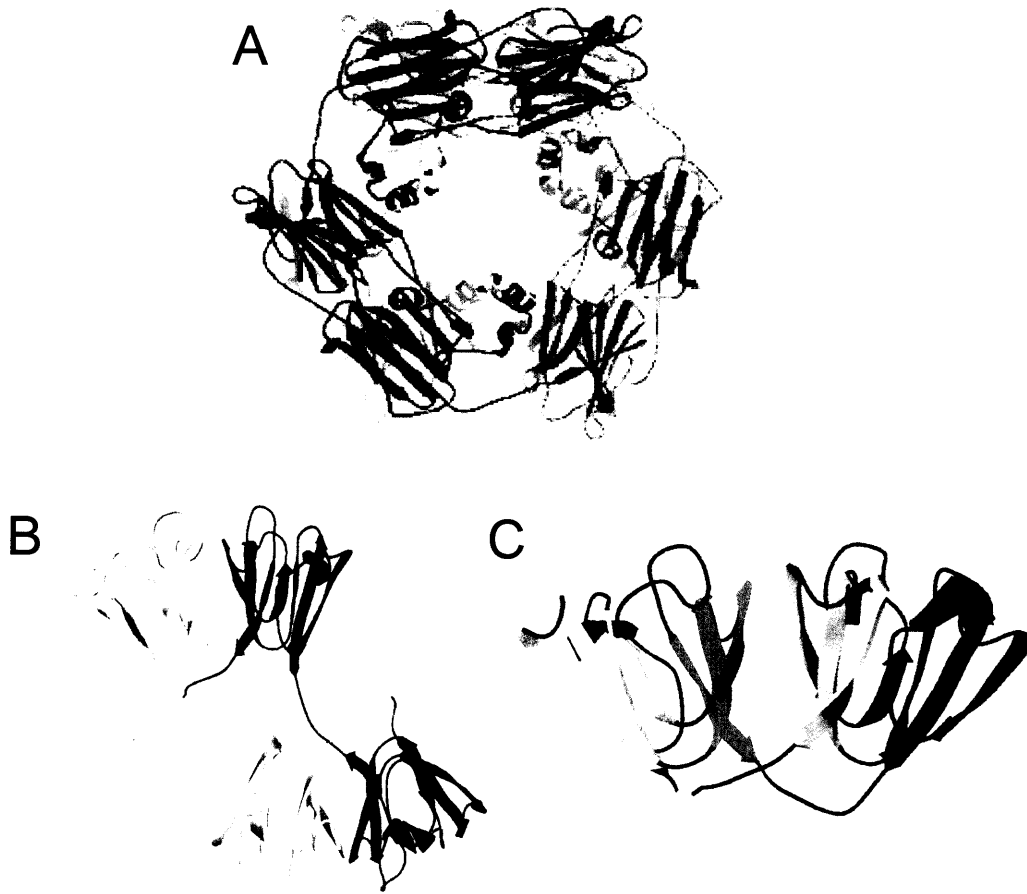


Figure 1-4. Ubiquitous vertebrate crystallin proteins (A) Wheat small heat shock protein (Hsps) polymer a model of α -crystallin oligomerization (B) X-ray crystallography structure of the β B2 crystallin (C) X-ray crystallography structure of the γ crystallin (Modified from Slingsby, C. and Clout, N.J. Eye (1999) Pt 3b (395-402))

1. α -crystallin

There are two α -crystallin proteins in the lens, α A (lens specific) and α B which are at a ratio of 3 α A:1 α B (Horwitz 2003). α -crystallin forms polydispersed heterooligomers with complexes ranging from 300 kDa to 1200 kDa (MacRae 2000). A crystal structure of the α -crystallin has yet to be solved likely due to its polydispersed nature. Both α A and α B are homologous to the small heat shock (sHsp) family of molecular chaperones (de Jong et al. 1998). Many members of this family also form polydispersed multimers.

The sHsps have three major regions, a variable N- and C- terminal extension and a highly conserved β -sandwich α -crystallin domain (Mornon et al. 1998). Circular dichroism (CD) and infrared spectroscopy (IR) analyses support the β -sandwich conformation with unstructured N-terminal and C-terminal regions (Bloemendal et al. 2004). An X-ray structure of a wheat sHsp dodecamer polymer demonstrates the formation of a higher order structure by association with the N- and C-terminal regions (Van Montfort et al. 2001; van Montfort et al. 2001, Fig. 1-4). In addition, cryo-electron microscopy reconstructions of α B-crystallin suggest that it forms a micelle-like structure \sim 19 nm in diameter, having a central cavity with a diameter of \sim 8 nm (Haley et al. 1998; Haley et al. 2000).

This proposed structure along with chaperone assays of α -crystallin *in vitro* has led to a plausible mechanism of α -crystallin function. In addition to its structural role, it is hypothesized that a function of α -crystallin *in vivo* involves binding misfolded proteins to prevent aggregation (Clark and Muchowski 2000; Ganea and Harding 2000). There is no evidence that α -crystallin refolds proteins by itself, therefore acting as an ATP-independent passive chaperone. The proposed cavity of the α -crystallin provides an attractive model in which the aggregation susceptible crystallins may be internalized into the cavity.

α -crystallin demonstrates lower stability than the $\beta\gamma$ -crystallins, with α A slightly more stable than α B (Sun et al. 1999). Thermal stability assays revealed a T_M of $\sim 60^\circ\text{C}$ and both α A and α B crystallin initiate unfolding at lower concentrations of denaturant than the $\beta\gamma$ -crystallins (Das and Liang 1997; Santini et al. 1992).

2. β - and γ -crystallins

The β - and γ -crystallins are proposed primarily to function as structural proteins. Structurally, the vertebrate β and γ -crystallins are similar in that they contain two highly symmetrical domains with an interface connected by an interdomain linker (Fig. 1-4). Other structural commonalities are the intercalated double anti-parallel β -sheet Greek Key motifs, in which strand d of the first motif is paired with strand c of the second motif (Bax et al. 1990). Also, a tyrosine corner present in each domain is thought to be stabilizing to this β -hairpin of the Greek Key (Bax et al. 1990). The tyrosine corner is a hydrogen bond between the tyrosine on strand c of the second motif with the backbone of strand d of the same motif and is thought to be common among β -sandwich Greek Key containing proteins (Hemmingsen et al. 1994). Only a few structural differences are observable between β and γ -crystallins such as the β -crystallins N- and C-terminal extensions as well as their ability to associate.

The β - and γ -crystallin sequences are only 30% conserved. However, the interdomain interface and the tyrosine corner have high sequence similarity. The domain interface has a hydrophobic region and peripheral amino acids that are situated above and below the hydrophobic residues. The peripheral residues may act as a barrier to the solvent surrounding the hydrophobic region (Flaugh et al. 2005a). Also, there is high conservation of the four tryptophans in the buried core of the crystallins and tyrosines throughout the crystallins, many of which are surface exposed.

In the human lens, the major crystallins are the closely related γ C, γ S and γ D (Fig. 1-5). Their amino acid composition is not unusual except possibly for a higher

percentage of Arg and a higher percentage of aromatic residues consistent with the other crystallins (Table 1-1). On the other hand, in γ D and γ C crystallin, the higher concentration of Arg residues is offset by a reduction in the Lys residues. It has been predicted that the Arg residues can contribute to solubility. Thus, the higher prevalence of Arg may assist in the solubility of the protein.

Table 1-1. Amino acids abundance in proteins compared to γ C, γ D, γ S crystallin composition. McCadon, P. et al., *Proteins: Structure, Function and Genetics* 1988, 4:99-122.

Amino Acid	Average	H γ C-crystallin	H γ D-crystallin	H γ S-crystallin
	Percent	Percent	Percent	Percent
Ala (A)	7.34%	2.30%	2.30%	3.90%
Arg (R)	5.2%	11.50%	12.10%	7.30%
Asn (N)	4.57%	2.30%	4.00%	2.80%
Asp (D)	5.12%	5.70%	6.90%	5.60%
Cys (C)	1.76%	4.60%	3.40%	3.90%
Gln (Q)	3.96%	6.90%	5.70%	5.10%
Glu (E)	6.22%	7.50%	5.70%	7.90%
Gly (G)	6.89%	6.30%	8.00%	7.90%
His (H)	2.26%	2.30%	3.40%	2.20%
Ile (I)	5.76%	3.40%	3.40%	5.60%
Leu (L)	9.36%	10.90%	9.80%	5.10%
Lys (K)	5.81%	1.70%	0.60%	5.60%
Met (M)	2.32%	3.40%	2.90%	3.40%
Phe (F)	4.12%	2.30%	3.40%	5.10%
Pro (P)	5.00%	4.60%	2.90%	4.50%
Ser (S)	7.38%	7.50%	9.80%	6.20%
Thr (T)	5.85%	2.90%	1.70%	3.90%
Trp (W)	1.34%	2.30%	2.30%	2.20%
Tyr (Y)	3.25%	8.00%	8.00%	7.90%
Val (V)	6.48%	3.40%	3.40%	3.90%

```

Human $\gamma$ C MG----KITFYEDRAFQGRSYETTTDCPNLQPYFSRCNSIRVESGCWMLYERPNIYQGQQY 56
Human $\gamma$ D MG----KITLYEDRGFQGRHYECSSDHPNLQPYLSRCNSARVDSGCWMLYEQPNYSGLQY 56
Human $\gamma$ S MSKTGTKITFYEDKNFQGRRYDCDCDCADFHTYLSRCNSIKVEGGTWAVYERPNIYAGYMY 60

Human $\gamma$ C LLRRGEYPDYQQWMGLSDSIRSCC--LIPQTVSHRLRLYEREDHKGLMMELSEDCPSIQD 114
Human $\gamma$ D FLRRGDYADHQQWMGLSDSVRSCR--LIPHSQSHRIRLYEREDYRGQMIEFTEDCSCLQD 114
Human $\gamma$ S ILPQGEYPEYQRWMGLNDRLLSSCRAVHLPSGGQYKIQIFEKGFSGQMYETTTEDCPSIME 120

Human $\gamma$ C RFHLSEIRSLHVLEGCWVLYELPNYRGRQYLLRPQYRRCQDWGAMDAKAGSLRRVVDLY 174
Human $\gamma$ D RFRFNEIHSLNVLEGSWVLYELSNYRGRQYLLMPGDYRRYQDWGATNARVGSLLRRVIDFS 174
Human $\gamma$ S QFHMREIHSCKVLEGVWIFYELPNYRGRQYLLDKKEYRKPIDWGAASPAVQSFRRIVE-- 178

```

Figure 1-5. Sequence alignment of the three abundant γ -crystallins in the human lens. γ C and γ D crystallins have 72% identity, 83% similarity; γ D and γ S crystallins have 50% identity, 69% similarity; γ C and γ S crystallins have 53% identity and 70% similarity.

Solved structures of the β -crystallins show two different ways that the β -crystallin monomers can form oligomers. In β B2, the dimer formation is a domain swapped conformation in which the N-terminal domain of molecule A interacts with the C terminal domain of molecule B. The domain linker is extended to allow for intermolecular interactions (Bax et al. 1990). On the other hand, in β B1, the linker is bent as in the γ -crystallins and the β B1 monomers interact intramolecularly. The two monomers associate forming another interface in the β B1 dimer (Bateman et al. 2001; Van Montfort et al. 2003). This other interface is the same interface observed in the β B2 tetramer, in which domain-swapped dimers associate (Bloemendal et al. 2004).

In addition to forming homo-oligomers, the β -crystallins can form hetero-oligomers. Bateman et al. has shown that β A3 and β A4 crystallins can undergo subunit-exchange to oligomerize with β B1 (2003). These hetero-oligomers and homo-oligomers formed by the β -crystallins can contribute to the polydispersity of the crystallins, inhibiting crystallization in the lens as well as promoting an even distribution of protein in the lens, leading to a higher refractive index.

Comparative experimental studies between β and γ -crystallins have shed some insight into what structural differences are involved in their association properties. Both have analogous interface residues; therefore, the linker composition between the domains has been previously studied to determine its role in oligomerization. For instance, differences between β B2 and γ B crystallin are an extended linker in β B2 between the domains and the N- and C- terminal extensions of β B2. To study the nature of this association empirically, the linker sequences of β B2 and Bovine γ B crystallins were exchanged and oligomerization was determined. Replacing the γ B linker with β B2 crystallin's linker changed the oligomerization of β B2 to a monomeric state (Trinkl et al. 1994). Substituting the β B2 linker in γ B did not change the monomeric state of γ B-crystallin (Mayr et al. 1994). These results emphasized the importance of the interdomain interface association in maintaining the monomeric state of the γ -crystallins. Thus, the γ -crystallin linker may serve primarily to lower the entropic cost for these

interactions to be favorable. Contrary to γB , these results suggest that for $\beta\text{B}2$, the conformation and length of the linker is crucial for the oligomerization and stabilization of the dimer. The intermolecular interface also plays an important role in the stabilization of the dimer.

3. Stability of the $\beta\gamma$ -crystallins

The β - and γ -crystallins have provided one of the major experimental systems for the investigation of domain-domain interactions in proteins and of domain-swapping in multimeric proteins (Bloemendal et al. 2004). Although the majority of $\beta\gamma$ -crystallins have the same structural features, their stabilities vary quite considerably. Several studies of the β - and γ -crystallins have exhibited differential domain stability in their individual domains. These isolated domain studies have revealed the importance of the domain interface in the overall conformational stability.

The $\beta\gamma$ -crystallin family exhibits very high stability, albeit varying in stability with β -crystallins considerably less stable than γ -crystallins. These $\beta\gamma$ -crystallin properties have been suggested to be due to the complex topology of the $\beta\gamma$ -crystallin Greek Key fold (MacDonald et al. 2005). Another contributor is the stronger interface contacts in the γ -crystallins (Mayr et al. 1994; Palme et al. 1997; Wieligmann et al. 1999). Mutations within the interface destabilized both bovine γB and human γD -crystallins (Flaugh et al. 2005b; Palme et al. 1997; Palme et al. 1998b). Prior to this thesis, isolated domain studies have been performed in γB , $\beta\text{B}2$, and γS crystallins.

Thermodynamic studies of bovine γB crystallin demonstrated a bimodal equilibrium transition at pH 2.0, in which the N-terminal isolated domain was found to be more stable than the C-terminal isolated domain (Rudolph et al. 1990). This difference in domain stabilities was attributed to electrostatic repulsions in the C-terminal domain. The C-terminal domain had a net charge of +16 compared to the N-terminal domain which had a net charge of +13 at acidic pH. At pH 7.0, closer to the isoelectric point of

both domains the thermodynamic stabilities were similar (Mayr et al. 1997). Thus at acidic pH, the C-terminal domain's repulsion led to destabilization.

For β B2 crystallin, biophysical studies have shown that the C-terminal domain is more stable than the N-terminal domain (Wieligmann et al. 1999). The N-terminal domain in isolation was found to be destabilized at low concentrations without denaturant. At higher concentrations, the N-terminus dimerized and exhibited marginal stability. This work suggested a folding pathway for β B2 in which C-terminal domain folds first and the N-terminal domain associates with its dimeric partner assisting in the folding of the protein. These studies suggest that the interface contributes to the overall stability of the protein.

Human and bovine γ S studies also demonstrated differential domain stability. Wenk et al. found that in both orthologs, the N-terminal domain was similar to the stability of the C-terminal domain (2000). The hypothesis suggested was that the interface interactions in γ S crystallin were not as strong as in other γ -crystallins. In this thesis, the analysis of the isolated domains demonstrated that the N-terminal domain is less stable thermally, thermodynamically, and kinetically than the C-terminal domain. The hypothesis discussed in this thesis is that the contribution of the interface is not seen in the equilibrium experiments since the (un)folding transitions of the individual domains overlay or because of the folding cooperativity exhibited by wild type γ S crystallin. However, in the context of the full-length protein, the N-terminal domain is stabilized by the presence of the C-terminal domain through its interdomain interactions.

4. *Microbial Crystallins*

Although separated as a different section in this thesis, the microbial crystallins have a β γ -crystallin fold. The most extensively studied microbial crystallins include Protein S and Spherulin 3a. Protein S is a spore coat protein in the soil bacterium, *Myxococcus xanthus*, that is activated upon starvation of the bacterium (Inouye et al. 1979; Wistow et al. 1985). Spherulin 3a is also a protein that is most abundantly

expressed upon starvation and stress in *Physarum polycephalum*, a slime-mold (Bernier et al. 1986; Bernier et al. 1987; Wistow 1990). It is interesting to note that these two microbial crystallins have stress-related roles.

Crystallography and nuclear magnetic resonance (NMR) studies of these two microbial crystallins reveal several structural features common to the vertebrate $\beta\gamma$ -crystallins. Protein S is also a two domain protein, although the interdomain interface differs from the vertebrate crystallins and is not symmetrical (Bagby et al. 1994a; Bagby et al. 1994b; Bagby et al. 1994c). Both domains of Protein S contain the tyrosine corner motif and similar hydrophobic core within each domain (Clout et al. 1997). Contrary to Protein S, Spherulin 3a is a single domain crystallin that dimerizes but differs from other $\beta\gamma$ -crystallins in that it does not contain a tyrosine corner (Clout et al. 2001; Rosinke et al. 1997). The dimer interface in the Spherulin 3a observed in the crystal structure is not the conserved interdomain interface of the vertebrate crystallins. However, the domain interface contacts are extremely strong since appreciable amounts of monomer are not present in equilibrium. Additional structural features of the microbial crystallins are symmetrical Ca^{2+} binding sites in each domain, thought not to be present in the vertebrate $\beta\gamma$ -crystallins (Clout et al. 2001). In Protein S, Ca^{2+} binding promotes polymerization of the protein forming the spore coat in *Myxococcus* (Kaiser et al. 1979).

Similar to the vertebrate crystallins, the microbial crystallins exhibit high stability which increases upon binding of Ca^{2+} . In attempts to decipher the folding pathway of Protein S, isolated domain thermodynamic analyses were performed. These studies from Protein S showed differential thermodynamic stability like the vertebrate β - and γ -crystallins, with the N-terminal domain being more stable than the C-terminal domain (Wenk and Jaenicke 1999). This difference is attributed to a weaker Ca^{2+} binding site in the N-terminal domain of the protein. Protein S also demonstrates kinetic stability with increased stability in the context of the full-length protein due to interface contacts (Wenk et al. 1998). The rates of unfolding of the individual N-terminal domain and C-terminal domain of Protein S were similar and exhibited a 100-fold increase compared to

the full-length protein suggesting that the interface contacts were important in the kinetic stability of the protein (Wenk et al. 1998).

Spherulin 3a has demonstrated kinetic and thermodynamic stability by DSC measurements, equilibrium unfolding/refolding and kinetic unfolding studies (Kretschmar and Jaenicke 1999; Kretschmar et al. 1999a; Kretschmar et al. 1999b). Kinetic unfolding studies of Spherulin 3a revealed even higher kinetic stability compared to Protein S with an extrapolated half-life of ~12 days in strong chaotropic denaturant (Guanidinium thiocyanate) (Kretschmar et al. 1999a). Interestingly, upon binding Ca^{2+} , Spherulin 3a exhibits fluorescence quenching in the native state, similar to other vertebrate β - and γ -crystallins. The structural, functional, and stability properties of the microbial crystallins suggest that they may be related to the ubiquitous vertebrate crystallin ancestral protein.

G. THE CRYSTALLIN PROTEINS AND LENS EVOLUTION

1. Vertebrate Ubiquitous Crystallins

The lens is not preserved in fossil specimens making it difficult to study its evolution through paleontology. However, advances in genomics allowing for comparative studies of lens related genes have provided insight into lens evolution. Studies of vertebrate lens evolution have advanced recently due to the discovery of an ancestral single domain $\beta\gamma$ -crystallin present in the urochordate sea squirt, *Ciona intestinalis* (Shimeld et al. 2005). Urochordates are closely related to the ancestor of modern vertebrates. Protein and structural alignment comparisons support a tree in which the β and γ crystallins evolved from this common ancestral protein (Fig. 1-6). Surprisingly, the regulatory promoter of this single domain had the ability to be expressed in the vertebrate, *Xenopus laevis*. It was previously believed that the lens arose after the split of the chordates to the vertebrate lineage; however, this new finding suggests that $\beta\gamma$ -crystallin was expressed in the neuroectodermal tissue before vertebrate lens evolution (Shimeld et al. 2005).

The discovery of the non-dimeric single domain protein confirms the idea that a gene duplication event occurred in the evolution of the crystallins. Phylogenetic trees comparing genes, intronic sequences/structures and protein structures have suggested that there are four different γ -crystallin classes (γ A-F, γ M, γ N, and γ S) (Fig. 1-6) (Wistow et al. 2005). γ N crystallin is the evolutionary connection between β - and γ -crystallins according to its gene and protein structure similarities to both groups. These diverse subgroups of γ -crystallins are important for the different optical properties required for a diverse set of organisms. While the sequence and structural alignments of the crystallin proteins have given us a phylogenetic lineage and information on the evolution of the crystallin family, biophysical properties may also give insight into the importance of the gene duplication of these proteins as well as the stability of these proteins.

Since $\beta\gamma$ -crystallins have structural and gene similarities with microbial stress proteins, it has been suggested that the crystallin ancestor had a primitive role in stress and was possibly recruited to the vertebrate lens for its stress-related properties. The crystallins are present in other tissues such as the retina and the presence of non-lens $\beta\gamma$ -crystallins such as the Absent in Melanoma (AIM) protein associated with a suppression of malignancy in melanomas, contributes to the idea of recruitment (Aravind et al. 2006; Rajini et al. 2003).

2. *Taxon-specific crystallins*

In addition to the vertebrate α , β , and γ crystallins, invertebrates and vertebrates have recruited other proteins for crystallins. These are called taxon-specific crystallins. Many of the taxon-specific crystallins are believed to have evolved through a gene-sharing mechanism (Piatigorsky and Wistow 1989). Gene sharing refers to using identical genes are utilized for two different functions.

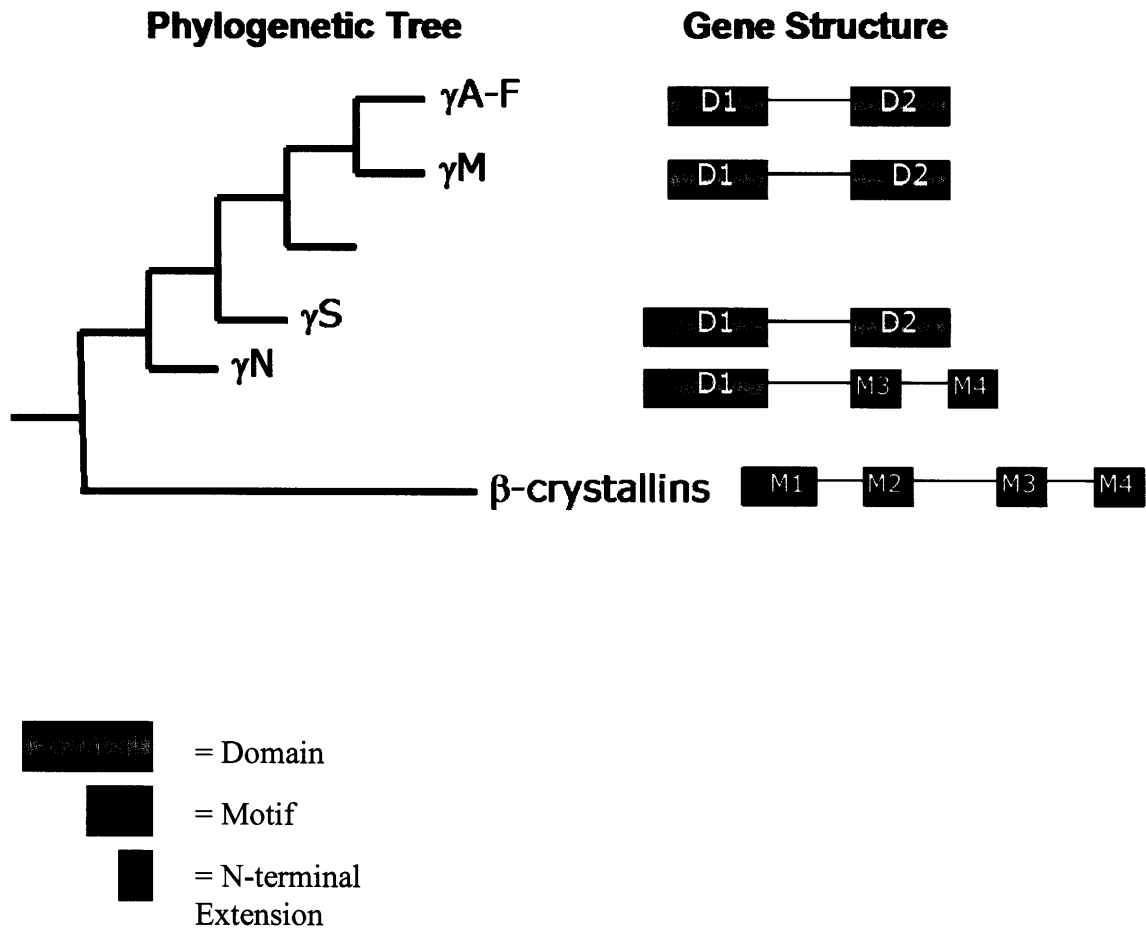


Figure 1-6. Phylogenetic tree of the β - and γ -crystallins and their gene structure. (Adapted from Shimeld, S.M. et al. *Current Biology* (2005), 15: 1684-1689)

An example of gene sharing in the lens is the lactose dehydrogenase (LDH)/ ϵ -crystallin of the avian lens and the argininosuccinate lyase (ASL)/ $\delta 2$ crystallin (Hendriks et al. 1988; Piatigorsky et al. 1988). These housekeeping proteins exhibit dual roles as both metabolic enzymes and refraction proteins in the lens. Some of the taxon-specific crystallins have retained their enzymatic activity. Another avian crystallin, $\delta 1$, does not have the same gene and seems to be a gene duplication of the ASL gene (Thomas et al. 1990). The avian $\delta 1$ -crystallin has adapted to the lens environment by sequence changes that increased its thermal stability (Voorter et al. 1993). The recruitment of enzymes to act as refractive proteins may be associated with the need for high expression of crystallins to produce the lens's high refractive index properties.

H. β -SHEET PROTEIN FOLDING

The folding mechanism of β -sheet proteins are still not well understood and are lagging far behind α -helical proteins. The structural properties of β -sheet proteins are far more complex than α -helices primarily because of the separation in linear sequence compared to the interacting residues of the β -sheet secondary structure. For example, although RNase has been extensively studied, pathways leading to the formation of the three β -strands still have not been elucidated. Although some of the available techniques can catch submillisecond transitions, studies of β -sheet model proteins still have difficulties analyzing the fast folding intermediates, which are thought to be mostly β -turns and β -hairpins. β -hairpins and β -turns seem to be the first to fold and unfold in a β -sheet structure (Searle and Ciani 2004). Protein engineering methods to create one β -hairpin or β -turn in conjunction with submillisecond techniques are being used to identify initial steps in β -sheet folding. The formation of these intermediates may involve non-native contacts. Additionally, computational studies are being utilized (Hughes and Waters 2006). In the future these techniques may provide further insights.

Model small globular β -sheet proteins have been utilized to studying the early intermediates formed during the folding pathway such as Interleukin-1 β and ubiquitin. In

studies of these model proteins, the first folding intermediates identified thus far have the β -strand already formed. In many cases, the first folding intermediates have non-native contacts. Multiple techniques have been employed in determining the structure of these intermediates such as hydrogen exchange/NMR (HX)/NMR or NMR at various equilibrium or kinetic phases. Further studies may lead to insight into the structure of the intermediates containing non-native interactions.

Interleukin-1 β is a β -trefoil protein with six two-stranded β -hairpins. CD studies have suggested a molten globule intermediate with 90% secondary structure forms on a millisecond timescale along the folding pathway (Ptitsyn et al. 1990). However, (HX)/NMR detected intermediates after 1 second demonstrating the rapid folding of β -sheets. After this time period, a transient intermediate formed after which a more native-like intermediate is structured (Varley et al. 1993). Taken together, these studies suggested that the fast non-native β -sheet intermediate is formed before the more native-like structured β -sheet on the folding pathway (Varley et al. 1993).

Ubiquitin is a small 76 amino acid protein that has a five β -strand structure consisting of three anti-parallel β -strands and one parallel β -strand as well as two α -helices. Although upon initial inspection, the folding kinetics appeared to be two-state, further experiments indicated that this protein may have kinetic folding intermediates (Khorasanizadeh et al. 1993). (HX)/NMR folding experiments indicated that the N-terminal region of the protein consisting of an α -helix and a β -hairpin was more structured than the C-terminal region (Went and Jackson 2005). This intermediate was able to be stabilized. More recent techniques including pressure unfolding and computational analysis also suggest that the folding nucleus of the protein may be this α -helix, β -hairpin region. It also supports the idea that these early transient intermediates may involve non-native contacts.

I. OTHER PROTEINS WITH GREEK KEY MOTIFS

A type of β -sheet fold is the Greek Key motif which obtained its name from the symbol found on Greek Attic vases which resembles the anti-parallel β -sheets of the protein fold. The Greek Key containing proteins consist of at least four anti-parallel β -sheets; for instance, with one Greek key motif strands 1 and 4 interact and strands 2 and 3 interact (Fig. 1-7). However, the presence of only one Greek Key motif in isolation has not been found in nature (Jaenicke and Slingsby 2001). Greek Key motifs are found in many β -barrel and β -sandwich structures which can form a diverse set of topologies. The β -barrels form a closed structure consisting of β -twists and coils and participate in hydrogen bonding of the first β -strand to the last β -strand. β -sandwiches are stacked β -sheets intertwined to maximize the interactions between the two sheets. Both types can contain right and left handed Greek Key motifs and can form multiple types of orientations (Zhang and Kim 2000). The mechanism of Greek Key folding is still under investigation. In theory there are a number of ways to obtain the configuration of the Greek Key. However, common conformations are observed in the β -sandwich proteins involving a five-stranded overlapping Greek Key, that is, the Greek Key is split between the two stacked sheets (Zhang and Kim 2000).

Examples of well-studied β -sandwich proteins are the immunoglobulins. The Greek Key motifs observed in these proteins are common to the Ig-like structural family of proteins with diverse functions. Some additional examples of this structural family include fibronectins and cadherins. It has been proposed that due to the similarities in the kinetics and stabilities of these proteins that they have a common folding nucleus (Clarke et al. 1999). Commonalities include interactions among some of the interacting β -strands of the proteins and were confirmed by computational methods (Mirny et al. 1998). This folding nucleus may be important for understanding the folding of the Greek Key β -sandwich proteins.

Another example of a folding nucleus in the Greek Key domains was found in NMR studies of the N-terminal isolated domain from the microbial crystallin, Protein S.

Since the N-terminal domain of Protein S exhibited non-coincidence in fluorescence spectroscopy and CD studies, this suggested that N-terminal domain was not a two state folder. NMR on equilibrium intermediates identified a β -hairpin region of the Greek Key that was structured. This region was in the second motif of the Greek Key (innermost Greek Key, closest to the interface) including the tyrosine corner and may act as a nucleus for the folding of the protein (Bagby et al. 1998).

In Chapter two of this thesis, refolding kinetics of the human γ D and γ S crystallins detected a partially folded intermediate in the formation of the isolated domains (two intercalated Greek Key motifs). This intermediate may indicate a sequential folding of each Greek Key motif within the domain or alternatively may be a combination of folding both Greek Key motifs. The results from the aforementioned Bagby et al. support the sequential Greek Key motif folding hypothesis in which the region containing the tyrosine corner in the second motif is the first to fold (1998). Similar to the Ig-like structural family, analysis of these refolding intermediates may give insight into how the Greek Key domains fold.

J. FOLDING, UNFOLDING AND AGGREGATION OF HUMAN γ D CRYSTALLIN

Fluorescence spectroscopy has been beneficial in determining the folding and stability of the crystallins. The γ -crystallins have four conserved tryptophans that are buried in the four different quadrants of the protein core. Two of the tryptophans are located in the upper regions of the two domain protein in near symmetrical positions while the other two are located in the bottom region (Fig. 1-8). Human γ D crystallin (γ D_{WT}) and many of the other β and γ -crystallins have tryptophans that are quenched in the native state. The tryptophan quenching phenomenon in γ D_{WT} has recently been elucidated. It involves a charge transfer mechanism in which the top tryptophans, Trps 68 and 156, undergo a charge transfer to the amide backbone polypeptide chain. The moderately fluorescent bottom tryptophans, Trps 42 and 130, undergo an energy transfer mechanism to their corresponding top tryptophans, Trps 68 and 156 (Chen et al. 2006).

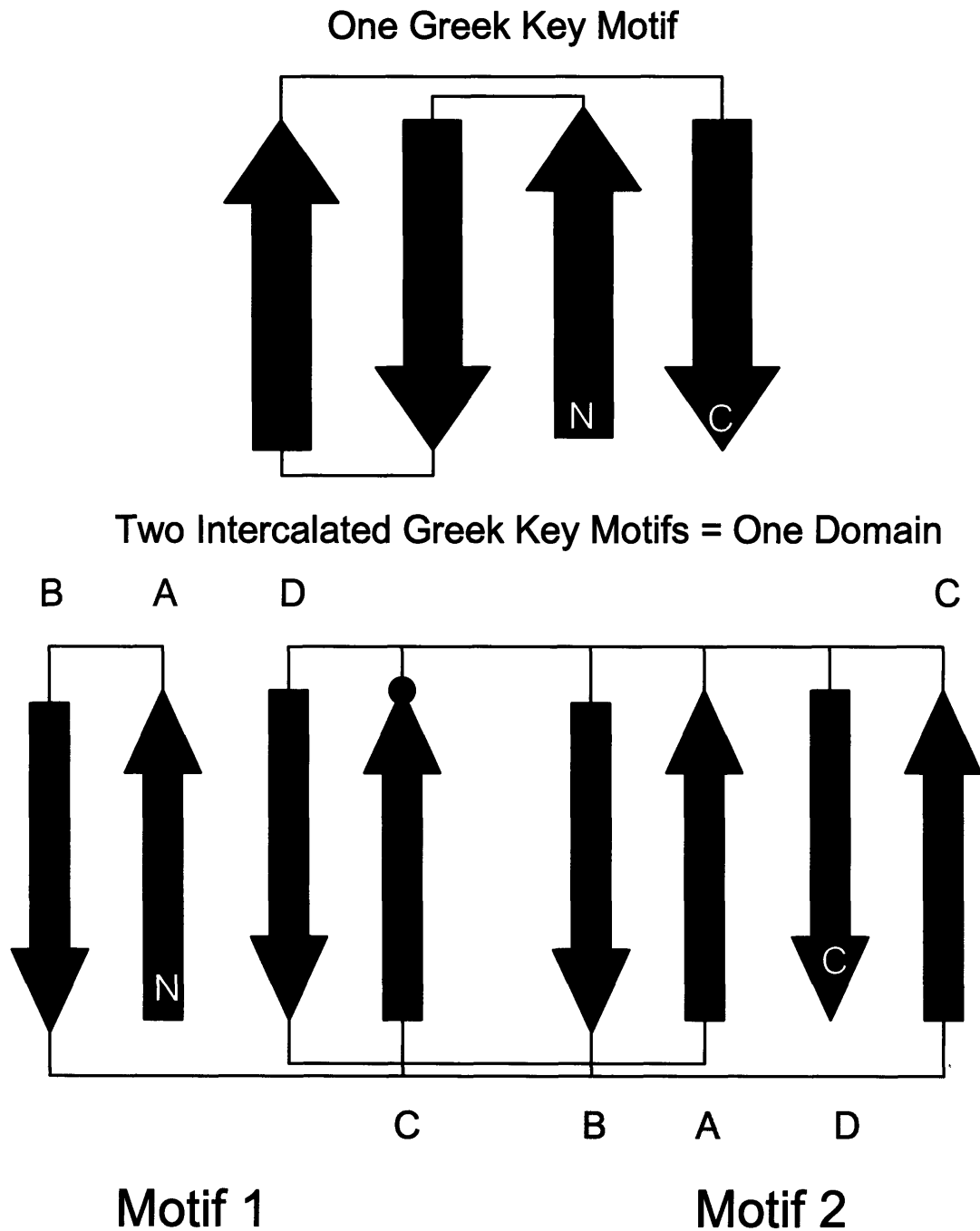


Figure 1-7. Greek Key topology two-dimensional schematic. N and C (white) represent the N-terminal and C-terminal domain. Top: One Greek Key motif. Bottom: Two intercalated Greek Key motifs as seen in one domain of the β - and γ -crystallins. A, B, C, and D represent the different β -strands of the each motif, red represents motif 1 and blue represents motif 2. The green circle represents the location of the tyrosine corner in motif 2 of the β - and γ -crystallins.

Therefore, when the tryptophans are buried in the protein's folded state, the fluorescence emission is lower than the exposed tryptophans in the unfolded state.

γD_{WT} has been cloned, expressed, purified and characterized biophysically (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). Previous kinetic and equilibrium studies have identified a folding pathway for γD_{WT} . By probing the structure of γD_{WT} , the unfolding and refolding of the protein was monitored. In work by Kosinski et al., triple tryptophan mutants of γD_{WT} were made in order to gain insight into the folding pathway of γD_{WT} , by monitoring the fluorescence of each individual tryptophan (2004). These studies showed that the C-terminal domain folded before the N-terminal domain.

In studies by Flaugh et al., mutations in the γD_{WT} interdomain interface demonstrated the importance of the interface for overall stability of the protein (2005). The C-terminal domain acts as a template for folding and stabilization of the N-terminal domain. These studies also identified an intermediate along the folding/unfolding pathway. The interface mutants were destabilized in the transition monitoring the N-terminal domain (Flaugh et al. 2005b). Thus, the populated equilibrium intermediate had likely the C-terminal domain folded and the N-terminal domain unfolded. The studies of the isolated domains in this thesis confirmed these observations.

The characterized partially folded intermediate with the N-terminal domain unfolded and the C-terminal domain folded provides insight for the folding of multi-domain proteins. Independent folding of each domain has been proposed as a mechanism of multi-domain protein folding. Although this is a viable hypothesis and many domains seem to fold independently, γD crystallin has revealed a nucleated folded process. This suggested that the N-terminal domain in the context of the full-length protein may be in a partially structured state such as a molten globule or a precursor to β -sheet formation that has yet to be described. However, the N-terminal domain of the protein can, on its own, fold into a native-like structure (Chapter 2).

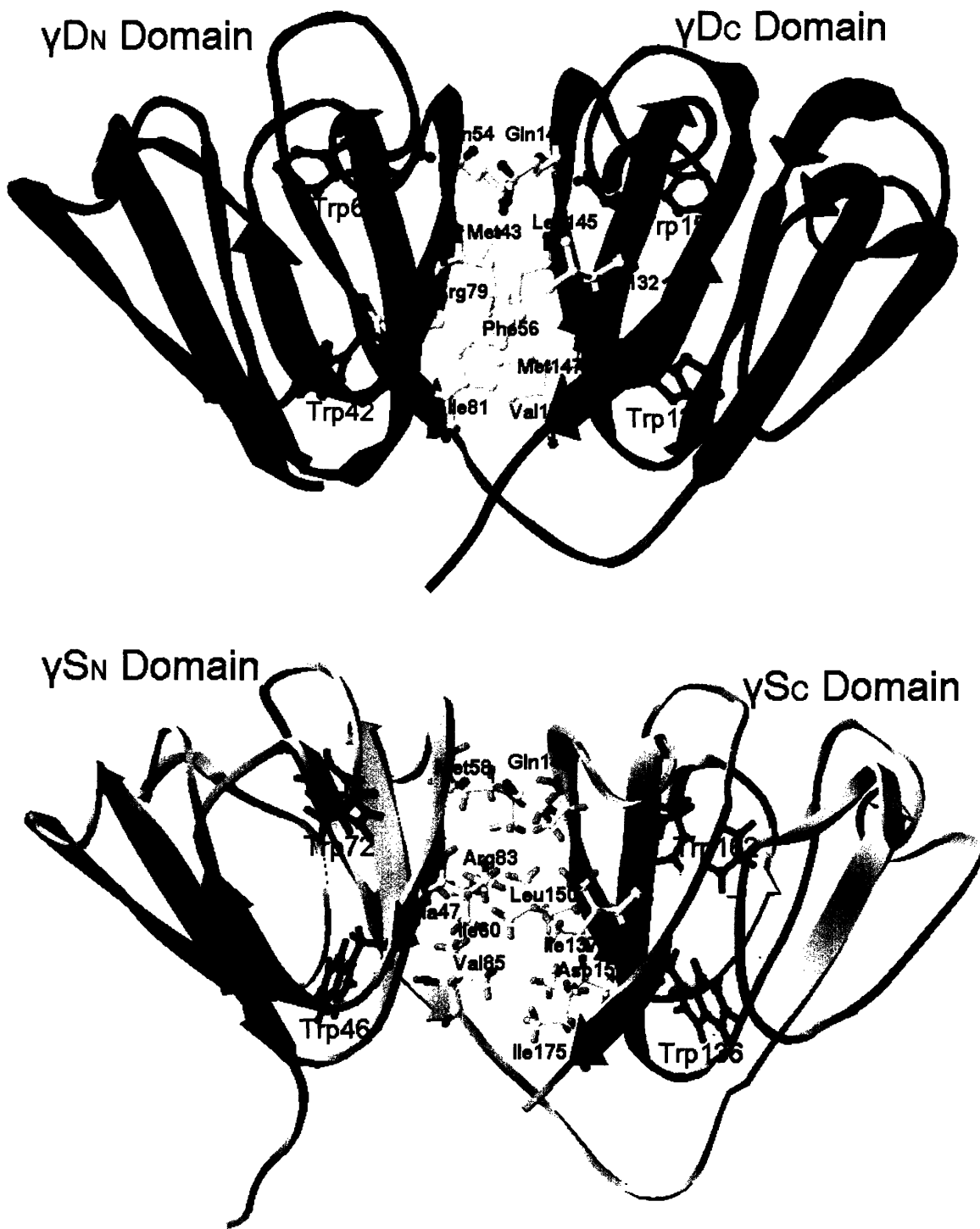


Figure 1-8. Human γ D and murine γ S high resolution structure. (PDB ID: 1HK0 and 1ZWO). The four buried conserved tryptophan residues and the interdomain interface residues are represented in ball and stick.

Recently, it has been also shown that deamidation mimetic mutations in the interface decreased this kinetic barrier to unfolding, indicating the importance of the interface in the kinetic stability of the protein (Flaugh et al. 2006). Deamidated crystallins have been detected in the insoluble fraction of lens homogenates and deamidation is thought to be one of the most abundant modifications of the crystallins (Wilmarth et al. 2006). These studies suggested a plausible model of how a post-translation modification of the crystallins by introducing a negative charge into the interface can cause destabilization of the protein. In the context of the lens, multiple species of a destabilized intermediate possibly, the N-terminal domain unfolded and the C-terminal domain folded, may associate in cataract formation or may bind α -crystallin.

K. BASIS FOR GLOBULAR PROTEIN STABILITY

Numerous weak interactions contribute to the marginal stability of folded proteins. Studies of a host of proteins *in vitro* have designated hydrogen bonding, the hydrophobic effect, the packing of the protein core, and electrostatic interactions as contributors to stability (Matthews 1993; Rose and Wolfenden 1993). Although more than 30,000 3-D structures of proteins are now solved, it is still difficult to predict particular interactions that are important for protein stability. For example, in the case of the crystallin proteins, β and γ crystallins structurally have few noticeable differences between them (Bax et al. 1990; Blundell et al. 1981). However, β B2 crystallin is a domain swapped dimer that is less stable than the monomeric γ B crystallin. By inspection of structure alone, the differences in stability and state of association are not explainable (Jaenicke and Seckler 1997). Hence, determining stability differences between proteins has been accomplished mostly by empirical methods. Individual mutations of specific amino acids have been important in determining their contribution to stability. The location of the specific interactions is also important in that the same interaction engineered in a different region of the protein would not have the same effect (Matthews 1993).

The contribution of hydrogen bonds to stability remains controversial. It was predicted that the stability of the H-bonds between polar groups is minimal due to the fact that these polar groups would H-bond with solvent in the unfolded state also. However, Fersht et al. using mutational analysis demonstrated that hydrogen bonds could contribute 0.5 to 1.8 kcal/mol with polar side chains and higher when the amino acid side chain was providing an ionic interaction (1985). Since the hydrogen bonds are intramolecular interactions as opposed to intermolecular interactions that would occur between solvent and the side chain, multiple hydrogen bonds could contribute to the overall stability of the protein. However, the surrounding environment of the intramolecular hydrogen bond, and also solvent variability could provide alternative results. In addition, these substitutions were alanine mutations which could disrupt other interactions due to the change in size and loss of atoms in the alanine mutation. Calculations of the free energy of the hydrogen bond through mutational analysis may be affected by these contributions from other forces such as reduced van der Waals interactions. In the crystallins, the symmetrical twisted β -sheet structures is thought to contribute to the stability by maximizing the hydrogen bonding between the strands (Jaenicke 1999).

The hydrophobic effect is generally accepted to be the major force of protein folding via the hydrophobic-collapse model. Mutational studies on the hydrophobic core of T4 lysozyme show a linear relationship between the decrease in stability and how much of a cavity is created by the substitution (Eriksson et al. 1992). However, there is variability in the amount of energy that each hydrophobic residue provides within the protein in the case of some non-destabilizing mutations (Eriksson et al. 1992). This discrepancy was suggested to be due to an adjustment of the protein structure when a mutation is made to prevent a destabilizing cavity (Karpusas et al. 1989; Matsumura et al. 1989). An extensive mutational study of the Arc repressor, demonstrated that several residues in the hydrophobic core of the protein were the most destabilizing (Sauer et al. 1996). In addition, although some polar and ionic residue substitutions were destabilizing, replacing these residues with non-polar residues improved the stability of the protein. Surface substitutions were more tolerant to mutations compared to the hydrophobic core (Brown and Sauer 1999).

The high packing density of globular proteins is also important for protein folding and stability. It has been extensively debated whether hydrophobic cores are fixed or tolerate more fluidity. There is strong evidence for both ideas; for example, several amino acid mutations in the core have been shown to have minor effects on stability albeit some do exhibit more destabilization than others (Lim and Sauer 1991; Matouschek et al. 1989; Matthews 1987). In contrast, packing density measurements of hydrophobic cores demonstrate well packed interiors (Richards 1977). An explanation for these different results is that the secondary structural components are arranged in the core limiting accessibility to solvent. Although mutations may not affect the stability of the protein, crystallography studies do show alterations in packing which are adjusted for by rearrangement in the core. Thus, the close packing of globular proteins contributes to their overall stability (Rose and Wolfenden 1993).

It has been difficult to assess the contribution of electrostatics in protein stability. It has been proposed that the effects of electrostatics are minimal in the stability of a protein (Akke and Forsen 1990; Sun et al. 1991a). For example, in T4 lysozyme, a basic protein, long range repulsion was tested by replacing Lys and Arg residues with Glu (Dao-pin et al. 1991; Sun et al. 1991b). These mutations did not affect the thermal stability or thermodynamic stability of the protein. Mutagenic analyses of salt bridges have also demonstrated minimal effects on stability (Erwin et al. 1990; Horovitz et al. 1990). Additionally, the dependence of stability on pH was unchanged when one charged residue was substituted for an uncharged residue (Sun et al. 1991b). The entropy needed to orient the ionic pair for a favorable interaction is expected to be the reason as to why these otherwise stabilizing interactions do not contribute more significantly to the protein's stability (Dao-pin et al. 1991). On the other hand, ionic networks have been proposed to be stabilizing perhaps by limiting the entropic cost required to obtain a favorable interaction (Nordberg Karlsson et al. 2003).

Stability assessments of the thermophilic proteins compared to mesophilic homologues have been important in attempts to determine interactions important for protein stability. Analysis of thermophilic proteins has suggested several contributions to

their stability including hydrophobics, local ionic interactions, overall surface attraction, secondary, tertiary, or quaternary structure, post-translational modifications and associations (Jaenicke 1996). In comparing thermophilic and mesophilic cold shock proteins (CSP), mutations of specific residues achieved high stability. These residues were involved in local electrostatic interactions important for the overall stability of the protein (Wunderlich et al. 2005).

Ionic networks and/or aromatic networks may be particularly important for the stability of the γ -crystallin proteins. The crystal structures of human γ D crystallin and γ S crystallin have extensive ionic networks that had been proposed previously (Salim and Zaidi 2003). Additionally, the high percentage of aromatic residues may act as a network to contribute to the overall stability observed in the γ -crystallins.

Most of the aforementioned studies have focused on thermodynamic or conformational stability which is a measure of the difference in the free energy of unfolded and native states. Another source of stability is kinetic stability which refers to the high kinetic barrier between the unfolded and native states. The basis for kinetic stability is still being investigated. In thermophilic proteins, electrostatic interactions have been suggested to be a source of high kinetic stability as well as thermodynamic stability (Jaenicke and Bohm 1998; Solis-Mendiola et al. 1998). Surface hydrophobic residues (Machius et al. 2003), addition of disulfide bonds (Mansfeld et al. 1997), and metal binding sites (Pozdnyakova et al. 2001) have demonstrated increased kinetic stability. Proteins which demonstrate high kinetic stability have common features such as β -sheets and oligomeric quaternary structures, albeit these features are not strictly required for this property (Manning and Colon 2004). For instance, α -lytic proteases and α_1 -antitrypsin both have α -helical components and are known for their high kinetic stability (Carrell and Huntington 2003; Jaswal et al. 2002). In addition, high kinetic barriers have been associated with preventing disease prone conformations. As an example, the oligomeric transition of Transthyretin from a tetramer to the amyloidogenic monomer has a high kinetic barrier (Johnson et al. 2005).

L. CATARACT AS A PROTEIN DEPOSITION DISEASE

Cataract, the leading cause of blindness in the world, is defined as an opacification of the lens. Cataracts afflict primarily older adults and diabetics (sugar cataracts). Congenital cataracts occur at a low frequency with approximately 40 cases per 100,000 births (Graw 2004). Congenital cataracts can affect individuals from birth or develop during childhood. Mature-onset cataract affects 46% of people worldwide who experience some form of blindness regardless of gender, race, or economical differences. Increasing global population and the increase in human lifespan has compounded this problem.

Analysis of cataractous lens tissue reveals an abundance of protein aggregates consisting of multiple lens protein species. Major components of these aggregates are the lens crystallin proteins (Fig. 1-3). Identification of post-translational modifications of crystallins have been performed by removing young, cataractous and age-matched lens and using tandem mass spectrometry to identify modifications. Comparative analyses on water soluble and insoluble fractions of these lenses have attempted to uncover which modifications lead to the disease state. Multiple studies have demonstrated that post-translational modifications of the crystallins included oxidation, carbamylation, UV-filter adducts, truncations, and deamidations (Hanson et al. 2000; Lampi et al. 1998; Ma et al. 1998; Robinson et al. 2006). Among the modified amino acids are methionine sulfoxides, deamidated glutamine and asparagines, and disulfide bonded cysteines. Recently, it was found that in every lens regardless of age, the insoluble crystallin fraction had higher percentages of deamidation modifications in all of the crystallins (Wilmarth et al. 2006).

Structural studies of the cataract are mostly through EM and Fourier transform analysis of removed lenses. Comparisons of EM thin sections between normal aged lenses and mature-onset cataractous lenses reveal numerous cellular abnormalities in the membranes of the lens fiber cells (Costello et al. 1993). The large cataracts appear to disrupt the membranes between lens fibers. Cataractous lens membranes within the

nuclear region were most disrupted consisting of vacuoles between membranes, multilamellar membrane aggregates, globular bodies, and highly convoluted membranes (Costello et al. 1992). Globules, vacuoles, membrane aggregates, and lens cell swelling all are observed in diseased lenses (Gilliland et al. 2001). Utilizing Fourier analysis to compare cataractous and transparent lenses demonstrated a correlated roughness with an increase in scatter (Freel et al. 2003). This may be indicative of the aggregation of the crystallin proteins.

Currently, there are several proposals as to the causes of mature-onset cataract. These models include age-related modifications of proteins such as oxidative or UV damage leading to protein unfolding, aberrant protein interactions, and/or disruptions in short range order, resulting in protein precipitation, insolubility, or aggregation (Benedek 1997). The remarkable stability of the crystallins and chaperone function in the lens is presumed to prevent these mishaps for a long period of time, i.e. 40 years. Thus, the high stability of the crystallins is accordingly a subject of much interest.

1. Mutations in Crystallin Genes Leading to Congenital Cataracts

Autosomal dominant single amino acid mutations in the crystallin genes have led to several different forms of congenital cataracts (Graw 2004; Heon et al. 1999; Santhiya et al. 2002). Mutations in human γ D crystallin include many Arg substitutions, such as R58H and R36S. Both R58H and R36S mutations enhance crystallization; as a result, R36H crystals are visible in the lens of patients (Pande et al. 2001). Contrary to the other mutations, R14C, develops disulfide bonded higher ordered aggregates progressively increasing early in life (Pande et al. 2000; Stephan et al. 1999). Another mutation P23T protein structure has been characterized further and indicates slight perturbations in structure leading to aberrant association and insolubility of the protein (Evans et al. 2004; Pande et al. 2005).

An additional congenital mutation, T5P, is found in the abundant, human γ C crystallin (Heon et al. 1999; Ren et al. 2000). Studies of the mutant protein *in vitro* demonstrated reduced stability and conformational change in the native state (Fu and

Liang 2002a; Fu and Liang 2003) Additionally, the binding of this mutant to other γ - and α -crystallins was reduced, confirming structural alterations caused by the mutation in the protein (Fu and Liang 2003).

Other congenital mutations have been identified in β and α -crystallins as well as gap junction and other structural proteins (Li et al. 2006; Vanita et al. 2006). For example, mutations in α -crystallins prevent binding of other crystallins, presumably inhibiting its chaperone function (Fu and Liang 2003).

2. *Potential Causes of Mature-Onset Cataract*

Partial protein unfolding and subsequent aggregation is an important model for the lens disease, cataract. In contrast to some congenital cataracts, a probable cause of mature onset cataract is the accumulation of modifications causing instability of the protein, leading to aggregation. For many protein deposition diseases, it has been difficult to determine the conformation of the partially unfolded aggregation prone conformer. In the protected environment of the lens, it has been difficult to assess which stresses are responsible for initiating the aggregation process. Modifications detected upon examination of the insoluble fraction of the lens could be an effect of cataract formation and not causative. Thus, it is important to study how these modifications affect proteins *in vitro*.

Recently, studies which mimic protein damage such as deamidation have shown destabilization of the protein *in vitro*. Previous studies on human γ D crystallin, demonstrated that deamidation of the interface residues led to destabilization of the protein due to introduction of a negative charge in the domain interface (Flaugh et al. 2006). Similarly, some deamidation mimetic mutations in both β B2 and β B1 crystallins have led to destabilization of these proteins. Studying the unfolding and aggregation of the crystallins *in vitro*, may provide insight into the cataract formation (Harms et al. 2004; Lampi et al. 2006, Fig. 1-9).

3. Role of α -crystallin

As the lens ages, proteasome activity decreases in the terminally differentiated fiber cells (Viteri et al. 2004). Therefore, at the high concentration of α -crystallin present in the lens, it is likely the primary mechanism in rescuing endangered crystallins. However, it acts as a passive chaperone binding partially unfolded proteins but not refolding them. This sponge-like mechanism may act as a quality control mechanism from childhood to middle age. However, α -crystallin is also susceptible to age-related modifications as with the $\beta\gamma$ -crystallins and it is found in the cataractous lens aggregate with multiple age-related modifications. One possible hypothesis has been that as the percentage of the partially unfolded proteins increases, the α -crystallin “sponge” becomes fully saturated and is integrated into the cataractous aggregate (Fig. 1-9). Accordingly, this confirms the high stability requirements as described in Chapter two and three, important for the crystallins to prevent accumulation of aggregation prone intermediates.

M. MECHANISMS OF PROTEIN AGGREGATION RELATED TO HUMAN DISEASE

Protein aggregation has been implicated in various diseases and is a major problem in the biotechnology industry and in biomedical research laboratories. Studies of aggregation pathways suggest that this reaction can occur through three different mechanisms. One mechanism is initiated by a slight conformational change in the native protein leading to a native-like, off-pathway conformer which interacts with identical conformers to initiate aggregation. The polymerization of sickle cell hemoglobin at low oxygen tension is a classic example.

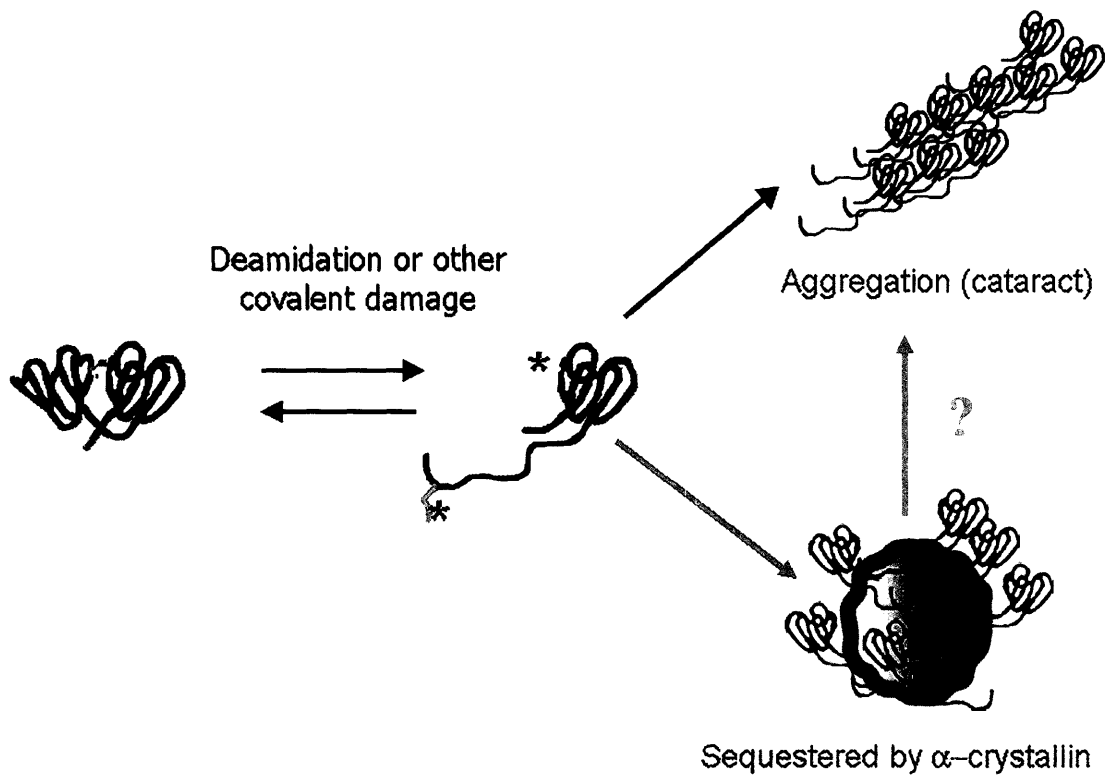


Figure 1-9. Model of cataract *in vitro*.

Secondly, partially folded or unfolded intermediates along the folding or unfolding pathway can be precursors to aggregation. This is the major source of the inclusion bodies that accumulate in *E. coli* after high-level expression of many cloned proteins (Betts and King 1998; Mitraki and King 1992). In the past, these mechanisms were thought to simply be random and non-specific interactions of exposed hydrophobic residues in these intermediates leading to off-pathway aggregation. However, there is increasing evidence that the early steps leading to aggregation may involve more specific interactions consisting of associations between similar structural conformations or particular amino acid interactions. The third and most extensively investigated mechanism is amyloidosis in which variable conformation changes lead to similar cross β -sheet structure regardless of the protein.

1. Amyloidosis

The accumulation of “cross- β ” amyloid fibers is implicated in several neurodegenerative, transmissible spongiform encephalopathies (e.g. Type 2 Diabetes), and prion related diseases (Zerovnik 2002). The final amyloid fibrillar state as well as the intermediates leading up to the amyloid fibrils have been proposed to be the cause of these diseases. Several of the amyloid diseases have been associated with specific precursor proteins. For example, the amylin is the component in the type 2 diabetes. Although amyloidogenic proteins do not have sequence similarity they contain similar structural characteristics, such as β -strands (Zerovnik 2002). Most proteins involved in amyloidosis exhibit a conformational shift that consists of β -sheets that are involved in amyloidogenesis. In addition, the specificity of these reactions is observed in the mixing experiments in which adding amyloidogenic proteins does not decrease the lag phase of the reaction. Understanding the pathway of amyloidosis is important for determining what causes the disease. Proposed models have suggested non-native conformers associate to form a “donut-like” shapes leading to protofibrils and then to extended amyloid fibrils (Caughey and Lansbury 2003). The regulated cross- β structure can be detected by binding to spectroscopic dyes, EM, and X-ray diffraction has enabled such studies.

Solid state NMR has been utilized to obtain the structural information on amyloid fibrils formed by many proteins (Tycko 2006). Using this technique provides valuable information not observed with X-ray diffraction. These studies have provided insight into the structural characteristics, such as the orientation, size of β -strand segments, and demonstrated that the cross- β structure maximizes hydrogen bond capacity (Tycko 2006). These studies have given us a detailed structural end point but have not provided the mechanistic perspective of how amyloids are formed. In addition to NMR structural studies, other studies such as isotope-edited infrared spectroscopy have been utilized to probe the structural properties and the mechanism of amyloidosis. Recent studies using isotope-edited IR spectroscopy monitored the amyloidogenic peptide of the Syrian hamster prion protein in amyloid fibrils. Modifications of the amyloid characteristic amide I (and II) bands by specific ^{13}C -labeled residues allowed for determination of the directionality and register of the Syrian hamster prion peptide (H1) (Petty et al. 2005). The wild type register peptides formed thin protofibrils followed by twisted fibril morphology versus the peptides in different register which exhibited only the thin protofibrils and did not form larger fibrillar structures (Petty et al. 2005). In addition, by utilizing isotope-editing in kinetic studies, the process of amyloid alignment was observable in a concentration dependent manner (Petty and Decatur 2005). Two mechanisms were proposed concerning the observed precise alignment and rearrangement between the peptides; one suggested a detachment/reannealing mechanism at low concentrations and the other a reptation-like mechanism at higher peptide concentrations, the latter mechanism takes longer to nucleate since it was more difficult for the peptides to get into register for nucleation (Petty et al. 2005). These two models demonstrated the multiple nucleation mechanisms in the initial steps of amyloid fibril formation.

2. Loop Sheet Insertion - The Case of the Serpins

The serine protease inhibitors, serpins, consist of a diverse set of proteins found in several organisms and even viruses (Silverman et al. 2001). The serpins include as the name implies several types of inhibitors but also proteins that participate in diverse

biological processes such as chromatin packing (Lomas and Carrell 2002). Although the serpins have diverse functions, the tertiary structure of all serpin family members is preserved. The inhibitory mechanism of the protease inhibitor serpins is described as a loop sheet insertion mechanism. Each inhibitory serpin contains a reactive loop (denoted R in the Fig. 1-10A) that binds to the active site of the protease mimicking its substrate. This binding is followed by a conformational change in the serpin whereby the reactive loop intramolecularly inserts into β -sheet A of the protein forming a β -strand. This dramatic structural shift of the reactive loop which is covalently linked to the protease, distorts the protease leaving it irreversibly inactive (Huntington et al. 2000, Fig. 1-10A). The serpins are metastable in the active form, relying on kinetic stability instead of thermodynamic stability to prevent unfolding of its structure. The insertion of the reactive loop into β -sheet A increases the serpin's stability (Gettins 2000).

Examples of inhibitory serpins include the α_1 -antitrypsin, anti-thrombin, C1 inhibitor, and plasminogen activator inhibitor 1. All of these proteins are important in the blood plasma inhibiting excessive protease activity for the circulatory and inflammation system. For example, α_1 -antitrypsin inhibits the neutrophil elastase from degrading of tissues such as the connective tissue of the lungs (Carrell and Lomas 2002). The loop sheet insertion mechanism of the serpins renders it susceptible to various diseases. For instance, certain point mutations in the reactive loop or within the β -sheet A region allows for insertion of another serpin molecule causing polymerization. One of these mutations has been extensively characterized and is prominent in people of North European descent. This mutation is a Glu > Lys substitution (called the Z variant) in one of the hinges above the β -sheet A of α_1 -antitrypsin promoting insertion of another α_1 -antitrypsin molecule instead of self-insertion (Lomas et al. 1992). Alternatively, it can form a closed loop leading to a latent conformation of α_1 -antitrypsin (Silverman et al. 2001).

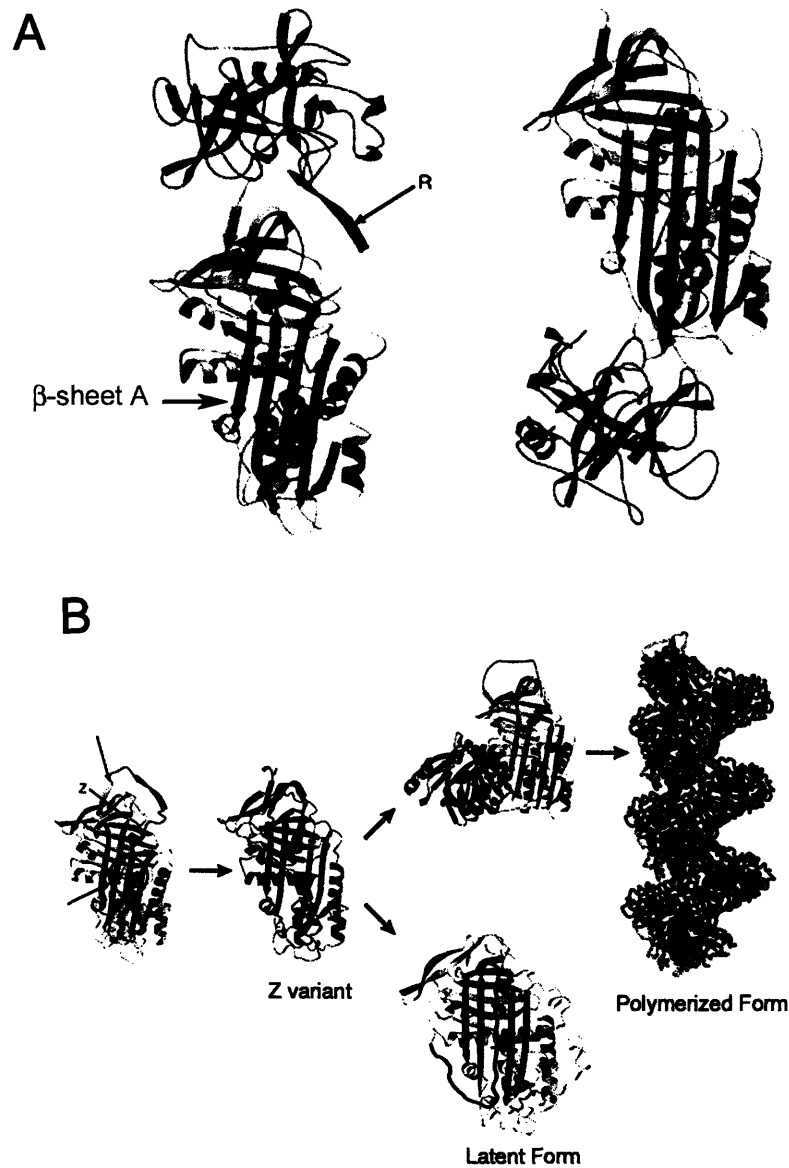


Figure 1-10. Serpin protease, α_1 -antitrypsin normal function and disease state
 (A) Inhibitory protease mechanism - Loop sheet insertion of reactive loop (R) intramolecularly into the β -sheet A (shown in green) deforms the trypsin protease (shown in blue) rendering it irreversibly inactive.
 (B) Point mutations in the α_1 -antitrypsin forms polymerized state. The Z variant mutation in the β -sheet A region allows for another molecule to insert into the region leading to a latent form or a polymerized state. (Modified from Lomas and Carrell, *Nat Rev Genet* (2002), 3: 759-768.

The extensive polymerization process leads to inclusions in liver cells and a loss of the α_1 -antitrypsin in the plasma contributing to cirrhosis and emphysema, respectively (Carrell and Lomas 1997). This polymerized form has been studied both *in vitro* and *in vivo* (Janciauskiene et al. 2002; Le et al. 1992). The α_1 -antitrypsin polymerization from the Z-variant is proposed to be a form of domain swapping mechanism (Fig 1-B). Crystallization of the Pittsburg mutant Met > Arg in the β -Sheet A region exhibits an open-ended domain swap mechanism (Bennett et al. 2006).

3. Domain Swapping – In general

There have been many examples of the domain swapping mechanism in functional protein structures as well as in protein aggregation. Domain swapping is sometimes referred to as 3-D domain swapping so as not to confuse it with protein engineered swapping of domains. In this thesis, domain swapping will refer to 3-D domain swapping. A protein is considered to have a 3-D domain swap when it maintains the same non covalent bond interactions as it does in the monomeric form but intermolecularly with another monomer to form a dimer or other oligomers (Schlunegger et al. 1997). Closed-ended domain swapping may result in an oligomeric structure, while open-ended domain swapping can lead to polymerization that may be pathologically detrimental. To date there are approximately 30 solved structures by NMR or crystallography of domain swapped proteins (Bennett et al. 2006). Some of these proteins were obtained through non-physiological conditions and their physiological relevance as domain swapped structures is unclear (Liu and Eisenberg 2002).

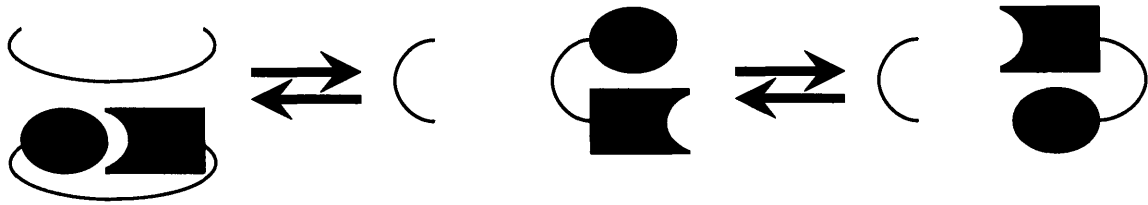
There are three different types of domain swapping that result in domain swapped oligomer structures. One type involves proteins that have both monomer and domain-swapped protein forms. The first defined domain swapped protein, Diphtheria toxin, is an example of this type. Diphtheria toxin has an active monomeric and an inactive dimeric form (Bennett et al. 1994). Little is still known about the mechanism of creating this domain swap. Mutations in the interface of *suc1*, a cell cycle protein, have shown a shift in equilibrium from monomer to dimer suggesting properties of the interface are

important in this mechanism. In addition, other structural features consist of proteins forming closed-ended linear and cyclic oligomers; open-ended and closed-ended trimers such as RNase A, which have been shown to form domain swapped closed trimers and linear trimers (Bennett et al. 2006).

Another type of domain swap mechanism involves groups of proteins which are products of divergent evolution, the sequence of the proteins are not identical but they have structural similarities in the monomeric form and domain swapped form. The β 2 crystallin and γ -crystallins are an example of this group. These proteins have ~30% sequence homology but are structurally similar. Lastly, there are intertwined oligomers that resemble domain swapped proteins but do not have an observable corresponding monomeric form. An example of this domain swap is RecA, a protein involved in DNA repair and recombination which has a N-terminal α -helical domain swap (Story et al. 1992).

In addition to domain swapped functional proteins, open-ended domain swapping is a mechanism in forming polymerized fibrillar structures. Lee and Eisenberg showed that the hamster prion protein PrP^{RDX}, can form a cross β -sheet amyloid structure (2003). This is caused by an open-ended runaway swaps polymerized to form a cross β -sheet structure. This is caused by a runaway swap in which one domain interacts with another molecule leaving one domain unsatisfied (Fig. 1-11). Alternatively, closed-ended domain swaps with all domains satisfied can form higher ordered polymerization (Bennett et al. 2006). An example of a closed-ended domain swap aggregate is formed by cystatin C, which forms amyloid domain swapped amyloid aggregates (Janowski et al. 2005).

A



B

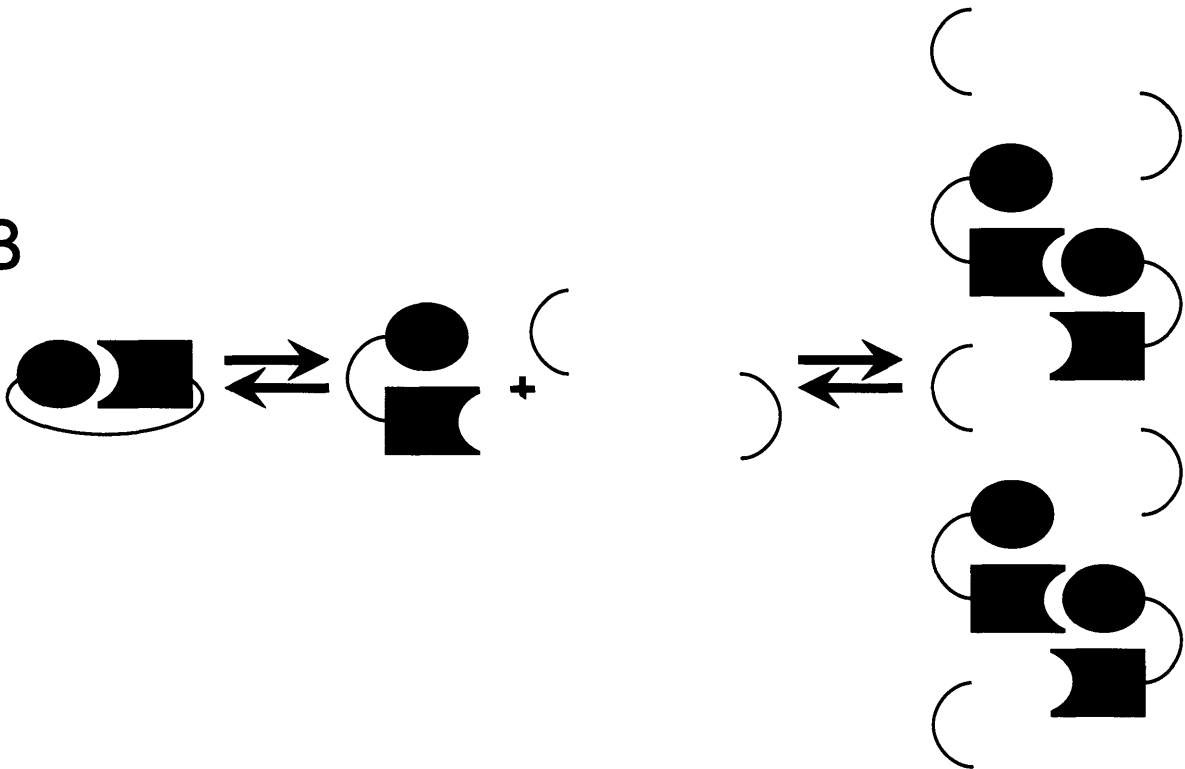


Figure 1-11. Mechanisms of domain-swapping. (A) A dimer formed by closed-ended domain swapping (B) Polymer formation by an open-ended runaway domain swap mechanism.

4. Light Chain Amyloidosis and Light Chain Deposition Disorder

The variable region of immunoglobulin chains shares a β -sandwich Greek key fold. They have the ability to dimerize through interface contacts which are stabilized through disulfide bonding and are covalently bonded to the heavy chain of the antibody. They are relatively stable thermally and the disulfide bonds contribute to the thermodynamic stability of the bonds. The immunoglobulin light chains have high sequence diversity due to the variable regions that are necessary for the function of antibodies. Given the random VDJ recombination responsible for the variability of the light chains, it is expected that a percentage of these proteins may be destabilized in the native state.

Light chain amyloidosis, a systemic disease, occurs when this amyloidogenic variable light (V_L) chain is expanded as a monoclonal light chain that has been selected for a particular antigen. Other related diseases such as the Light Chain Deposition Disease (LCDD) involve aggregation of the V_L chain. In contrast to Light Chain Amyloidosis, LCDD is a non-amyloidogenic aggregate. The Greek Key nature of the V_L chain has proposed domain swapping as a model in the aggregation of these chains (Wetzel 1997). There is heterogeneity in the kinds of aggregates that can be formed with these same proteins.

In Chapter four of this thesis, the γD crystallin aggregation reaction is investigated as a model of cataract *in vitro*. Domain swapping has been proposed for many aggregation mechanisms. Characterization of γD and γS domain-exchanged chimeras are consistent with a domain swap model for γD crystallin aggregation.

N. THE BIOLOGICAL CONTEXT OF THIS THESIS

All of the experiments in the following chapters consist of the biophysical characterizations of purified crystallins. However, reflecting on my background as a cell biologist and geneticist the motivating questions in this thesis were the evolution of these

proteins and their biological roles. All known crystallins in modern day vertebrates have duplicated domains. One question that has arisen from this observation is the basis for gene duplication in these organisms? Additionally, is this gene duplication important for the crystallin biophysical characteristics such as protein stability? A second question was why numerous crystallins are needed and are differentially expressed in different regions of the lens.

These biophysical properties of the crystallins may give us insight into understanding the underlying molecular mechanisms that may lead to cataract as well as provide more insight into the evolution of the crystallins, leading to insight into lens evolution in general.

CHAPTER TWO:

**FOLDING AND STABILITY OF THE ISOLATED GREEK KEY DOMAINS OF
THE LONG-LIVED HUMAN LENS GAMMA D CRYSTALLIN.**

A. INTRODUCTION

Crystallins are the major proteins in the elongated fiber cells of vertebrate eye lenses, present at concentrations of 200-450 mg/ml (Fagerholm et al. 1981a; Siezen et al. 1988; Slingsby and Clout 1999). The crystallins are responsible both for the transparency and high refractive index of the lens (Delaye and Tardieu 1983; Fernald and Wright 1983). The β - and γ -crystallins are thought to be primarily structural proteins while α -crystallin possesses an additional chaperone activity. Terminally differentiated fiber cells in the lens are enucleated and devoid of all other organelles, and are unable to degrade damaged crystallins or to synthesize new ones (Oyster 1999). Thus, the crystallins must remain soluble for decades despite high concentrations of protein, continual UV exposure and potential oxidative stress. This is particularly true of the crystallins in the central lens nucleus, which are synthesized *in utero* (Aarts et al. 1989; Harding JJ, Crabbe MJC 1984; Lampi et al. 2002)

Cataracts interfering with light transmission represent an aggregated and insoluble form of the crystallins. Mature-onset cataract affects more than 40% of people who experience some form of blindness, regardless of gender, race, or economic status. The protein composition of the insoluble fraction of aged and cataractous lenses includes the α , β , and γ lens crystallins, with many carrying a variety of oxidative modifications and truncations (Hanson et al. 2000; Lampi et al. 1998; Lapko et al. 2005; Searle et al. 2005; Wilmarth et al. 2006).

Models for the causes of mature-onset cataract include oxidative or UV-induced protein damage leading to protein unfolding, aberrant activation of fiber proteases, and saturation of α -crystallin chaperone function. Any or all of these could result in protein precipitation, insolubility, or aggregation. The remarkable stability of the crystallins and chaperone function in the lens is presumably to prevent the onset of these aggregated states. The biochemical basis of the very high stability of the crystallins, and the nature of the misfolded, modified, or aggregated states, are thus of considerable importance in understanding the etiology of cataract.

All β and γ -crystallins have two homologous Greek key domains. They are believed to have evolved by gene duplication and fusion from an ancestral single domain $\beta\gamma$ -crystallin. A candidate for this precursor has recently been found in the urochordate sea squirt, *Ciona intestinalis*. Gene and structural alignment comparisons suggest that β and γ -crystallins evolved from this common ancestral protein (Shimeld et al. 2005).

Each domain consists of two anti-parallel β -sheet Greek Key motifs, which are intertwined interacting with each another. The two highly symmetrical domains interact through a hydrophobic interface and are connected by an interdomain linker. The γ -crystallins are monomeric, whereas the β -crystallins can form multimers (Bateman et al. 2003; Lampi et al. 2001; Werten et al. 1999). Differences between these two groups are attributed to N and/or C-terminal extensions in the β -crystallins, an extended linker in β B2 crystallins such that the composition of this linker is important for oligomerization, and intermolecular and intramolecular associations (Bateman et al. 2001; Mayr et al. 1994; Norledge et al. 1997; Trinkl et al. 1994).

The stability of the β and γ -crystallin proteins have been suggested to be due to the complex topology of the double Greek Key (MacDonald et al. 2005), in which strand d of the first motif is paired with strand c of the second motif (Fig. 2-1). Although the majority of β and γ -crystallins have these features, there are differences in the intrinsic stabilities among the crystallins and their different domains.

An additional distinctive feature of the vertebrate $\beta\gamma$ -crystallin family is the interface between the Greek key domains. In human γ D-crystallin, this is composed of a hydrophobic patch of six residues, three from each domain, plus a pair of glutamines shielding the hydrophobic domain from solvent, and an arginine and methionine at the base of the interface near the linker (Basak et al. 2003). These residues are highly conserved among vertebrate crystallins. The contribution of these residues to overall stability has been assessed by characterizing proteins with alanine substitutions of the interface residues. Substitutions of both the hydrophobic and the polar residues of the

interface significantly destabilized the native protein (Flaugh et al. 2005a; Flaugh et al. 2005b). Substitutions in the interface of bovine γ B crystallin (B γ B-Crys) have also been found to be destabilizing (Palme et al. 1997). Liu and Liang have shown that polar substitutions of hydrophobic residues in the β -strands of the domain interfaces of human β B2 crystallins significantly destabilized the protein (2006).

Human γ D crystallin and human γ S crystallin are two of the most abundant proteins in the human lens. These proteins share 69% sequence similarity and 50% sequence identity and are ~21 kDa in size. Human γ D crystallin is found at highest levels in the central nucleus, and is primarily synthesized *in utero*. Thus the long term solubility and stability of this protein is particularly important for maintaining lens transparency. Families carrying single amino acid substitutions in this protein exhibit juvenile-onset cataracts (Heon et al. 1999; Pande et al. 2000; Santhiya et al. 2002; Smith et al. 2000; Stephan et al. 1999). One of these mutations P23T reduces the stability and solubility of the mutant protein *in vitro* (Evans et al. 2004; Pande et al. 2005). On the other hand, γ S crystallin is more prevalent in the outer regions of the lens, primarily the cortex, which continues to grow throughout life (Bron et al. 2000; Cook et al. 1994; Wistow et al. 2002). Both proteins synthesis and turnover is thought to occur in these regions.

The γ D_{WT} crystal structure has been solved as well as the C-terminal domain of γ S_{WT} in isolation. The murine γ S crystallin structure has been resolved by NMR methods. There are few noticeable differences between the two proteins (Fig. 2-1) though γ S crystallin has a four amino acid N-terminal extension and possibly two additional amino acids in its interdomain linker. Crystal structure and modeling studies have shown that both γ D crystallin domains have high structural similarity (Basak et al. 2003; Blundell et al. 1981). The NMR solved structure of murine γ S crystallin also demonstrated high structural similarity among domains (Wu et al. 2005). Therefore, both γ -crystallins have similar structures and their individual domains display high structural similarity between each other.

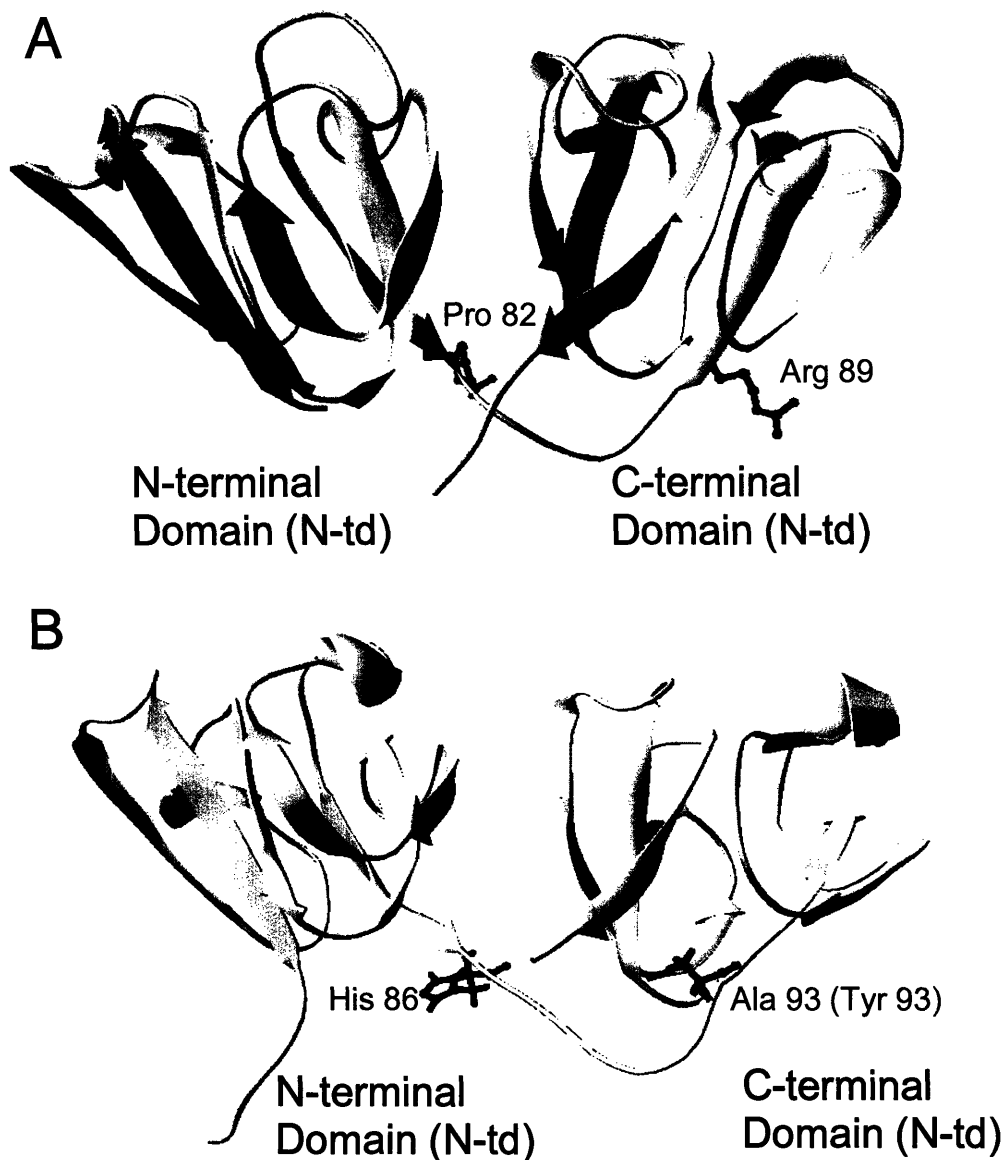


Figure 2-1. Crystal structure of $\gamma_{D_{WT}}$ and NMR structure of $\gamma_{S_{WT}}$ (mouse NMR). Both full-length proteins are ~20 kDa in size. (A) A ribbon diagram of the $\gamma_{D_{WT}}$ X-Ray crystal structure (Basak et al. 2003) (PDB ID: 1HK0). The isolated γ_{D_N} protein ends at Pro82 (highlighted in gray) not including the short interdomain linker. The isolated γ_{D_C} protein begins at Arg 89 (highlighted in gray) at the beginning of the β -sheet. (B) A ribbon diagram representing the NMR structure of the murine γ_S crystallin protein (Wu et al. 2005) (PDB ID: 1ZWO). The His 86 position is the same amino acid in the human sequence while that Ala 93 position is replaced with Tyr in the human sequence. Sequence alignment between human and murine γ_S crystallin shows 89% identity and 96% similarity (bl2seq, Blosum 62 matrix).

Human γ D-crystallin (γ D_{WT}) has been cloned, expressed, purified and characterized with respect to its folding and unfolding *in vitro* (Kosinski-Collins and King 2003; Kosinski-Collins et al. 2004). Similarly, the unfolding and refolding of human γ S-crystallin (γ S_{WT}) *in vitro* has also been characterized ((Wenk et al. 2000).

The unfolding and refolding of γ D_{WT} is a three state process. Kinetic and equilibrium studies have identified a major partially folded intermediate in its unfolding and refolding pathway, *in vitro*. This species has its C-terminal domain folded and its N-terminal domain unfolded or at least disordered. From the three state melting transitions, it was clear that the N-terminal domain was much less stable than the C-terminal domain. Further information on the interaction between the domains was obtained from site-specific mutations of the residues forming the domain interface. Not surprisingly substitutions of the N-terminal residues contributing to the interface destabilized the N-terminus. However, unexpectedly, substitutions of the C-terminal residues in the interface had little effect on the C-terminus itself, but also destabilized the N-terminal domain (Flaugh et al. 2005b). This suggested that C-terminal domain provided a template for the folding of the N-terminal domain. This raised the possibility that the N-terminal domain of γ D_{WT} could not fold independently.

Recently, it has been also shown that Gln > Glu mutations mimicking deamidation, in the γ D_{WT} interface decreased γ D stability, indicating the importance of the interface in the unfolding barrier (Flaugh et al. 2006). Deamidation in the interface of the dimeric β B2-crystallin also destabilizes the intact protein, supporting an important role for the domain interface (Lampi et al. 2006).

Wenk et al. showed that the unfolding and refolding of bovine and γ S_{WT} followed a two state transition, implying that the stability of the two domains was similar (2000). The isolated domains from these studies suggested that the γ S N-terminal domain and C-terminal domain had slightly different stabilities suggesting that the domain interface may not be important for the stability of the protein and that this protein folds not sequentially like γ D_{WT} but more cooperatively. Other studies have shown that the

isolated domains of bovine γ Bcrystallin and β B2 crystallin are not as stable as the full-length protein (Mayr et al. 1997; Wieligmann et al. 1999).

To explore more deeply the question of the contribution of domain interactions to overall protein stability, we have prepared and investigated the properties of the isolated domains of γ D_{WT}. We have included in these experiments the analogous isolated N- and C-terminal domains of γ S_{WT}. Efforts to crystallize the isolated N-terminal domain of γ S crystallin have been unsuccessful; raising the possibility that its N-terminus in isolation was also not in the native fold. The studies below attempt to address the question of the importance of the domain interface in the evolution of the two domain vertebrate crystallins.

B. MATERIALS AND METHODS

1. Preparation of Constructs for Isolated Domains and γ S_{WT}

A previously subcloned vector of γ D_{WT} as described in (Kosinski-Collins and King 2003) was used to prepare γ D_N. This vector includes an N-terminal His-tag for purification with Ni-NTA affinity chromatography (Qiagen). PCR primers (IDT-DNA) were designed to introduce a stop codon (QuikChange Site-Directed mutagenesis, Stratagene) at position 83 to create the γ D_N consisting of residues G1- P82. The rest of the single domains were prepared using PCR primers (IDT-DNA) designed at appropriate positions in the sequence. The resulting PCR products were subcloned into the pQE-1 vector using the blunt-ended PvuII (adaptable for other blunt-ended restriction enzymes) and the HindIII enzyme restriction sites. All vectors included a N-terminal His-tag, MKHHHHHHQA, to aid in purification.

γ D_C consisted of residues R89 – S174 (based on numbering in PDB:1HK0). The γ S_N consisted of residues S1 - H86 and γ S_C consisted of residues Y93 - E177. The γ D_{WT} template was a generous gift from S. Helber at Commonwealth Biotechnologies, Inc.

The vectors were all sequenced to confirm and to ensure no erroneous mutations, additions, or deletions occurred in the sequences (MGH DNA Core, Cambridge, MA). All protein sequences are included in the appendix.

2. *Expression and Purification of Proteins*

Recombinant full-length and variant proteins were prepared as described in (Kosinski-Collins et al. 2004). Briefly, all aforementioned vectors were transformed into *E. coli* M15[pRep4] cells (Qiagen), utilized for tightly regulated protein expression. The cells were lysed by conventional methods and purified by Ni-NTA resin (Qiagen) affinity chromatography using a Pharmacia FPLC apparatus. The purity and size of each protein was confirmed by SDS-PAGE. The identities of $\gamma_{D_{WT}}$, γ_{D_N} and γ_{D_C} were additionally confirmed by Mass spectrometry (CCR Biopolymers Laboratory, MIT). This purification protocol produced proteins with a purity of >90%.

Protein concentrations were determined by unfolding of proteins in 5.5 M GuHCl and measuring absorbance at 280 nm using their respective protein extinction coefficients; $\gamma_{D_{WT}}$, γ_{D_N} , γ_{D_C} , $\gamma_{S_{WT}}$, γ_{S_N} and γ_{S_C} , 41,040 $\text{cm}^{-1}\text{M}^{-1}$, 20,580 $\text{cm}^{-1}\text{M}^{-1}$, 21,555 $\text{cm}^{-1}\text{M}^{-1}$, 41,040 $\text{cm}^{-1}\text{M}^{-1}$, 21,860 $\text{cm}^{-1}\text{M}^{-1}$, 19,180 $\text{cm}^{-1}\text{M}^{-1}$, respectively.

3. *Analytical Size Exclusion Chromatography*

All samples were prepared by diluting to a final protein concentration of 80 $\mu\text{g}/\text{ml}$ in 10 mM Ammonium Acetate buffer, pH 7.0. The column was equilibrated with 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer. The samples were loaded onto the Superdex™200 10/300 GL (Pharmacia Biotech) column using a FPLC apparatus (Pharmacia). Molecular weight standards were utilized to determine relative elution times for various protein sizes. Overlays of each spectrum were made using the Unicorn program and subsequent analysis in Excel (Microsoft). Peak fractions were collected and SDS-PAGE analysis confirmed the presence of each protein.

4. *Circular Dichroism*

All experiments were performed on the Aviv Model 202 CD spectrometer with an internal Peltier thermoelectric controller used to maintain constant 37°C temperature (Lakewood, NJ). Protein samples were prepared at a 100 µg/ml protein concentration in a degassed 10 mM sodium phosphate buffer (no EDTA or DTT to prevent absorption at lower wavelengths) and equilibrated overnight at 37°C. Each sample was placed in a 1mm quartz cuvette (Starna, Inc.), allowed to equilibrate in the CD spectrometer for 1 minute and data recorded in the 260 – 195 nm wavelength range, with each wavelength averaged over a 10 second period. The CD spectrum of the buffer was subtracted from all spectra and the molar ellipticity was subsequently calculated. Deconvolution of the CD spectrum was performed using CDPro suite software package consisting of the CONTILL, CDSSTR, and SELCON3 programs (<http://lamar.colostate.edu/~sreeram/CDPro/>, Sreerama and Woody 2000; Sreerama et al. 2000). The IBasis 1 parameter (Johnson 1999) was chosen since it analyzed a larger range for the recorded data, had a large reference set (29 proteins) and had a lower RMSD and NRMSD for most of the data. The results from each program were averaged to obtain the overall secondary structure percentages.

5. *Thermal Denaturation*

All experiments were performed on the Aviv Model 202 CD spectrometer with an internal Peltier Thermoelectric controller (Lakewood, NJ). A quartz four sided screw top cuvette with a bandwidth of 4 mm was used to prevent loss of sample due to evaporation as the temperature increased. Samples were prepared at a concentration of 100 µg/ml in a degassed 10mM phosphate buffer, pH 7.0. Samples were equilibrated for 1 minute for each 1°C increase in temperature and all data points were averaged over a 3 second period. Specifically, decrease in the β-sheet secondary structure minimum at 218 nm versus increase in temperature was monitored. Buffer was subtracted for each data point. Due to aggregation of the proteins at high temperatures, fraction native of each protein was calculated by the following equation.

$$F_N = (y - y_U) / (y_N - y_U)$$

Where F_N = Fraction Native, y = Ellipticity at 218 nm, y_U = the unfolded/aggregation baseline, y_N = the native baseline. All experiments were repeated three times, calculating averages and standard deviation.

6. *Fluorescence Spectroscopy*

Fluorescence emission spectra were taken using a Hitachi F-4500 fluorimeter equipped with a temperature control circulating water bath to maintain 37°C. The fluorimeter parameters were a bandpass of 10 nm for excitation and emission monochromators, scan speed 60 nm/min, and a 2 s response time. All proteins were analyzed at a concentration of 10 µg/ml in 100 mM phosphate, 1 mM EDTA, and 5 mM DTT buffer, pH 7.0 and 5.5M GuHCl for the unfolded protein samples. The protein samples were excited at 295 nm and the emission fluorescence was recorded from a wavelength range of 310 – 400 nm.

7. *Equilibrium Unfolding and Refolding*

Equilibrium unfolding samples were diluted to a protein concentration of 10 µg/ml with increasing concentrations of GuHCl (0M – 5.5M) in 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer (Guanidine Hydrochloride solution, 8 M (GuHCl), Sigma®). In equilibrium unfolding experiments, all samples were equilibrated for 24 hours in order to reach equilibrium. For equilibrium refolding experiments, a 10X protein solution was unfolded at 5.5 M GuHCl for 5 hours. The unfolded protein was then diluted 10-fold into various concentrations of GuHCl (0 M-5.5 M) giving a lowest GuHCl concentration of .55 M GuHCl. These samples were also allowed to equilibrate for 24 hours. Exact GuHCl concentrations of each sample were determined by refractometer readings. Emission spectrum was recorded from wavelength range 310 – 400 nm. All spectra were corrected for buffer. Equilibrium unfolding/refolding curves were fit to a two state model according to the methods of (Greene and Pace 1974), or a three state model according to the methods of (Clark et al.

1993). Calculations of thermodynamic parameters were performed on 360 nm emission data and 360/320 nm emission ratio data using Kaleidagraph software version 4.0 (Synergy Software). Both analyses were comparable and within standard error of one another. 360/320 nm is shown for visual clarity of equilibrium transitions. Single wavelength 360 nm data was used to calculate m and ΔG values. Each experiment was repeated three times to determine averages and standard deviation parameters.

8. Productive Refolding Kinetics

Productive refolding kinetics experiments were performed by first unfolding 10X protein in 5.5 M GuHCl for 5 hours to guarantee complete unfolding of the protein. The 10X unfolded protein sample was diluted 10-fold into agitated 100 mM phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer using the injection port system with a dead time of ~1 s. Temperature was maintained at 18°C utilizing the circulating water bath feature. The samples were excited by 295 nm wavelength with 10nm bandpass. Emission at 350 nm wavelength with 10 nm bandpass was recorded over time. Unfolded and native control spectra were recorded at the beginning of the experiment and at the end of the experiment. The refolding kinetic data was analyzed using different exponential model equations to fit by (Fersht 1999) and residual distribution to determine the best fit. The data was analyzed by fitting it to equations depicting two state, three state, or four state models by Kaleidagraph 4.0 software. All experiments were performed three times to calculate average kinetic rates and standard deviation. All refolding data is depicted as normalized fluorescence data for comparison.

C. RESULTS

1. Protein Purification and Characterization

The isolated domains of γD_{WT} were constructed by examining the crystal structure and including the sequences of the protein without the linker. The γD N-terminal domain (γD_N) construct included residues Gly 1 - Pro 82 at the end of the β -strand before the

linker. The γD_N construct was created by introducing a stop codon at residue 83 (His 83 stop) into the γD_{WT} sequence. The γD C-terminal domain (γD_C) construct was created by cloning into the pQE1 vector (Qiagen), sequences corresponding to Arg 89 after the linker and at the beginning of the β -strand, continuing to Ser 174 at the end of the protein (Fig. 2-2).

At the time of creating the γS isolated domain constructs, there was no structure available of the γS N-terminal domain or linker. Thus, it was difficult to predict what sequences should be included in the constructs. The γS N-terminal domain (γS_N) construct was created by introducing a stop codon at residue 87 (Leu 87 stop) into the γS_{WT} sequence. Recently the NMR structure of the murine γS predicted a linker region of γS_{WT} (residues 85-93) according to Entrez protein database (# NP_060011). Therefore, γS_N starts with Ser 1 and includes His 86, one residue within the proposed linker. The γS C-terminal domain (γS_C) was created by cloning into the pQE1 vector, sequences corresponding to Tyr 93 to Glu 177 at the end of the protein (Fig. 2-2). The solved γS C-terminal domain crystal structure also began with Tyr 93 in the C-terminal domain (Purkiss et al. 2002).

All full-length and isolated domain proteins were expressed by inducing *E. coli* cell culture with IPTG and incubating at 37°C for several hours. The full-length recombinant crystallins were soluble, accumulating in the cell supernatant. The isolated domains also behaved as soluble subunits, accumulating in the cell supernatants. The isolated domain recombinant crystallins were purified using Ni-NTA affinity chromatography, following the same protocols used for the full-length γD and γS crystallins. The elution of the isolated domains was similar to the full-length protein elution during column purification. γS_{WT} and its isolated domains eluted from the Ni-NTA column with lower concentrations of imidazole than γD_{WT} and its isolated domains. In general, the isolated domains behaved similarly to the full-length crystallins during protein expression and purification.

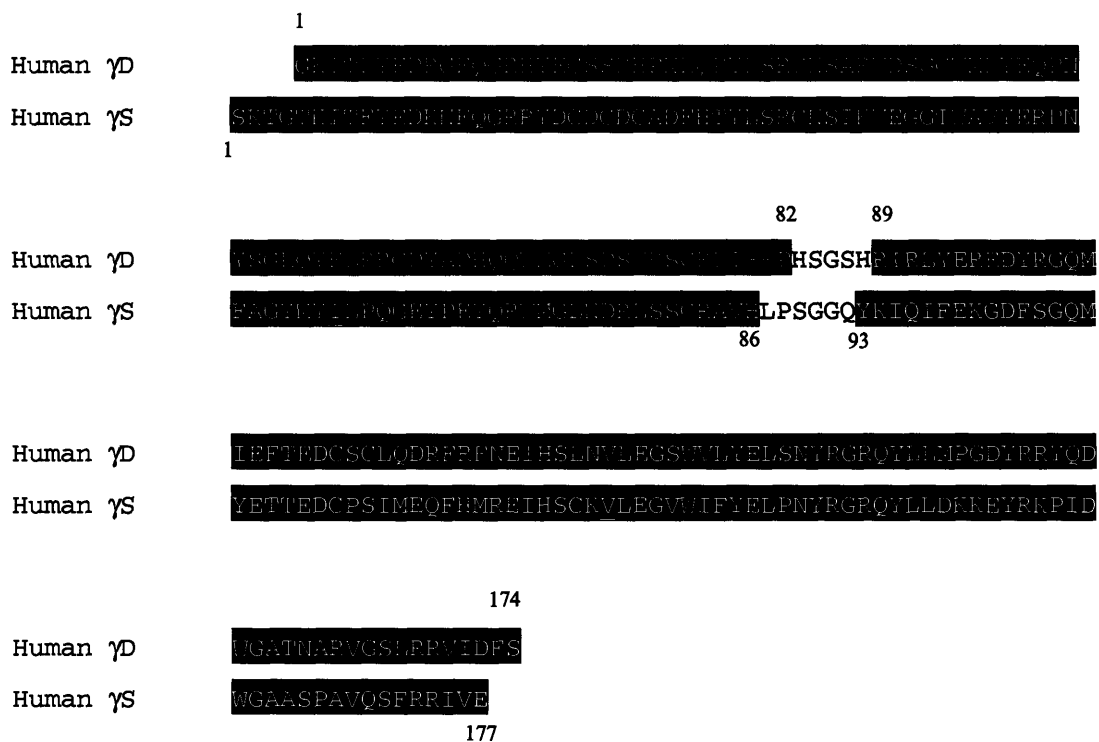


Figure 2-2. Protein sequence alignment of γ D_{WT} and γ S_{WT}. The regions of the protein included in each domain are highlighted, γ D_N G1 – P82 (blue), γ D_C R89 – S174 (red), γ S_N S1 – H86 (green), γ S_C Y93 – E177 (gray). Upper numbers represent the residues in γ D_{WT}, lower number represent the residues in γ S_{WT}.

Analysis of His-tag γD_{WT} revealed no difference in the kinetic, equilibrium, and secondary structural characteristics compared to recombinant protein without His-tag (Kosinski-Collins and King 2003). In addition, analysis of His-tag γS_{WT} compared to no His-tag recombinant γS_{WT} detected no difference in the secondary structure as analyzed by CD spectroscopy or in kinetic and equilibrium data as analyzed by fluorescence spectroscopy (Kosinski-Collins et al. unpublished results).

2. Analytical Size Exclusion Chromatography

Previous studies with isolated domains of $\beta B2$ N-terminal domain showed the possibility of dimerization occurring between isolated domains (Wieligmann et al. 1999). In addition, the microbial single domain Spherulin 3a crystallin forms dimers at physiological concentrations (Kretschmar et al. 1999b). In order to determine if the isolated domain dimerized under experimental conditions, analytical size exclusion chromatography (SEC) was utilized. Protein was loaded onto the column at a concentration of 80 $\mu\text{g/ml}$. Molecular weight standards elution volumes were determined to estimate elution of different sized proteins. All of the isolated domain proteins eluted after the 13.7 kDa protein standard (Ribonuclease A) indicating that all proteins were in the monomer form (Fig. 2-3). Moreover, the isolated domains did not overlay with the wild type proteins confirming that the isolated domains did not form stable dimers in appreciable amounts.

γD_{WT} protein eluted from the column at peak volume of 17.86 ml while γS_{WT} eluted at 17.4 ml. This result is expected since γS_{WT} is 1 kDa larger than γD_{WT} . The γD_C isolated domain eluted at a peak volume of 18.45 ml and the γD_N isolated domain at a peak volume of 18.84 ml confirming γD_C is ~ 1 kDa larger than γD_N (Fig. 2-3A). The isolated domains of γS eluted at similar volumes compared to γD isolated domains (Fig. 2-3B). Eluted peak volumes of γS_N and γS_C were 18.54 ml and 18.04 ml, respectively.

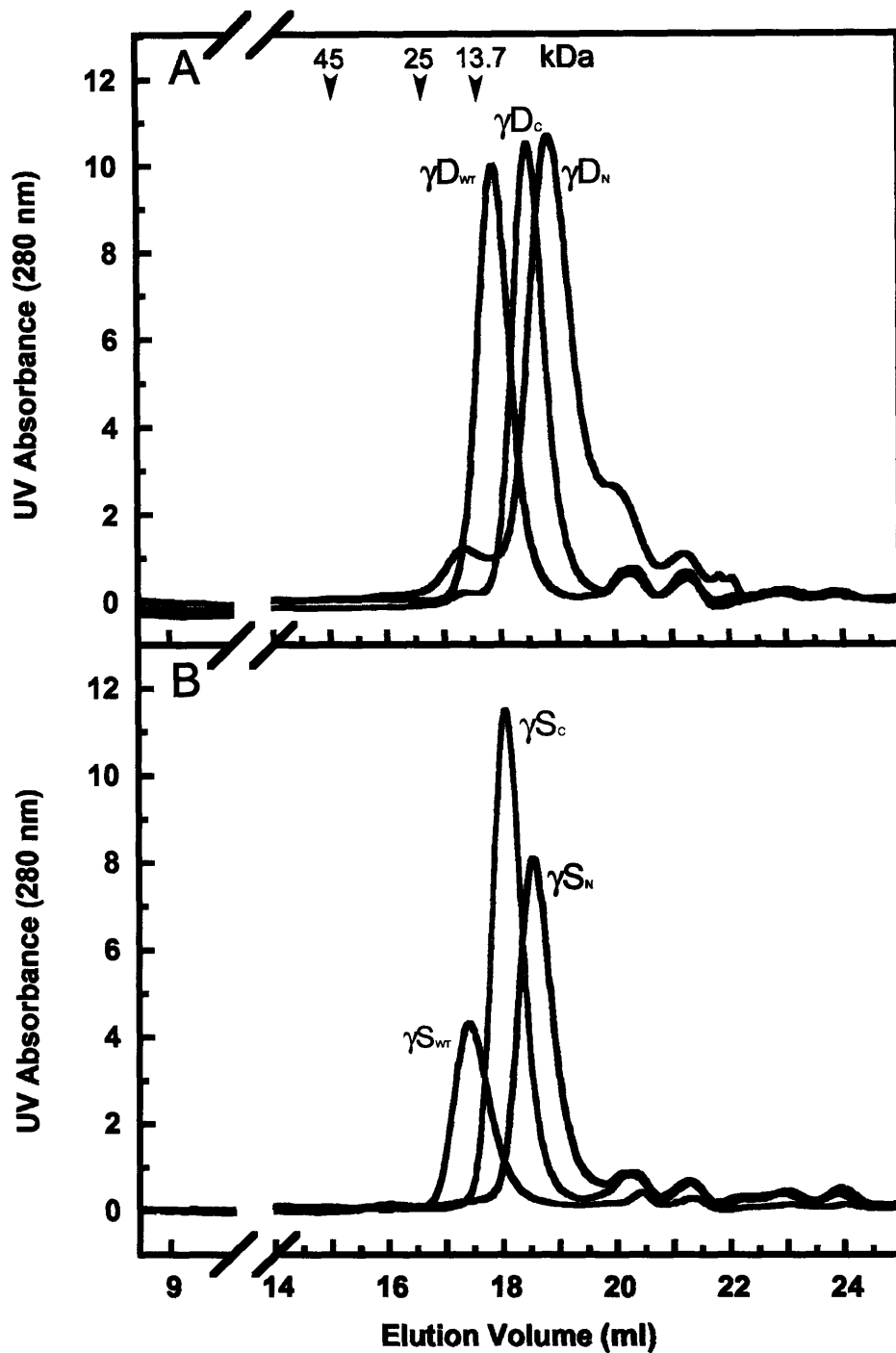


Figure 2-3. Analytical Size Exclusion Chromatography profiles of the isolated domains and wild type proteins. All samples were loaded onto a Superdex™200 10/300 GL column at a protein concentration of 80 $\mu\text{g/ml}$. The molecular weight standards were Ovalbumin (45 kDa), Chymotrypsinogen A (25 kDa), Ribonuclease A (13.7 kDa) and are noted by arrow. (A) γD_{WT} (black), γD_N (dark blue), and γD_C (red) (B) γS_{WT} (light blue), γS_N (green), and γS_C (orange).

3. Circular Dichroism (CD) Spectroscopy

To determine if the isolated domains were folded into native-like conformations, we examined them by CD and fluorescence spectroscopy. The secondary structure of the isolated domains and the full-length proteins were analyzed by Far-UV CD spectroscopy at 37°C (Fig. 2-4). CD spectrum of the complete γD_{WT} protein demonstrated the characteristic β -sheet ellipticity minimum at 218 nm. The spectrum of γD_C domain also indicated the characteristic β -sheet structure as well and was not distinguishable from γD_{WT} (Fig. 2-4A). However, the γD_N spectrum exhibited a possible increase in random-coil structure as seen by an increase in negative molar ellipticity at 204 nm. These CD spectra were further analyzed by deconvolution software, CDPPro Suite, to determine quantitative percentages of secondary structure (Sreerama and Woody 2000). Deconvolution of the CD spectra agreed with qualitative observations that the γD_{WT} and γD_C domain β -sheet structure were not distinguishable (Table 2-1). γD_{WT} analysis showed a ~40% β -sheet, ~6% α -helical, mostly 3¹⁰ α -helical structure (consistent with 3D structure), ~21% turn, and ~31% unordered. γD_C deconvolution was similar to γD_{WT} with ~41% β -sheet, ~5% α -helical, ~22% turn, and ~32% unordered. The differences in percentages are probably within experimental error of one another. However, the CD deconvolution of γD_N domain suggested it had a decrease in β -sheet (~30%) and an increase in α -helix (~11%), turns (~24%) and unordered (~35%) secondary structure (Table 2-1).

The γS_{WT} also exhibited the characteristic β -sheet structure with a minimum at 218 nm and its isolated γS_C domain had a similar spectrum to full-length γS_{WT} (Fig. 2-4B). The γS_N isolated domain spectrum showed the largest difference when compared to the γS_{WT} CD spectrum. The deconvolution of γS_N yielded ~29% β -sheet, ~4% α -helical, ~25% turns, and ~40% unordered compared to ~33% β -sheet, ~6% α -helical, mostly 3¹⁰ α -helical structure (consistent with 3D structure), ~24% turns, ~36% unordered for γS_{WT} . Again, γS_C deconvolution was similar compared to γS_{WT} with ~32% β -sheet, ~9% α -helical, ~26% turns, and ~34% unordered (Table 2-1). Secondary structure similarity

between $\gamma_{\text{S}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{C}}}$ is consistent with atomic structure similarity observed between the murine NMR $\gamma_{\text{S}_{\text{WT}}}$ structure and the crystal human $\gamma_{\text{S}_{\text{C}}}$ structure (Purkiss et al. 2002; Wu et al. 2005).

Crystallin proteins consist of Greek Key anti-parallel β -sheets, a topology that contains a considerable amount of twisted β -sheets (Bax et al. 1990; Blundell et al. 1981). It has been previously observed that the deconvolution of CD spectra of these proteins are difficult to assess due to the fact that the twisted anti-parallel β -sheets have similar optical dispersions as unordered peptides as well as the lack of reference sets for these proteins (Sreerama and Woody 2003). This is supported by the high percentage of unordered structure in our deconvolution analysis. Nonetheless, we were able to ascertain quantitative overall relatively small differences in $\gamma_{\text{D}_{\text{N}}}$ and $\gamma_{\text{S}_{\text{N}}}$ secondary structure while both C-terminal domain secondary structures remained similar to their respective full-length proteins.

4. Fluorescence Spectroscopy of the Purified Proteins

Fluorescence spectroscopy was utilized to monitor the tertiary structure of both crystallins. $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$ both have 4 conserved tryptophans, 2 buried within the hydrophobic core of each domain. The tryptophans are at position 42 and 68 in the $\gamma_{\text{D}_{\text{N}}}$ (46 and 72 in $\gamma_{\text{S}_{\text{N}}}$) and 130 and 156 in the $\gamma_{\text{D}_{\text{C}}}$ (136 and 162 in $\gamma_{\text{S}_{\text{C}}}$). The fluorescence of these tryptophans are highly quenched in the folded state (Chen et al. 2006; Kosinski-Collins et al. 2004). Trps 68 and 156 are quenched through a charge transfer to the polypeptide chain backbone, while Trps 42 and 130 undergo an energy transfer mechanism to Trps 68 and 156, respectively (Chen et al. 2006). $\gamma_{\text{S}_{\text{WT}}}$ may be quenched in the native state by a similar mechanism. As a result, tryptophan fluorescence is a sensitive reporter of the native-like state of these proteins.

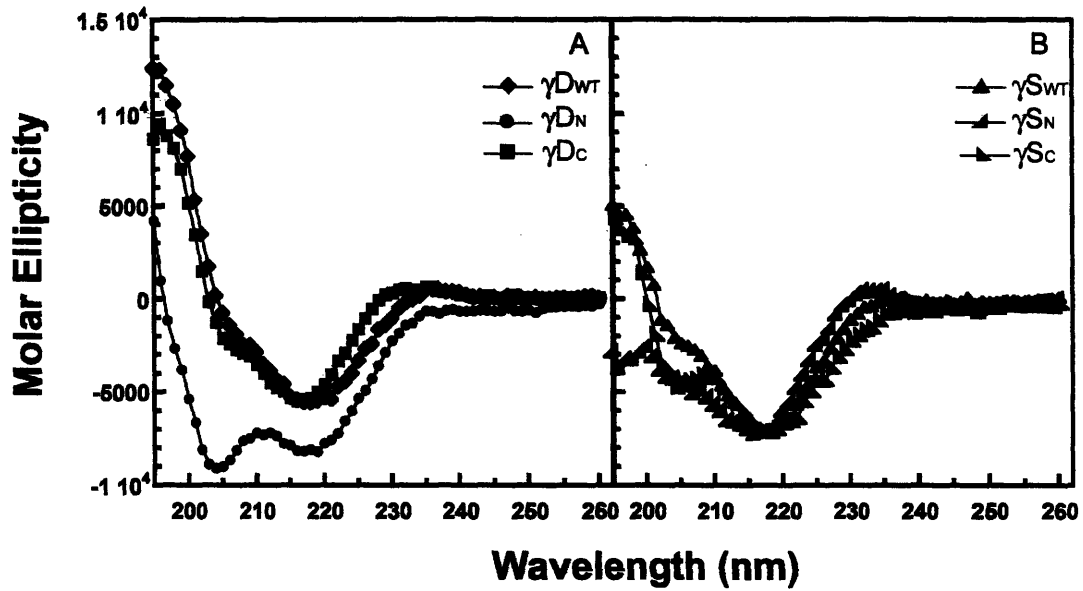


Figure 2-4. Far-UV CD spectroscopy of isolated domains and full-length proteins. Samples are at a protein concentration of 100 $\mu\text{g/ml}$ in 10 mM sodium phosphate buffer, pH 7.0 at 37°C. (A) CD spectra recorded from 195-260 nm wavelengths for $\gamma_{D_{WT}}$ (black \blacklozenge), γ_{D_N} (dark blue \bullet), and γ_{D_C} (red \blacksquare) (B) CD spectra recorded from 195-260 nm wavelengths for $\gamma_{S_{WT}}$ (light blue \blacktriangle), γ_{S_N} (green \blacktriangle), and γ_{S_C} (orange \blacktriangle)

In addition to the four tryptophans, there are 14 semi-conserved tyrosines (71% identity, 93% similarity based on 32 diverse γ -crystallin sequences) located throughout the protein. There are 7 Tyr in each domain of γD_{WT} . For γS_{WT} , there are 8 Tyr in γS_N and 6 in γS_C . In order to monitor preferentially the structure surrounding the tryptophan residues to probe distinct locations of the proteins, we excited the proteins at 295 nm and recorded the fluorescence emission spectra from wavelengths 310 – 400 nm.

The fluorescence emission spectra of the isolated domains at 37°C shown in Figure 2-5 were similar to those of the full-length proteins. γD_{WT} has a quenched native maximum of ~326 nm and a red-shifted unfolded maximum of 350 nm (Table 2-1). The native maximum of γD_C domain is similar to full-length and has quenched fluorescence intensity as well, indicating a native-like structure. The γD_N domain fluorescence had a higher quantum yield than both γD_{WT} and γD_C indicating that it is not as quenched in isolation possibly due to disruptions in structure around Trp 42. However, upon denaturation in GuHCl, the fluorescence intensity of the γD_N domain increased indicating it was native-like in the absence of denaturant.

γS_{WT} had a quenched native emission maximum of ~329 nm and an unfolded maximum of 350 nm (Fig. 2-5D, Table 2-1). Its C-terminal domain exhibited an increase in fluorescence intensity, also indicating that the C-terminal domain had a higher quantum yield and is not as quenched in the native state (Fig. 2-5F). γS_N domain had a similar peak and fluorescence intensity compared to γS_{WT} , although the overall shape of the spectra indicates there may be some differences in tertiary structure (Fig. 2-5E).

In conclusion, all domains in isolation though exhibiting slight alterations in secondary and tertiary structures are relatively structurally similar to their respective full-length proteins as well as to each other.

Table 2-1. Deconvoluted CD spectra and fluorescence emission spectra maximums for γ D and γ S wild type and isolated domain proteins

PROTEIN	CD Spectra				Fluorescence Emission	
	% β -sheet	% α -Helix	% Turns	% Unordered	Native (nm)	Unfolded (nm)
γ D _{wt}	40	6	21	31	326	350
γ D _N	30	11	24	35	326	350
γ D _C	41	5	22	32	326	350
γ S _{wt}	33	6	24	36	329	350
γ S _N	29	4	25	40	329	350
γ S _C	32	9	26	34	329	350

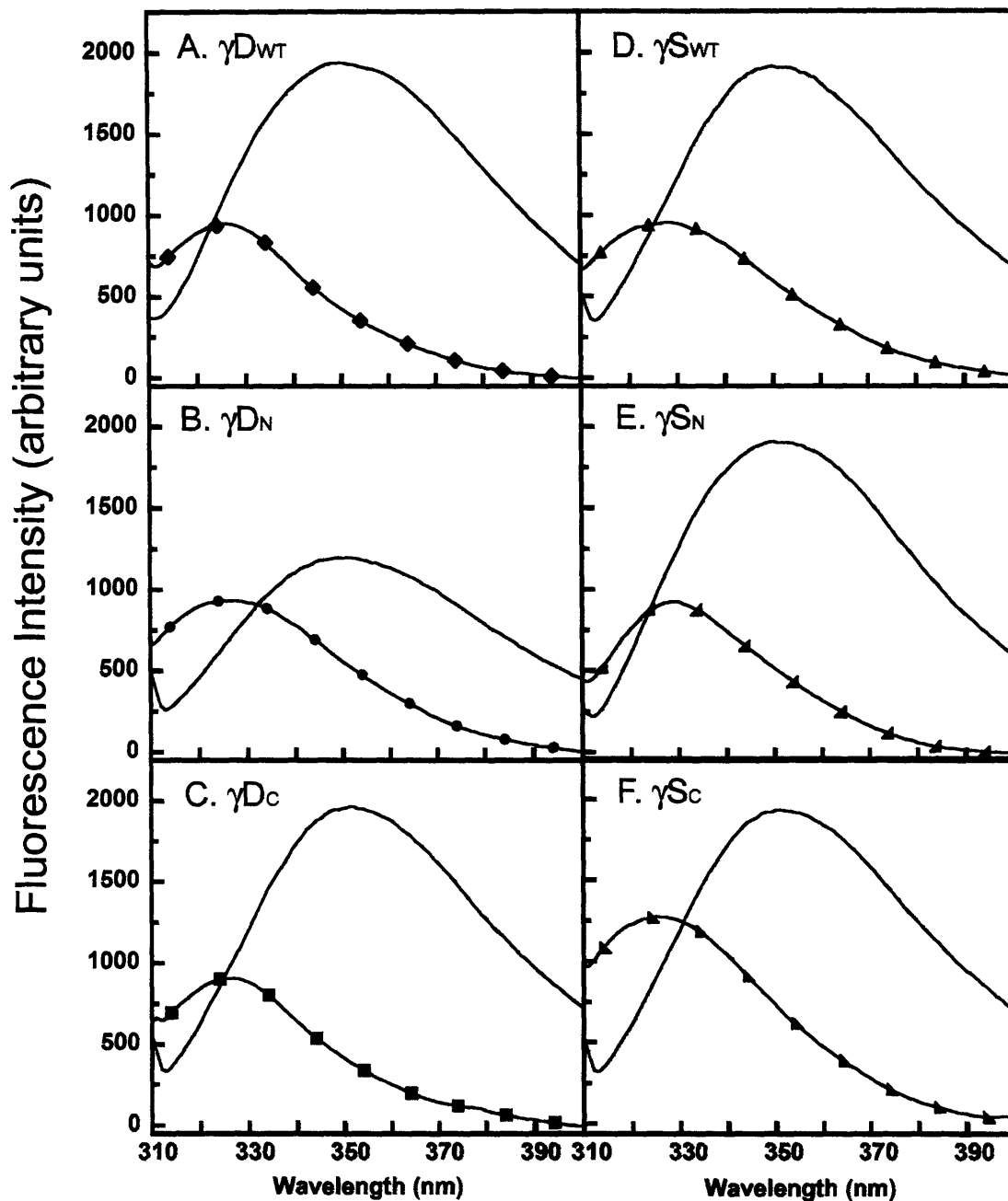


Figure 2-5. Fluorescence emission spectra of native and unfolded of the isolated domains and full-length γD_{WT} and γS_{WT} . All proteins excited at 295 nm and emission recorded from 310–400 nm wavelengths. Samples consisted of 10 $\mu\text{g/ml}$ protein in 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and additionally 5.5M GuHCl for unfolded samples equilibrated at 37°C.

(A) Native spectra for γD_{WT} (◆) and Unfolded spectra for γD_{WT} (line); (B) γD_N Native (●) and γD_N Unfolded (line); (C) γD_C Native (■) and γD_C Unfolded (line); (D) γS_{WT} Native (▲), γS_{WT} Unfolded (line); (E) γS_N Native (◄) and γS_N Unfolded (line); (F) γS_C Native (▴), and γS_C Unfolded (line).

5. Thermal Denaturation Indicates Differential Domain Stability

Circular dichroism and fluorescence emission spectra indicated that the isolated domains were in native-like conformations. To assess the stability of these domains qualitatively we examined their thermal denaturation. Although thermal denaturation is qualitative, this tests the stability in a more physiological manner than denaturant experiments.

Figure 2-6 shows the thermal melting behavior monitored by CD. All four of the isolated domains remained folded until 60°C or higher and exhibited a cooperative melting transition consistent with thermal denaturation unfolding transitions. We monitored the effects of increasing temperature by the loss of β -sheet structure determined from Far UV CD spectroscopy. Thermal denaturation was an irreversible reaction demonstrated by visible aggregation upon the completion of the experiment. Since the calculation of meaningful thermodynamic parameters is limited by the lack of reversibility, we report only the fraction native as a function of increasing temperature (Fig. 2-6).

Both full-length γD_{WT} and γS_{WT} were extremely stable with a T_M of 84.5°C and T_M of 74.1°C, respectively. The differences in stability between the individual domains were evident in the thermal experiments; the T_M of γD_N domain was 64°C, while the T_M of γD_C domain was 77°C (Fig. 2-6A). The T_M of the γS_N domain was 69.1°C and γS_C domain was 75.1°C (Fig. 2-6B). The T_M difference between γD isolated domains was larger than the γS isolated domains confirming the idea that the γD_{WT} interface is essential in increasing the thermodynamic stability of its N-terminal domain. In addition, full-length γD_{WT} has a 7°C increase in T_M compared to the isolated γD_C domain, reinforcing that both domains are necessary for overall conformational stability. The interface of γS_{WT} even though essential in increasing the stability of its N-terminal domain may not be as crucial in the overall stability of the protein as suggested by various other studies (Wenk et al. 2000; Zarina et al. 1994).

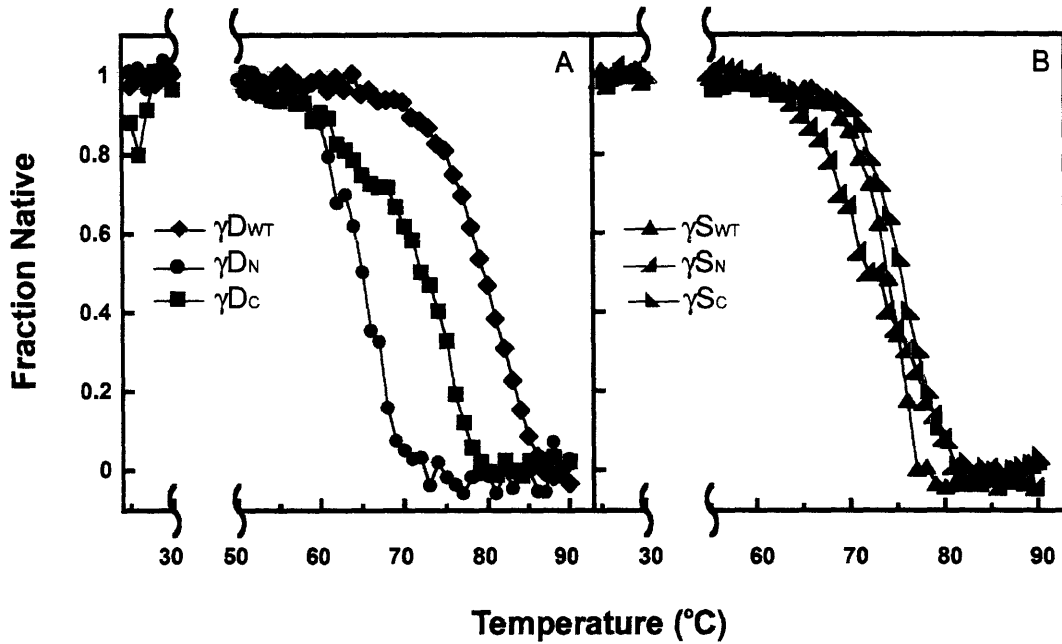


Figure 2-6. Thermal denaturation of γD_{WT} , γS_{WT} , and their respective individual domains. Samples were prepared at 100 $\mu\text{g/ml}$ protein concentration in 10 mM sodium phosphate buffer, pH 7.0 and CD wavelength 218nm was monitored as the temperature increased from 25-90°C. Data normalized and calculated as Fraction Native. (A) γD_{WT} (black \blacklozenge), γD_N (dark blue \bullet), and γD_C (red \blacksquare) (B) γS_{WT} (light blue \blacktriangle), γS_N (green \blacktriangleleft), and γS_C (orange \blacktriangleright)

6. Equilibrium Unfolding and Refolding In Vitro

The polypeptide chains of the isolated domains described above folded into their native-like state within *E. coli*. To determine if the isolated domains would refold *in vitro*, we conducted equilibrium unfolding/refolding experiments. These results also allowed us to estimate quantitatively the stability of each domain under the conditions previously described (Flaugh et al. 2005b). The denaturant GuHCl was used instead of urea because previous analysis had shown that $\gamma_{D_{WT}}$ is resistant to up to 8 M urea denaturant (Kosinski-Collins and King 2003).

As seen in Fig. 2-7, the isolated γ_{D_N} and γ_{D_C} domains refolded at high yield when diluted out of denaturant. As reported by (Wenk et al. 2000), the isolated γ_{S_N} and γ_{S_C} domains also refolded at high yield when diluted out of denaturant. In contrast to $\gamma_{D_{WT}}$, we observed no evidence for aggregated states competing with productive refolding for any of the isolated domains and $\gamma_{S_{WT}}$ (Figs. 2-7 and 2-8).

To compare the stabilities of the isolated domain proteins, the analyses of γ_{D_N} , γ_{D_C} , $\gamma_{S_{WT}}$, and its γ_{S_N} and γ_{S_C} domains were also performed under these same conditions. The equilibrium unfolding/refolding curves were determined by calculating the ratio of the fluorescence emission at 360 nm and 320 nm versus denaturant concentration (Figs. 2-7 and 2-8). Since the irreversibility of $\gamma_{D_{WT}}$ can cause extrapolation errors in determining free energy (H_{20}), we used transition midpoint C_M values (GuHCl concentration at which 50% of the protein is denatured) of the equilibrium unfolding/refolding curves as a measure of stability. However, since all of the other proteins exhibited full reversibility under these experimental conditions, we have extrapolated $\Delta G_{H_{20}}$ values and have included m -values in Table 2-2 for $\gamma_{D_{WT}}$ and Table 2-3 for $\gamma_{S_{WT}}$. These calculations confirmed the C_M comparisons of $\gamma_{D_{WT}}$, $\gamma_{S_{WT}}$ and their respective individual domains.

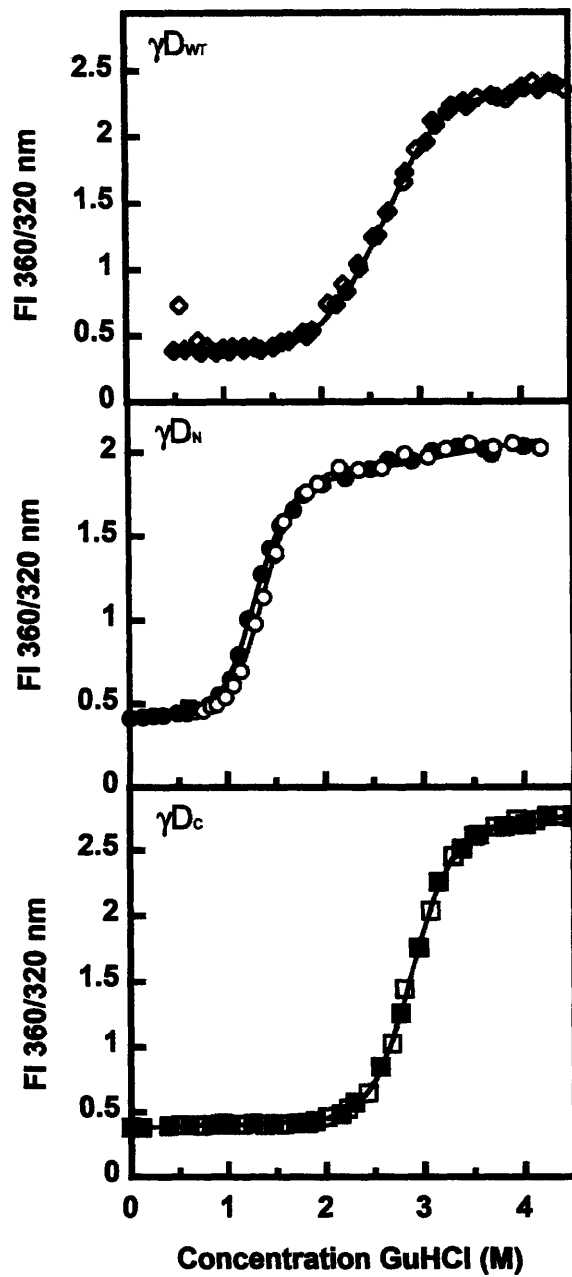


Figure 2-7. Equilibrium unfolding (closed symbols) and refolding (open symbols) for γD_{WT} (\blacklozenge), γD_N (\bullet), and γD_C (\blacksquare). Samples consisted of 10 $\mu\text{g/ml}$ protein concentration, 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and various concentrations of GuHCl at 37°C. Fluorescence emission at 360 nm and 320 nm ratio were calculated. Equilibrium data fit indicated by solid black line.

The equilibrium unfolding/refolding curves for γD_{WT} was best fit to a three state transition (Flaugh et al. 2005b). The C_M of the first transition of the γD_{WT} was 2.2 M GuHCl. The C_M of the second transition was 2.8 M GuHCl. The corresponding ΔG_{H2O} values were 7.7 and 8.9 kcal* mol^{-1} , respectively (Flaugh et al. 2005b). Both domains in isolation fit best to a two-state transition with no detectable equilibrium intermediates along the folding pathway. The γD_N domain was destabilized in isolation with a C_M of 1.2 M GuHCl, and ΔG_{H2O} of 3.7 kcal* mol^{-1} . The γD_C domain stability parameters were C_M of 2.7 M GuHCl and ΔG_{H2O} of 8.7 kcal* mol^{-1} in isolation, comparable to the C_M of 2.8 M observed in the second transition in γD_{WT} . Therefore, this result confirms that the second transition is monitoring the unfolding of the γD_C domain while the first transition is monitoring the unfolding of the γD_N domain. The overall ΔG of γD_{WT} can be estimated by the addition of the ΔG of both individual domains plus the ΔG of the γD_{WT} interface, depicted in the following equation.

$$\Delta G_{\gamma D_{overall}} = \Delta G_{N-td} + \Delta G_{C-td} + \Delta G_{int}, \text{ thus } \Delta G_{int} = \Delta G - (\Delta G_{N-td} + \Delta G_{C-td})$$

Utilizing this equation we estimate that the γD_{WT} interface contributes approximately 4.2 kcal* mol^{-1} to the overall free energy of γD_{WT} . In other words, in the absence of the γD_C and its interface contacts, γD_N was 4.2 kcal* mol^{-1} less stable compared to the γD_{WT} .

7. γS_{WT} Equilibrium Unfolding/Refolding Analysis Also Demonstrates Differential Domain Stability

Initially we did not know if γS_{WT} would refold like γD_{WT} efficiently under these conditions, pH 7 and 37°C. Previously, Wenk et al. analyzed the stability of the human and bovine γS crystallin as well as their isolated N-terminal domains and C-terminal domains at 20°C (2000). Equilibrium unfolding/refolding experiments were performed to analyze the stability of γS_{WT} and its individual domains. We were able to detect complete refolding of γS_{WT} out of denaturant at pH 7 and 37°C as observed in the overlay of native and refolding equilibrium curves (Fig. 2-8).

Table 2-2. Equilibrium unfolding/refolding at 37°C and thermal unfolding parameters of γ D wild type and isolated domain proteins.

Protein	Thermal Unfolding Transition	Equilibrium Transition 1			Equilibrium Transition 2				
		T_M^a	ΔT_M^b	C_M^c	Apparent m value ^d	Apparent $\Delta G_{N-I}^{\circ e}$	C_M^c	Apparent m value ^d	Apparent $\Delta G_{I-U}^{\circ e}$
γ D ^f		83.8 ± 1.3	-	2.2 ± 0.1	3.6 ± 0.1	7.7 ± 0.2	2.8 ± 0.1	3.1 ± 0.4	8.9 ± 1.3
γ D _N		64.5 ± 0.6	19.3	1.2 ± 0.1	3.3 ± 0.9	3.7 ± 0.7	-	-	-
γ D _C		76.2 ± 0.2	7.6	-	-	-	2.7 ± 0.09	3.2 ± 0.1	8.7 ± 0.5

^a Midpoints of melting transitions monitored by 218 nm FarUV CD in units of °C.

^b $\Delta T_m = T_{m \text{ Full-length}} - T_{m \text{ Single Domain}}$ in units of °C.

^c Transition midpoints in units of M GuHCl.

^d Apparent m values in units of $\text{kcal}^* \text{mol}^{-1} * \text{M}^{-1}$.

^e Free energy of unfolding in the absence of GuHCl in units of $\text{kcal}^* \text{mol}^{-1}$.

^f Thermal denaturation parameters of γ D_{WT} are from (Flaugh et al. 2006), Equilibrium parameters of γ D_{WT} are from (Flaugh et al. 2005b).

Therefore, we analyzed the stability of the γS_{WT} as well as its N-terminus and C-terminus in isolation under these previous conditions. The results from these experiments also allowed us to compare γD_{WT} and its isolated domains to γS_{WT} and its isolated domains.

All of the unfolding and refolding transitions of the full-length γS crystallin and the γS isolated domains were best fit to a two state model. The C_M value observed for γS_{WT} was 2.3 M GuHCl while the ΔG_{H2O} was 10.5 kcal* mol^{-1} . The difference in stability between the γS isolated domains was significant but was not as great a difference as seen in the γD_{WT} isolated domains. The equilibrium unfolding experiments yielded a C_M of 1.7 M GuHCl and a ΔG_{H2O} of 4.9 kcal* mol^{-1} for γS_N . The γS_C C_M and ΔG_{H2O} values were 2.3 M GuHCl and 8.2 kcal* mol^{-1} , respectively. By observing the slight differences in N and C-terminal domain C_M values, the C_M of γS_{WT} is likely an average of both of the domain stabilities. Thus, the unfolding/refolding of γS_{WT} may populate intermediates that are not detectable by this method. Since there are differences between the N and C-terminal domains, there is a possibility that the one domain folded, one domain unfolded intermediate described for other γ -crystallins may also occur in γS_{WT} . Further experiments consisting of interface mutants such as those performed in γD_{WT} will be essential in determining the stability of the interface as well as if the folding pathway consists of an intermediate of one domain folded and one domain unfolded.

Comparison between the γD and γS crystallin isolated domains show similar stability results as the thermal denaturation experiments. The γS_N domain was less stable than the γS_C domain; however, the isolated γS_N domain is more stable than the isolated γD_N domain. In contrast, the isolated γD_C domain is more stable than the isolated γS_C domain. In general, from these experiments we detect differential domain stability among both crystallin's N and C-terminal domains. Yet, the sequence homology between both crystallin's N and C-terminal domains is higher than the N and C terminal domains within the γD or γS crystallin.

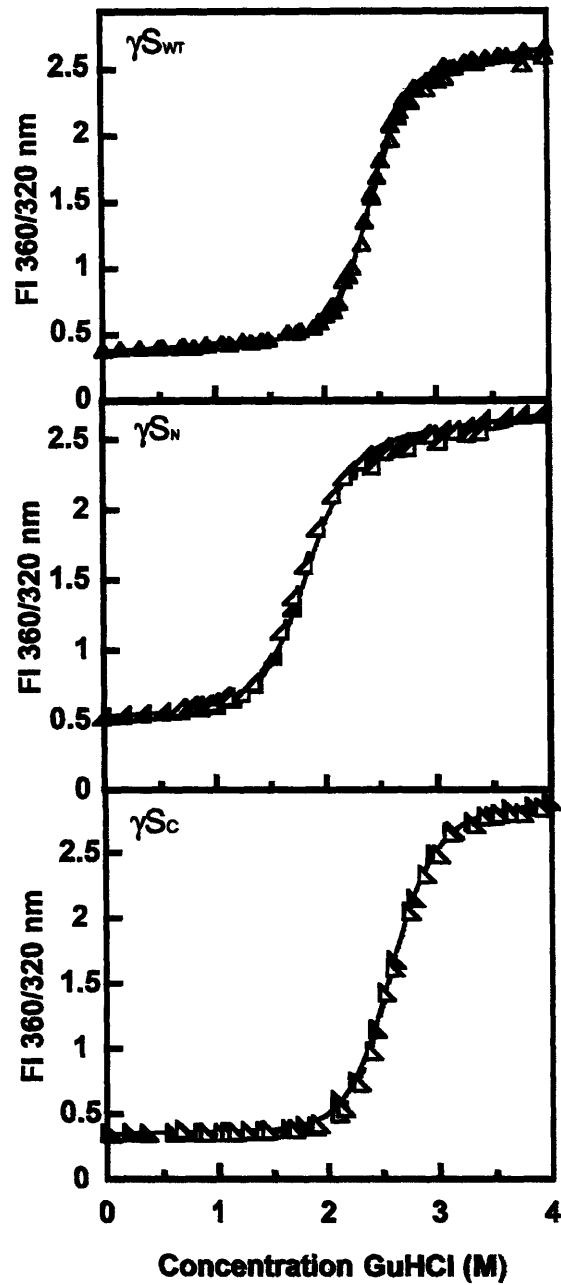


Figure 2-8. Equilibrium unfolding (closed symbols) and refolding (open symbols) for γS_{WT} (\blacktriangle), γS_N (\blacktriangleleft), and γS_C (\blacktriangleright). Samples consisted of 10ug/ml protein concentration, 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and various concentrations of GuHCl at 37°C. Fluorescence emission at 360 nm and 320 nm ratio were calculated. Equilibrium data fit indicated by solid black line.

Table 2-3. Equilibrium unfolding/refolding at 37°C and thermal unfolding parameters of γ S wild type and isolated domain proteins.

Protein	Thermal Unfolding Transition			Equilibrium Transition		
	T_M^a	ΔT_M^b	C_M^c	Apparent m value ^d	Apparent ΔG°_{N-U} ^e	
γ WT	74.1 ± 0.2	-	2.3 ± 0.02	4.6 ± 0.4	10.5 ± 0.9	
γ N	69.1 ± 2.4	5	1.7 ± 0.08	2.9 ± 0.4	4.9 ± 0.8	
γ C	75.1 ± 0.75	-1	2.3 ± 0.07	3.5 ± 0.4	8.2 ± 0.6	

^a Midpoints of melting transitions monitored by 218 nm FarUV CD in units of °C.

^b $\Delta T_m = T_{m \text{ Full-length}} - T_{m \text{ Single Domain}}$ in units of °C.

^c Transition midpoints in units of M GuHCl.

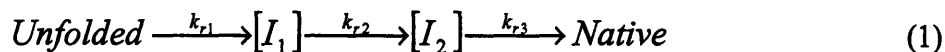
^d Apparent m values in units of $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$.

^e Free energy of unfolding in the absence of GuHCl in units of $\text{kcal} \cdot \text{mol}^{-1}$.

8. Productive Refolding of γD_{WT} , γS_{WT} and Their Respective Isolated Domains

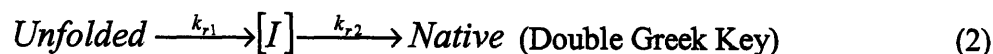
All full-length and single domain proteins had the ability to refold after complete unfolding in GuHCl and subsequent dilution out of denaturant. Protein refolding kinetics was monitored using tryptophan fluorescence spectroscopy to measure the quenching of the tryptophans as they folded into the native state. To better resolve the intermediate steps, these experiments were performed at lower temperature, 18°C. All proteins refolded in a time frame of ~300 s with the exception of γD_{WT} and γS_{WT} (Figs. 2-9 and 2-10). γD_{WT} was productively refolded in 1 M GuHCl since it exhibited off-pathway aggregation when refolded in < 1 M GuHCl (Kosinski-Collins and King 2003). The rest of the proteins were allowed to refold in < 1 M GuHCl.

γD_{WT} refolding kinetic measurements was best fit to a four state transition in these experiments with an average calculated $t_{1/2}$ of 10 s, 36 s, 252 s for the three exponentials (k_{r1} , k_{r2} , k_{r3}) (Fig. 2-9 and Table 2-4) at 18°C (Equation 1).



Mutations in the interface affect the refolding rate of γD_{WT} (Flaugh et al. 2005a; Flaugh et al. 2005b; Flaugh et al. 2006). These mutations suggested that two refolding intermediates are populated; the first intermediate is the C-terminal domain Greek Key motif 4 (one closest to the interface) is folded and the motif 3 of the C-terminal domain and N-terminal domain unfolded. The second intermediate is likely the C-terminal domain fully folded and the N-terminal domain unfolded.

All of the refolding reactions of the isolated domains were best fit by a three state model indicating one observable intermediate for each refolding pathway. For the isolated domains, this kinetic refolding intermediate is likely to be a partially folded individual Greek Key domain.



The kinetic refolding parameters for isolated γD_N and γD_C domains were $t_{1/2}$ of 3 s and 4.5 s for refolding transition 1 (k_{r1}) and 17 s and 14 s for refolding transition 2 (k_{r2}), respectively (Equation 2).

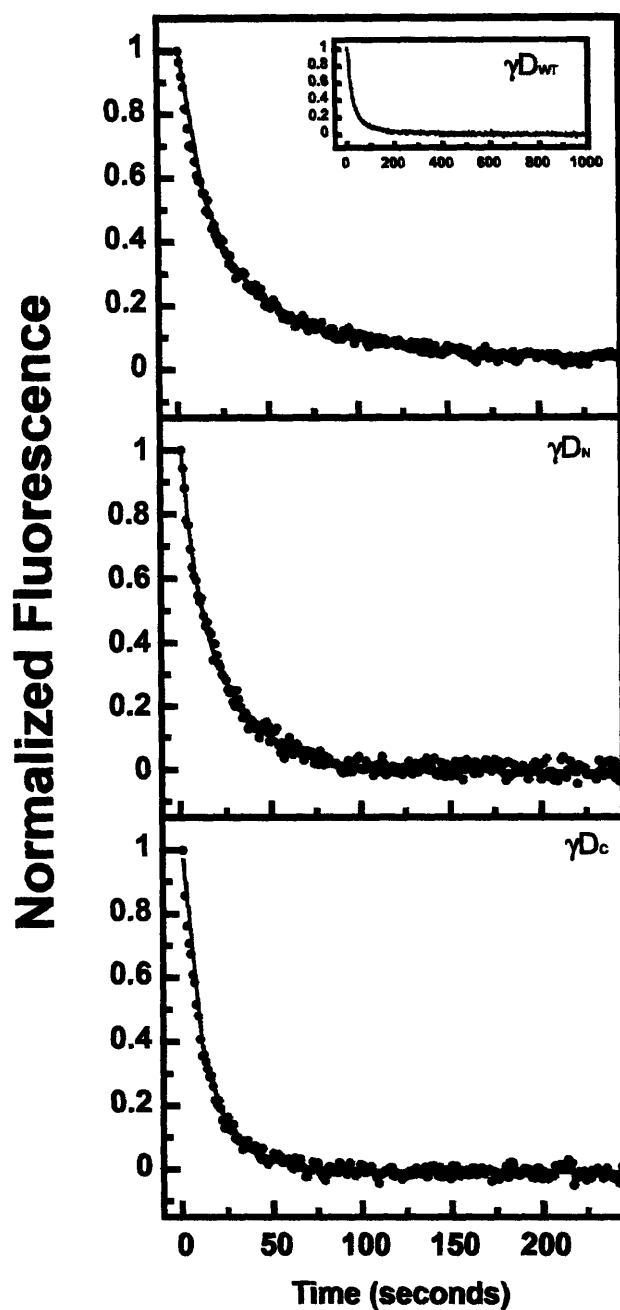


Figure 2-9. Kinetic refolding of isolated domains of γD_{WT} (inset shows completion of γD_{WT} refolding kinetics reaction), γD_N , γD_C . Protein was unfolded at high GuHCl concentration, then diluted into 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer for a final protein concentration of 10 $\mu\text{g/ml}$. Protein tryptophan fluorescence emission at 350 nm was recorded every second and data normalized for comparison. All experiments performed at 18°C. Kinetic refolding reactions of all proteins in 1 M or > 1 M GuHCl concentration (see text for details).

Both individual domain kinetic refolding parameters were within standard deviations of each other for both kinetic refolding transition rates. The first kinetic refolding transition $t_{1/2}$ was 14.5 s for γS_N and ~ 12 s for γS_C and for the second kinetic refolding transition $t_{1/2}$ was 190 s for γS_N and 122 s for γS_C .

These parameters were also within standard deviation of one another for both kinetic refolding transitions (Table 2-4). γS_{WT} refolding kinetics was best fit to a four state model, indicating two intermediates (Fig. 2-10). The refolding rates were considerably slower than γD_{WT} with calculated average $t_{1/2}$ of 20 s, 194 s, and 1043 s for k_{r1} , k_{r2} , and k_{r3} respectively.

D. DISCUSSION

All known eye lens β and γ -crystallins have two homologous Greek key domains interacting through a tight interface, presumably evolved from an ancestral single domain protein. Given the importance of long term solubility and stability for the crystallins, it seems likely that the evolution of a two domain form is related to the need for very long term stability of the lens crystallins. Previous evidence that in γD_{WT} protein, the C-terminus served as a template for the folding of the N-terminus, raised the possibility that the N-terminal domain could not fold or be stable on its own, due to the absence of the domain interface. On the contrary, the results reported here show that the N-terminal domain of γD does fold on its own *in vivo* within the *E. coli* cytoplasm, and remains folded through purification and storage. In addition, both equilibrium and kinetic experiments confirmed that these isolated domains were able to refold to a native-like state upon dilution out of denaturant. The isolated N- and C-terminal domains of γS crystallin also folded *in vivo* and refolded *in vitro*, as previously reported (Wenk et al. 2000).

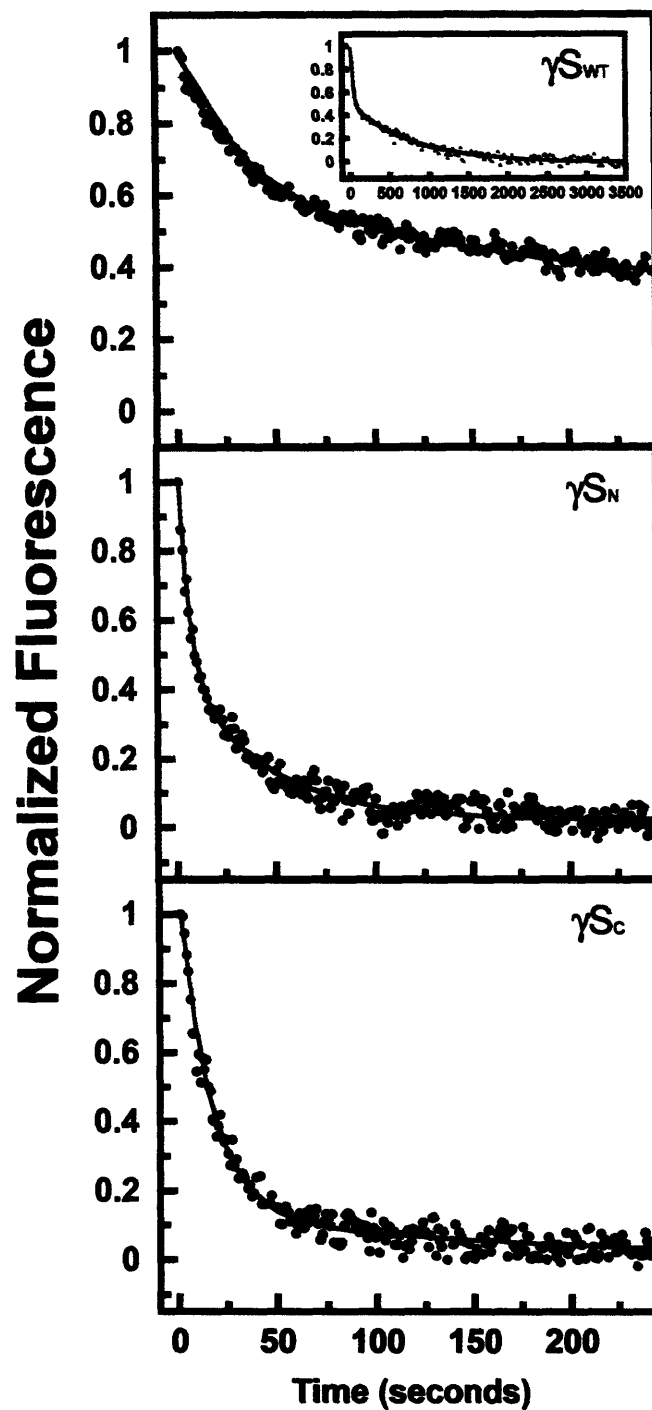


Figure 2-10. Kinetic refolding of isolated domains of γ_{SWT} (inset shows completion of γ_{SWT} refolding kinetics reaction), γ_{SN} , and γ_{SC} . Protein was unfolded at high GuHCl concentration, then diluted into 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer for a final protein concentration of 10 $\mu\text{g/ml}$. Protein tryptophan fluorescence emission at 350 nm was recorded every second and data normalized for comparison. All experiments performed at 18°C. Kinetic refolding reactions of all proteins >1 M GuHCl concentration.

Table 2-4. Productive refolding kinetic parameters of γ D and γ S wild type and isolated domain proteins at 18°C.

Protein	Productive Refolding Kinetics					
	First Transition		Second Transition		Third Transition	
	k_{r1}^a	$t_{1/2}^b$	k_{r2}^a	$t_{1/2}^b$	k_{r3}^a	$t_{1/2}^b$
γ D _{WT}	0.07 ± 0.016	10 ± 2	0.019 ± 0.004	36 ± 7	0.003 ± 0.0007	252 ± 63
γ D _N	0.22 ± 0.07	3 ± 1	0.04 ± 0.001	17 ± 0.4	-	-
γ D _C	0.16 ± 0.03	4.5 ± 0.8	0.05 ± 0.01	14 ± 3	-	-
γ S _{WT}	0.03 ± 0.001	20 ± 0.6	0.0045 ± .002	194 ± 125	0.00066 ± 0.0002	1043 ± 346
γ S _N	0.05 ± 0.01	14.5 ± 5	0.004 ± 0.002	190 ± 84	-	-
γ S _C	0.06 ± 0.008	11.8 ± 1.8	0.006 ± 0.001	122 ± 32	-	-

^a Kinetic refolding rates in units of s⁻¹.

^b Half-life in units of seconds

Nonetheless, the isolated N-terminal domain of γ D was considerably less stable than the isolated C-terminal domain, in addition to being less stable in the full-length protein. Comparison of the stabilities of the isolated domains and the full-length protein indicated that the domain interface contributes ΔG_{H2O} of $\sim 4.2 \text{ kcal}\cdot\text{mol}^{-1}$ to the overall stability of the complete two domain protein. Similarly, isolated domain and full-length protein analysis of bovine γ B crystallin at pH 2 determined that the interface contributed a ΔG_{H2O} of $\sim 3.8 \text{ kcal}\cdot\text{mol}^{-1}$ (Mayr et al. 1997). The differential domain stability was also observed for the γ S crystallin isolated domains, although the stability differences between the two domains were not as significant as γ D_{WT} isolated domains.

1. Native-like Folded Conformation of the Isolated Domains

The unusual fluorescence quenching of the buried tryptophans provides a very sensitive reporter of the folded state of the crystallins (Chen et al. 2006). Both emission maximum and intensity of the fluorescence spectra of the isolated domains indicated that they were in the native-like fold. The γ S_C Trps are highly fluorescent consistent with the increased fluorescence spectra observed when probing only C-terminal domain Trps in γ S crystallin N-terminal Trp > Phe mutants (Chen et al. unpublished). Circular dichroism spectroscopy confirmed the native-like structure of the isolated domains, indicating only minor structural differences between the domains themselves and their full-length counterparts. Analytical size exclusion demonstrated that all of the isolated domain chromatographs showed monomeric species that did not form stable dimers or multimeric species.

2. Stability of the Full-length γ -Crystallins and Their Isolated Domains

Comparison between both proteins demonstrated that γ D_{WT} is more stable than γ S_{WT}. This result was suggested empirically previously since γ S_{WT} could be unfolded in high molar urea whereas γ D_{WT} could not (Kosinski-Collins and King 2003; Wenk et al. 2000). By comparing the γ D and γ S isolated domains, we observed that the γ D_N domain

is less stable in isolation than γS_N domain while γD_C domain is more stable than the γS_C domain (Table 2-2 and 2-3).

The thermal denaturation qualitative data agreed with our unfolding/refolding equilibrium studies concluding that the N-terminal domains (γD 64.5°C, γS 69.1°C) are less stable than the C-terminal domains (γD 76.2°C, γS 75.1°C) of both γD_{WT} and γS_{WT} . Also overall γD_{WT} (83.8°C) is more stable than γS_{WT} (74.1°C). The γD_{WT} T_M was higher than both individual domains emphasizing that both domains contribute to the overall stability of the full-length protein. The isolated γS_C domain T_M is similar compared to full-length γS_{WT} T_M . Although the average of the T_M measurements of γS_C is higher than its full-length protein, it is within standard deviation of the average of the γS_{WT} . Therefore, the intrinsic stability of the C-terminal domain contributes most of the overall stability of the γS_{WT} . Thermodynamic parameters were not calculated from these experimental results because of the irreversibility of the reaction due to protein aggregation.

The thermal denaturation experiments represent the stability of the protein in more physiological conditions than in high concentrations of GuHCl. During thermal denaturation the unfolding molecules aggregate irreversibly corresponding to the events that are believed to occur within the lens. Exposure to high heat among glass blowers and chain welders has been correlated with a high incidence of cataract (D'onofrio and Mosci 1960; Vos and van Norren 1998; Vos and van Norren 2004). Thermally unfolded proteins could lead to partially unfolded species that are aggregation prone and subsequently initiate cataract formation.

3. Kinetic Refolding of the Isolated Domains and Their Respective Full-length Proteins

Equilibrium unfolding/refolding experiments at 37°C demonstrated full reversibility establishing that the isolated domain crystallins could refold *in vitro* efficiently in the absence of chaperones. The refolding kinetics of the isolated domain crystallins were best fit to three state models indicating one intermediate. The γD

individual domain's kinetic rates were similar and within standard deviation of one another. The γ D isolated domains exhibited higher kinetic rates compared to the γ S isolated domains, a ~ 3 fold increase for k_1 and a ~ 11 fold increase for k_2 . γ S isolated domains also had similar kinetic rates to one another.

Since each crystallin domain contains two Greek Key motifs, one may predict that the transitions of the isolated domains may be monitoring the refolding of each Greek Key motif sequentially. Previous studies with a $\beta\gamma$ -crystallin fold protein, has suggested that the interface Greek Key motif and stabilizing tyrosine corner can act as a nucleation center for folding of the second outer Greek Key motif (Bagby et al. 1998) (Fig. 2-1). Thus, the first transition of the refolding kinetics may monitor the innermost Greek Key motif while the second transition monitors the outermost Greek Key. In contrast, since there is no interface available for these proteins, the first transition may monitor the refolding of the outermost Greek Key or the combination of both Greek Key motifs. The two tryptophans in different regions of the domain monitoring the folding of the proteins both contribute to fluorescence quenching rendering it difficult to sparse out the regions of the protein sequentially folding (Chen et al. 2006).

γ D_{WT} kinetic refolding data was best fit to a four state model, indicating two intermediates. Previous analysis of γ D_{WT} refolding kinetics indicated that the first transition monitors the refolding of the C-terminus innermost Greek key motif. The second transition is thought to monitor the refolding of the second, outermost C-terminus Greek Key motif while the third transition is complete refolding of the N-terminus (Flaugh et al. 2006). This analysis was based on interdomain interface mutants which mostly affected the third refolding transition, suggesting that the interface was important for the folding of the N-terminus.

The refolding of γ S_{WT} was best fit to a four state model indicating two partially folded intermediates along the folding pathway. The overall refolding $t_{1/2}$ was considerably longer than γ D_{WT}, with a 4 fold increase in the overall refolding half-life (Table 2-3). The folding pathway of γ S_{WT} has not been elucidated so it is not possible to

compare its transitions to γD_{WT} . However, one possible model for γS_{WT} refolding is the two domains could fold separately monitored by the first transition and the second transition monitoring the interactions within the interdomain interface. As mentioned above, previous studies have suggested that the interface Greek Key motif and the stabilizing tyrosine corner can act as a nucleation center for folding of the second Greek Key motif. Thus, it is possible that the first and second transition is monitoring the refolding of these two innermost Greek Key motifs and interface interactions while the third transition is monitoring the folding of the outermost Greek Keys.

4. Biochemical Basis of Stability Differences of the Isolated Domains

Although the chain fold in γD and γS N and C-terminal domains is highly homologous, the primary sequences are 39% identical, 52% similar for γD domains and 30% identical, 47% similar for the γS domains. However, upon structural and sequence comparison of the two domains, it is not obvious what causes the differences in the stability. There is also variation in stability between different β and γ -crystallins, although structurally the double Greek Key domains are homologous. In this study, we observed that γD_{WT} and γS_{WT} have different conformational stabilities, when analyzed under the same conditions. Moreover, in high concentrations of urea, γS_{WT} completely unfolds, which is not the case for γD_{WT} (Kosinski-Collins and King 2003; Wenk et al. 2000). In comparing $\beta B2$ and Bovine γB crystallin, $\beta B2$ crystallin can unfold in low urea concentrations while γB crystallin requires acidic pH. Without experimental evidence, there are no obvious reasons for these drastic stability differences when comparing these two structures (Jaenicke and Seckler 1997). Furthermore, one would suspect that the oligomeric protein would be more stable than the monomeric protein (Jaenicke and Sterner 2003).

The γD_{WT} interface consists of six hydrophobic residues, three in each domain (M43, F56, I81, V132, L145, V170) and two pairs of peripheral residues (Q54 and Q143, R79 and M147) at the top and bottom of the hydrophobic interface. The interdomain

interface of γS_{WT} is similar to γD_{WT} , the major difference is within the hydrophobic region Phe>Ile at position 56 in γD . Among alignments of 35 various γ -crystallins sequences, Phe at position 56 is 80% conserved. Mutations of Phe>Ala have demonstrated destabilization in both bovine γB and human γD crystallin (Flaugh et al. 2005b; Palme et al. 1997; Palme et al. 1998a). The introduction of a smaller hydrophobic group in this position might contribute to less hydrophobic packing in the interface (Palme et al. 1998b). Additional mutations of Phe>Trp and Phe>Asp at position 56 in the bovine γB crystallin lead to destabilization of the protein emphasizing the importance of the Phe at this position in the interface (Palme et al. 1997).

One possible contributor to the stability of β and γ -crystallins could be the packing of hydrophobic residues, particularly aromatic residues. Both γD_{WT} and γS_{WT} have high percentages of aromatic residues with 4 Trp, 14 Tyr, 6 Phe (γD_{WT}) and 9 Phe (γS_{WT}), consisting of ~13-14% of residues compared to ~8% in other small globular proteins (McCaldon and Argos 1988). Most of the aromatic residues are strongly conserved (>80% identity) when comparing a species diverse set of 35 γ -crystallin sequences with the exception of a few Tyr and Phe. Aromatic-aromatic interactions identified in high resolution structures are predicted to stabilize the protein's native state (Burley and Petsko 1985). Long range hydrophobic interactions have been computed by structure and sequence comparison analysis in the cellular retinoic acid binding protein and have been found to be important for its stability (Gunasekaran et al. 2004).

In addition to the stabilizing interdomain interface, the tyrosine corner of the γ -crystallin proteins has been proposed to be a contributor to stability. The tyrosine corner is present between strand c and strand d of the second Greek Key motif (Bax et al. 1990; Hemmingsen et al. 1994). The tyrosine corner has a slight difference in conformation in γS C-terminal domain as compared to other solved γ -crystallin structures such as γD_{WT} (Purkiss et al. 2002). In the murine γS crystallin NMR structure, γS_N did not appear to have a tyrosine corner present (Wu et al. 2005). It is possible that the lack of the tyrosine corner in the γS_N has an effect on the stability of this domain.

Electrostatic interactions on the surface of the protein may also be contributing to the overall intrinsic stability of the protein. A surface charge network of human and bovine γ A-D crystallins has been observed through comparative and homology modeling (Salim and Zaidi 2003). These electrostatic interactions range from ionic pairs to interaction clusters, some of which are conserved and are located in particular regions of the Greek Key motif. For instance, one cluster is found on the loop regions in motifs 3 and 4 of γ A-D crystallin's C-terminal domain. Many of these ionic interactions have been confirmed through the γ D_{WT} structure (Basak et al. 2003). By comparing potential ionic interactions in the γ S C-terminal domain crystal structure, surface electrostatic interactions are also observed albeit these interactions seem to be at a longer range ($>4\text{\AA}$) than in the γ A-D crystallins.

Although surface electrostatic residues are thought to have a limited effect on stability due to water's dielectric constant, thermophilic proteins have ionic networks essential for stability. We calculated the net charge of each γ D and γ S crystallin isolated domain by utilizing <http://zbio.net>. Although, all proteins were close to their isoelectric points, there are slight differences in the net charge of the γ D_{WT} and the γ S_{WT} isolated domains at pH 7. The calculated net charge (Δz) was 2.2, 1.9, and 1.6 for γ D_{WT}, γ D_N and γ D_C, respectively. While, the calculated net charges (Δz) for γ S_{WT}, γ S_N, and γ S_C, were 0.9, 1.7, and 0.7, respectively. As a protein increases in net charge, the stability of the protein becomes more sensitive to ionic changes in the environment (Negin and Carbeck 2002). Since there were slight differences in net charge as well as possibilities of non-measurable local charge interactions, we performed thermal denaturation experiments in a low ionic strength buffer to better access intrinsic stability. In the thermal denaturation experiments, differential domain stability observed in the equilibrium studies was maintained as the C-terminal domains were more stable than the N-terminal domains. These results suggest that electrostatic interactions may have only minor effects in the intrinsic conformational stability of the γ D and γ S crystallins isolated domains. This is contrary to bovine γ B crystallin studies, which demonstrated that at neutral pH the

individual domains were similar in stability (Mayr et al. 1997). In bovine γ B N and C-terminus studies, the C-terminal domain was greatly destabilized at acidic pH and not at neutral pH. This suggested that the stabilizing interface contacts were not absent at neutral pH but difficult to access using equilibrium unfolding/refolding spectroscopic methods (Mayr et al. 1997).

The NMR structure of the murine γ S crystallin also indicates some differences in the region of the interface and linker that may indicate less packing of the interface (Wu et al. 2005). However, the authors cautioned against making conclusions without experimental evidence as to the packing of the γ S crystallin interface. Therefore, we cannot rule out the possibility that human γ S crystallin also folds one domain at a time with stabilizing interface contacts, having an intermediate with the C-terminal domain folded and the N-terminal domain unfolded. Nonetheless, the interdomain interface and/or linker of γ S_{WT} may not be as important for stability as the interdomain interface of γ D_{WT}.

5. Comparisons with Other Crystallins

Expression of β and γ -crystallins is specific to lens fiber cells, but their expression varies during development and in the different lens regions. In particular, γ A, γ B, γ C, and γ D crystallins (γ E and F are pseudogenes in humans) and β B1 crystallins are expressed *in utero* and are localized to the lens nucleus (Aarts et al. 1989; Chambers and Russell 1991; Lampi et al. 2002). In contrast, β B2 and γ S crystallin, are postnatally expressed in the secondary fiber cells and are expressed throughout life (Peek et al. 1992a; Ueda et al. 2002; Wistow et al. 2000). These physiological differences in the crystallins may be indicative of the stability differences observed biophysically. For example, the observation that γ S_{WT} is less stable than γ D_{WT}, may be related to the location in the lens: γ S_{WT} localized to a region that exhibits some protein synthesis due to the continual lens growth throughout life.

In addition to the aforementioned studies of γ B crystallin, isolated domain studies of the domain-swapped dimeric β B2 crystallin found that the N-terminal domain was marginally less stable than its C-terminal domain in isolation. The N-terminal domain in the monomeric form had a higher propensity to unfold. These results suggested that the unfolding of the N-terminal domain promotes dissociation of the N-terminal domain from the C-terminus of its partner molecule, forming a monomeric intermediate in which the N-terminus is unstructured and the C-terminal domain is folded. Thus, the second equilibrium transition monitors the unfolding of the C-terminal domain (Fu and Liang 2002b; Wieligmann et al. 1999).

β B1 crystallin forms dimers and oligomers not by domain swapping but by intramolecular association (Bateman et al. 2001; Lampi et al. 2001; Van Montfort et al. 2003). Studies using spin-labeling of the β B1 N-terminus demonstrated that the N-terminal domain of β B1 unfolded first. Therefore, the N-terminal domain was hypothesized to be less stable than the C-terminal domain (Kim et al. 2002). All of these studies demonstrated the importance of the domain interface in the stability of the β -crystallins.

The high thermal stability observed in our thermal denaturation experiments is also consistent with other β and γ -crystallins previously analyzed. For example, calorimetry experiments on bovine γ B, γ F, γ E, and γ D, established T_M s of 71.5°C, 70°C, 73°C, and 74°C, respectively (Sen et al. 1992). β -crystallins have lower thermal stabilities such as β B2 whose T_M is 67°C. Also, α -crystallins have lower thermal stabilities with a T_M s of ~ 61°C ((Das et al. 1997; Raman and Rao 1997; Surewicz and Olesen 1995). The crystallins' calculated T_M 's are within the range of structural proteins of thermophilic organisms (Jaenicke 1996).

6. Gene Duplication in the Crystallins

The crystallins are thought to originate from a single domain ancestor supporting evidence of a gene duplication and fusion event in the crystallin lineage (Lubsen et al. 1988; Piatigorsky 2003; Shimeld et al. 2005). It is interesting to note that one domain remains more stable than the other indicating one domain may have evolved increasing stability or the other domain accrued destabilizing properties. The significance of this observation is unknown in γS_{WT} , since the conformational stability of γS_C domain is similar to the stability of the full-length protein in both thermal denaturation and equilibrium experiments. However, in γD_{WT} , the addition the γD_N domain and the interdomain interface to the γD_C contribute to the overall stability of the full-length protein. This additive stability of both domains is demonstrated in both equilibrium and thermal denaturation experiments on γD_{WT} . The gene duplication event in the modern crystallins could be attributed to the necessity for crystallins with higher stabilities in longer lived organisms.

CHAPTER THREE:

**DIFFERENTIAL KINETIC STABILITY OF GAMMA D AND GAMMA S, AND
THEIR ISOLATED GREEK KEY DOMAINS**

A. INTRODUCTION:

As discussed in Chapter one, the crystallin protein must remain stable and soluble to maintain lens transparency and proper vision. Loss of solubility of the crystallin can lead to cataract formation. The lens fibers in the core nuclear region of the lens have no protein synthesis since these cells are enucleated and devoid of all organelles. In the cortical regions of the lens, the younger cells that are formed postnatally have protein synthesis and degradation.

The differential expression of crystallins is thought to have an effect on the overall properties of the lens. The composition of the different type of crystallins may be related to the gradient of refraction and visual acuity necessary for diverse organisms. γ S crystallin, the only crystallin in the soft lens of birds is primarily in the soft regions of the lens (Bloemendal et al. 2004). Moreover, the γ S crystallin gene is common to all vertebrates while the γ A-F genes are mostly mammalian. The γ A-F crystallins are primarily in the water poor regions of the lens (Bloemendal et al. 2004). In humans, there is low expression of the γ A and γ B genes while the γ E and γ F are pseudogenes (Brakenhoff et al. 1990). In the lens, the γ A-F crystallins are expressed in the core regions of the lens *in utero* while γ S is expressed in the cortical regions surrounding the core postnatally.

The differential expression may reflect difference in the stabilities or lifetimes of the diverse β and γ crystallins. Several β and γ -crystallins exhibit very high stability, though with β -crystallins considerably less stable than γ -crystallins. This high stability has been suggested to be due to the complex topology of the β and γ -crystallin Greek Key fold (MacDonald et al. 2005). This topology includes the intercalated double Greek Key domains, the interdomain interface and linker, and in some cases the tyrosine corner. Characterizing the high conformational stability of the crystallins may be important in understanding their properties necessary for their function in the lens.

One source of stability for the crystallins is kinetic stability, a large activation energy between the folded native states and intermediates or unfolded state along the refolding or unfolding pathway. High kinetic stability for unfolding has been observed though not quantitatively measured in bovine γ F crystallin (Das and Liang 1998). More quantitative unfolding kinetics have been performed in microbial crystallins, Proteins S and Spherulin 3a, which have a similar Greek Key fold to the vertebrate ubiquitous β and γ crystallins, although it is not clear if these proteins are ancestral (Bagby et al. 1994a). These microbial $\beta\gamma$ -crystallins exhibit both high conformational and kinetic stability that increases upon binding Ca^{2+} (Kretschmar and Jaenicke 1999; Kretschmar et al. 1999a; Wenk et al. 1998; Wenk and Jaenicke 1999; Wenk et al. 1999).

In addition to the crystallins, other proteins have been found to be stabilized kinetically. This phenomenon is seen in extracellular bacteria proteases that need to remain resistant to self-proteolysis and stable in their bacteria's harsh extracellular environment. For instance, the native state of the α -lytic protease does not have high thermodynamic stability; in fact the unfolded state has a lower free energy than the native state of the protein (Cunningham and Agard 2003). In order to remain folded and stable, the α -lytic protease has a high barrier to unfolding (Cunningham et al. 1999). However, this high kinetic barrier provides an additional problem for folding the α -lytic protease into the native state (Anderson et al. 1999). The solution is that the pro region of the protein acts as a catalyst for the folding of the protease into the native state (Jaswal et al. 2002). After the folding is accomplished, the pro region is cleaved and the native state remains stable through kinetic stability with calculated half-life of ~ 1 year (Sohl et al. 1998).

Other proteins can utilize kinetic stability in the formation of multimers from monomers. Transthyretin is a human plasma protein that has a β -sandwich fold and in an altered monomer form is a precursor to senile systemic amyloidosis and familial amyloid neuropathy (Colon and Kelly 1992). Mutations in humans have been found to produce this amyloidogenic altered monomeric state. In efforts to study why this protein forms an amyloidogenic monomer, denaturation equilibrium unfolding and refolding

studies were performed. The equilibrium transition exhibited strong hysteresis, attributed to a high activation energy between the tetramer and monomer transition (Lai et al. 1997). Further studies demonstrated that transthyretin variants as well as some molecules bound to wt transthyretin, provides a large energy barrier that maintains the tetramer conformation, preventing formation of the amyloidogenic partially unfolded monomer (Johnson et al. 2005). Other proteins such as the legume lectins, the carbohydrate binding proteins, also have high kinetic barriers between monomer and multimeric states (Ghosh and Mandal 2006).

In Chapter two, the conformational stability of the two homologous crystallins γ D and γ S was analyzed. The results confirmed that the γ D crystallin populated an unfolding intermediate in which the C-terminus was folded and the N-terminal domain was unfolded. Equilibrium unfolding/refolding studies demonstrated that γ D crystallin, the lens nucleus localized crystallin, was more stable than the γ S crystallin.

In this chapter, an extensive comparison of the kinetic stability of two homologous yet divergent crystallins, γ D and γ S, and their respective isolated domains was performed. The very high kinetic stability of human γ D crystallin may reflect its function as the longest lived crystallin in the eye lens. The nomenclature in this chapter will be the same as Chapter 2, with the full-length proteins of human γ D crystallin and human γ S crystallin noted as γ D_{WT} and γ S_{WT}, respectively. The isolated domains will be noted with a subscript N or C to state which domain in isolation is referred (e.g., γ D_N or γ D_C).

B. MATERIALS AND METHODS:

1. Preparation of Constructs for wild type and isolated domain

As previously mentioned in Chapter two, the wild type and isolated domains were cloned into the pQE1 His-tag containing vector (Qiagen). The γ D_{WT} and γ S_{WT} consisted

of G1 – S174 for γ_D and S1 – E177 for γ_S . The γ_{D_N} consisted of residues G1- P82 and γ_{D_C} consisted of residues R89 – S174 (based of a numbering in PDB ID: 1HK0). The γ_{S_N} consisted of residues S1 – H86 and γ_{S_C} Y93 – E177 (PDB ID: 1ZWO).

2. Expression and Purification of Proteins

Recombinant full-length and variant proteins were prepared as described in Chapter 2. Briefly, all aforementioned vectors were transformed into *E. coli* M15[pRep4] cells (Qiagen), utilized for tightly regulated protein expression. The cells were lysed by conventional methods and purified by Ni-NTA resin (Qiagen) affinity chromatography using a Pharmacia FPLC apparatus. The purity and size of each protein was confirmed by SDS-PAGE and Mass spectrometry. This purification protocol produced proteins with a purity of > 90%.

Protein concentrations were determined by unfolding of proteins in 5.5 M GuHCl and measuring absorbance at 280 nm using their respective protein extinction coefficients; $\gamma_{D_{WT}}$, γ_{D_N} , γ_{D_C} , $\gamma_{S_{WT}}$, γ_{S_N} and γ_{S_C} , are $41,040 \text{ cm}^{-1}\text{M}^{-1}$, $20,580 \text{ cm}^{-1}\text{M}^{-1}$, $21,555 \text{ cm}^{-1}\text{M}^{-1}$, $41,040 \text{ cm}^{-1}\text{M}^{-1}$, $21,860 \text{ cm}^{-1}\text{M}^{-1}$, $19,180 \text{ cm}^{-1}\text{M}^{-1}$, respectively.

3. Equilibrium Unfolding and Refolding

Equilibrium unfolding/refolding experiments were performed at 18°C. Each unfolding equilibrium sample consisted of 10 $\mu\text{g/ml}$ of protein with increasing concentrations of GuHCl (0 M – 5.5 M) in 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer (Guanidine Hydrochloride solution, 8 M (GuHCl), Sigma®). All samples were equilibrated for 24 hours in order to reach equilibrium except for $\gamma_{D_{WT}}$ which required a 192 hour equilibration to reach equilibrium. For equilibrium refolding experiments, a 10X protein solution was unfolded at 5.5 M GuHCl for 5 hours. The unfolded protein was then diluted 10-fold into various concentrations of GuHCl (0 M-5.5 M) giving a lowest GuHCl concentration of ~ 0.55 M GuHCl. Samples were excited at a wavelength of 295 nm. Fluorescence emission spectrum was recorded from wavelength

range 310 – 400 nm and buffer corrected. Calculations of thermodynamic parameters were performed on 360 nm emission data and 360/320 nm emission ratio data using Kalediagraph software version 4.0 (Synergy Software). Single wavelength 360 nm data was used to calculate m and ΔG values. Each experiment was repeated three times to determine averages and standard deviation parameters.

4. Unfolding Kinetics

Unfolding kinetic experiments were performed by first equilibrating 10X of purified protein in 100 mM phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer at 18°C. Then the protein sample was diluted 10-fold to a constantly agitated 3.5 M or 5.5 M GuHCl buffer using an injection port syringe feature with a dead time of ~1 s. Temperature was maintained at 18°C utilizing the circulating water bath feature. Each sample was excited at 295 nm and the emission wavelength of 350 nm was recorded over time. The parameters for the fluorimeter were 10 nm bandpass for emission and excitation monochromators. Unfolded protein and native protein spectrum controls were recorded to ensure that the native state value was represented in the beginning of the experiment and that the protein was completely unfolded at the end of the experiment. The data was analyzed by fitting it to equations depicting two state, three state, or four state models by Kaldeidagraph 4.0 software. The best fit was determined by analyzing the most random residuals of the different fits. These experiments were repeated three times and the data was averaged with standard deviations determined. Normalized fluorescence data is depicted to allow for comparison between various proteins.

5. Linear Extrapolation of Unfolding Kinetics

Kinetic unfolding experiments were performed for each protein in various concentrations of GuHCl at 18°C. The concentration of GuHCl was above the equilibrium midpoints determined for each protein at 18°C. Natural logarithmic kinetic unfolding rates for different kinetic transitions were plotted versus concentration of GuHCl. A linear regression was fit between all points and extrapolated to the y-axis

using Kaleidagraph. The R-value of each fit was determined and is designated on each graph. The y-axis intercept, the extrapolated kinetic rate, was utilized to calculate the half-lives of each kinetic transition.

C. RESULTS

1. Equilibrium unfolding/refolding experiments at different temperatures and equilibration times resulted in hysteresis

Previous equilibrium unfolding/refolding experiments with γD_{WT} demonstrated a hysteresis under the initial equilibrium conditions of 37°C for 6 hours incubation time (Flaugh, et al. unpublished observations, Fig. 3-1B). This observation was striking in that to be at true equilibrium both unfolding and refolding curves are expected to be superimposed. In particular, the unfolding of the protein required higher GuHCl concentration than the refolding of the protein. This difference was observed specifically in the first transition of the unfolding protein at 37°C. Equilibrium unfolding/refolding experiments performed at 25°C with a equilibration time of 24 hours, increased the hysteresis ((Kosinski-Collins and King 2003, Fig. 3-1A). Under these conditions, both unfolding transitions were resistant to higher concentrations of GuHCl than the refolding transitions. This result was consistent with a kinetically controlled event along the unfolding pathway. If the unfolding transitions are kinetically controlled, one would expect that an increase in equilibration time would alleviate the hysteresis. When the γD_{WT} equilibration time was increased to 24 hours at 37°C, there no longer was a hysteresis, and the equilibrium unfolding and refolding transitions were indistinguishable indicating that the protein had reached equilibrium (Fig. 3-1C). Thus, a kinetic barrier from the unfolded state to the previously identified partially folded intermediate was indicated.

Contrary to γD_{WT} , γS_{WT} did not exhibit a hysteresis at lower temperatures or at shorter equilibration times. It was interesting that although structurally similar, these proteins exhibited differences in their unfolding kinetic barriers.

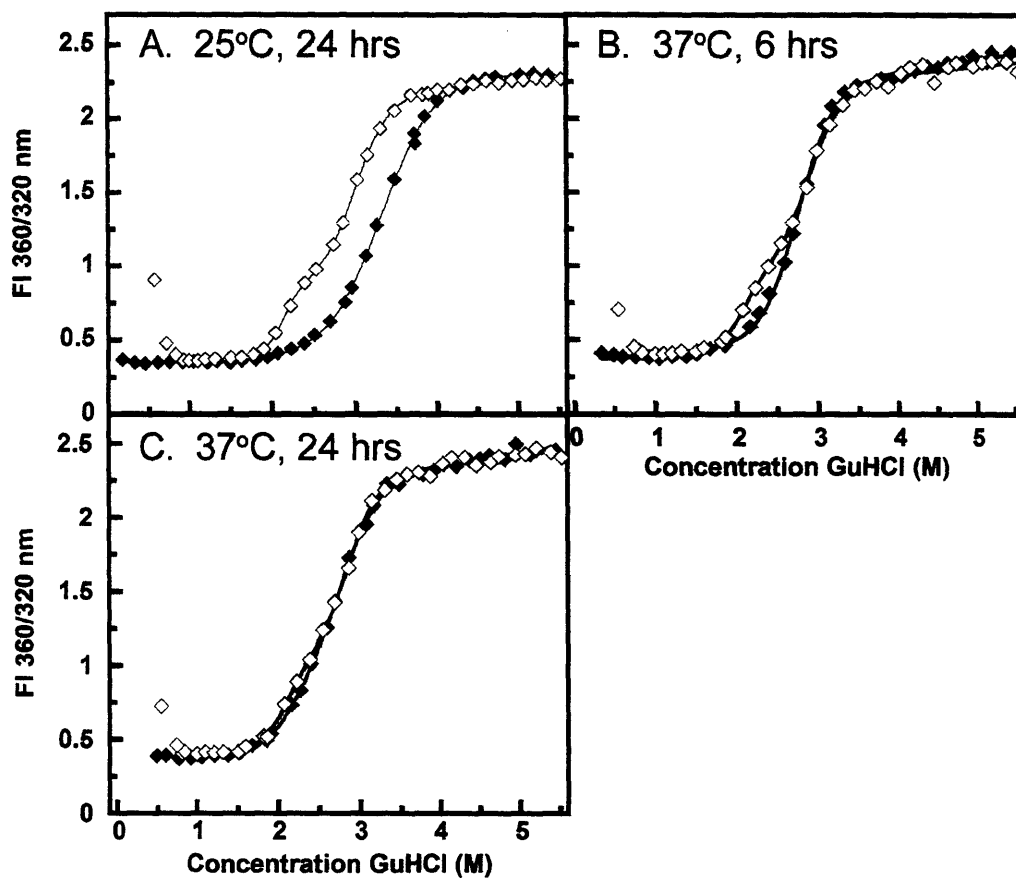


Figure 3-1. Equilibrium unfolding (closed symbols) and refolding (open symbols) experiments of γD_{WT} at different equilibration times and temperatures. Solid line indicates data fit.

(A) 24 hour equilibration, 25°C

(B) 6 hour equilibration, 37°C

(C) 24 hour equilibration, 37°C

2. *Equilibrium Unfolding/Refolding of Wild type and Isolated Domain Proteins*

We examined whether the equilibrium unfolding/refolding of the isolated domains exhibited a hysteresis. These experiments were carried out at a lower temperature of 18°C. As with the results at 25°C, $\gamma_{D_{WT}}$ had a hysteresis with a more extensive separation between the equilibrium unfolding and refolding reactions. $\gamma_{D_{WT}}$ required incubation for 192 hours (8 days) in order to achieve equilibrium. In contrast, $\gamma_{S_{WT}}$ and all of the isolated domains did not exhibit a hysteresis after the 24 hour equilibration time.

All four of the isolated domains as well as the full-length wild type proteins demonstrated reversibility indicating they had the ability to unfold and refold at 18°C (Fig. 3-2). Upon refolding out of denaturant $\gamma_{D_{WT}}$ exhibited an off-pathway aggregation reaction, visible in the elevated refolding points at low GuHCl concentration. The equilibrium unfolding of $\gamma_{D_{WT}}$ was best fit to a three state model indicating that an intermediate was still populated at 18°C similar to 37°C. $\gamma_{S_{WT}}$ was best fit to a two state model with a more cooperative transition. All of the isolated domains were best fit to two state models. Differential stability was still displayed at the lower temperature. However, the GuHCl concentration midpoint (C_M) increased about the same increment for every protein at 18°C compared to 37°C (Table 3-1). Taken all together, these properties of these proteins were comparable to the equilibrium unfolding/refolding experiments at 37°C.

3. *Unfolding Kinetic Analysis of the Wild Type and Isolated Domain Proteins*

In order to understand the kinetic barrier that exists in $\gamma_{D_{WT}}$, we tested individual domains to see if a large barrier to unfolding was associated with the N-terminus as predicted, or with the C-terminus or with both domains. $\gamma_{S_{WT}}$ and its isolated domains were compared to $\gamma_{D_{WT}}$ since equilibrium studies at lower temperature and at shorter equilibration times did not exhibit an unfolding hysteresis.

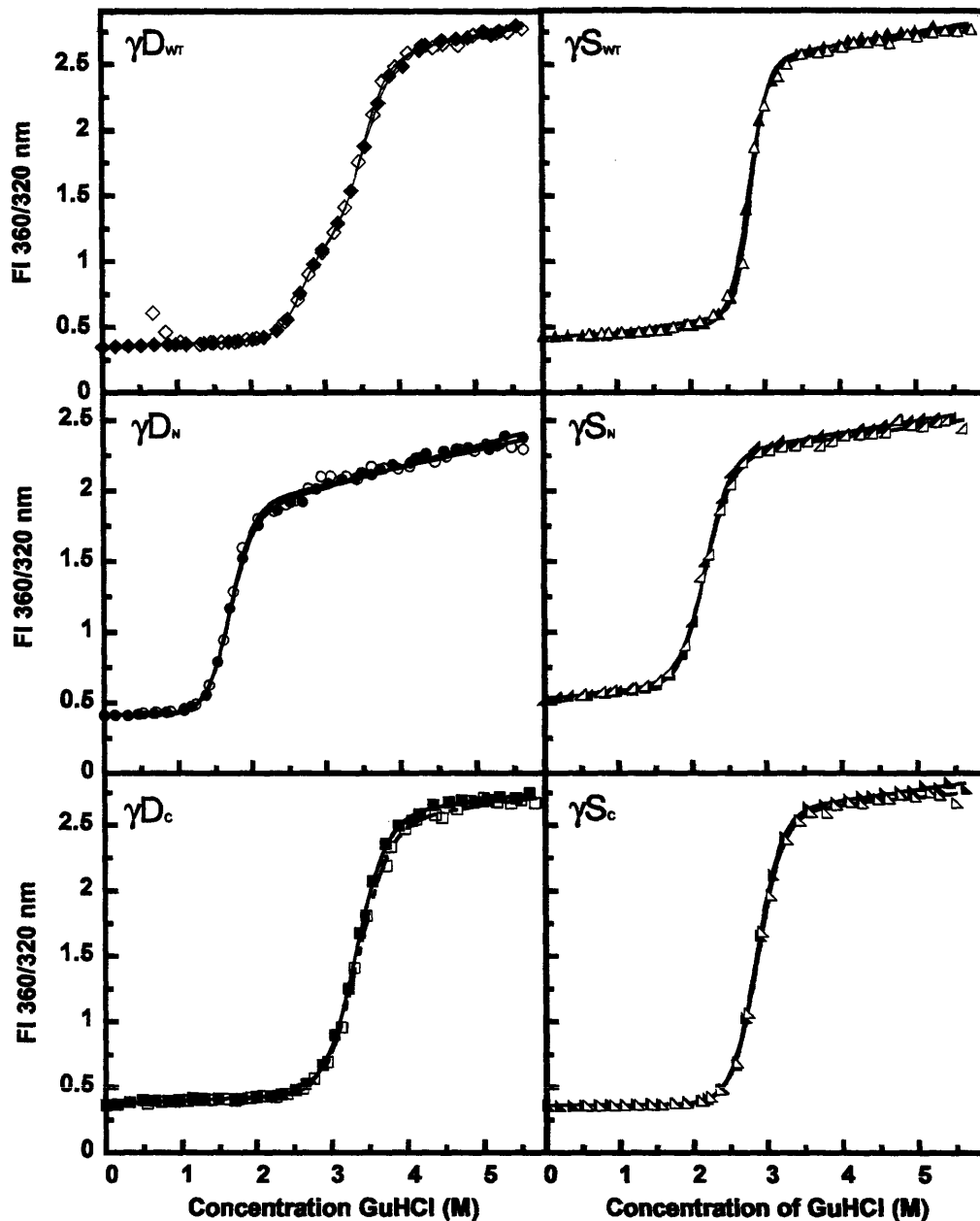


Figure 3-2. Equilibrium unfolding (closed symbols) and refolding (open symbols) for γD_{WT} (\blacklozenge), γD_N (\bullet), and γD_C (\blacksquare), γS_{WT} (\blacktriangle), γS_N (\blacktriangleleft), and γS_C (\blacktriangleright). Samples consisted of 10 $\mu\text{g}/\text{mL}$ protein concentration, 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and various concentrations of GuHCl at 18°C. Equilibration time 24 hours, except for γD_{WT} , which had an equilibration time of 192 hours. All proteins excited at 295 nm and emission at 360 nm and 320 nm calculated as a ratio. Equilibrium data fit indicated by solid black line.

Table 3-1. Equilibrium unfolding and refolding parameters for γ D and γ S crystallin wild type and isolated domains at 18°C.

Protein	Equilibrium Transition 1			Equilibrium Transition 2		
	C_M^a	Apparent m value ^b	Apparent $\Delta G_{N-U}^{\circ c}$	C_M^a	Apparent m value ^b	Apparent $\Delta G_{N-U}^{\circ c}$
	18°C					
γ D _{WT}	2.6 ± 0.035	4.0 ± 0.6	10.5 ± 1.5	3.5 ± 0.02	2.9 ± 0.03	10.1 ± 0.2
γ D _N	1.6 ± 0.08	3.5 ± 0.8	5.5 ± 1.3	-	-	-
γ D _C	-	-	-	3.2 ± 0.05	3.2 ± 0.3	10.4 ± 0.7
γ S _{WT}	2.7 ± 0.04	4.7 ± 0.7	12.6 ± 2.0	-	-	-
γ S _N	2.12 ± 0.03	3.3 ± 0.8	6.9 ± 1.7	-	-	-
γ S _C	2.7 ± 0.02	3.5 ± 0.4	9.6 ± 1.0	-	-	-

^a Transition midpoints in units of M GuHCl.

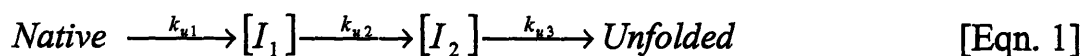
^b Apparent m values in units of kcal* mol^{-1} * M^{-1} .

^c Free energy of unfolding in the absence of GuHCl in units of kcal* mol^{-1} .

The unfolding kinetic experiments were performed by utilizing an injection port syringe containing each protein. The protein samples were injected into a GuHCl containing phosphate buffer, immediately recording tryptophan emission fluorescence after the dead time of ~ 1 s. The conformational changes in the protein were monitored by fluorescence spectroscopy, preferentially exciting the tryptophan residues at a wavelength of 295 nm and reading the emission at a wavelength of 350 nm.

Upon unfolding of the protein, native state quenching is alleviated creating an increase in tryptophan fluorescence emission. In order to monitor the fluorescence of the initial native state for the isolated domains, all kinetic experiments were performed at 18°C. We initially recorded kinetic data as the protein unfolded in 5.5 M concentration of GuHCl.

The kinetic unfolding experiments performed at 5.5 M GuHCl demonstrated that γD_{WT} protein was extremely stable requiring close to 3 hours to reach an unfolded baseline (Fig. 3-3, Table 3-2). Under these conditions, the data fit to a three state model, different from the four state model fit in our previous paper (Flaugh et al. 2006). This discrepancy may be due to not enough data to fit this initial burst phase or because the first two transitions at this temperature are too similar to resolve by this method. Previously using triple tryptophan mutants, it was shown that the γD_N domain unfolded completely before the γD_C domain (Kosinski-Collins et al. 2004). These results suggested that the following equation 1 (Eqn. 1):



In which, k_{u1} represented the partial unfolding of the N-terminal domain, k_{u2} represented complete unfolding of the N-terminal domain, and k_{u3} represented complete unfolding of the C-terminal domain. In these experiments, we suspect that k_{u1} and k_{u2} are hard to resolve at this temperature. The following equation for the two kinetic transitions (equation 2, Eqn.2) provided an adequate fit.



The first unfolding transition, k_{u1} , represented the complete unfolding of the N-terminus and k_{u2} represented complete unfolding of the C-terminus. The average half-life ($t_{1/2}$) for these two kinetic transitions was 251 s and 3455 s, respectively.

In comparison, $\gamma\text{S}_{\text{WT}}$ was completely unfolded within minutes at 5.5 M GuHCl concentration. The $\gamma\text{S}_{\text{WT}}$ kinetic data was best fit to a three state model, yielding one observable intermediate (Fig. 3-3, inset). The average calculated $t_{1/2}$ for k_{u1} and k_{u2} were 0.5 s and 37 s, respectively (Table 3-2). $\gamma\text{D}_{\text{C}}$ kinetic analysis was also performed at 5.5 M GuHCl concentration, demonstrating a two-state transition, with a calculated average $t_{1/2}$ of 1143 s (Fig. 3-3 and Table 3-2).

At temperatures from 37°C - 20°C most of the single domains unfolded partially or completely within the dead time of the experiment (~1 s). This result alone demonstrated that there was differential stability difference between the full-length γD and its isolated domain proteins. Even at the lower temperature, the $\gamma\text{S}_{\text{C}}$, $\gamma\text{D}_{\text{N}}$, and $\gamma\text{S}_{\text{N}}$ all were partially or completely unfolded at the high GuHCl concentration of 5.5 M, thus, all of these proteins were analyzed at 3.5 M GuHCl concentration. (The kinetic unfolding analysis was not performed at lower GuHCl concentration for $\gamma\text{D}_{\text{WT}}$ and $\gamma\text{D}_{\text{C}}$ since at 18°C these proteins are not completely unfolded.) $\gamma\text{S}_{\text{WT}}$ kinetic unfolding was performed at lower GuHCl and was fit to a three-state model with calculated average $t_{1/2}$ of 308 s for the first kinetic transition and a $t_{1/2}$ of 804 s for the second transition (Table 3-2).

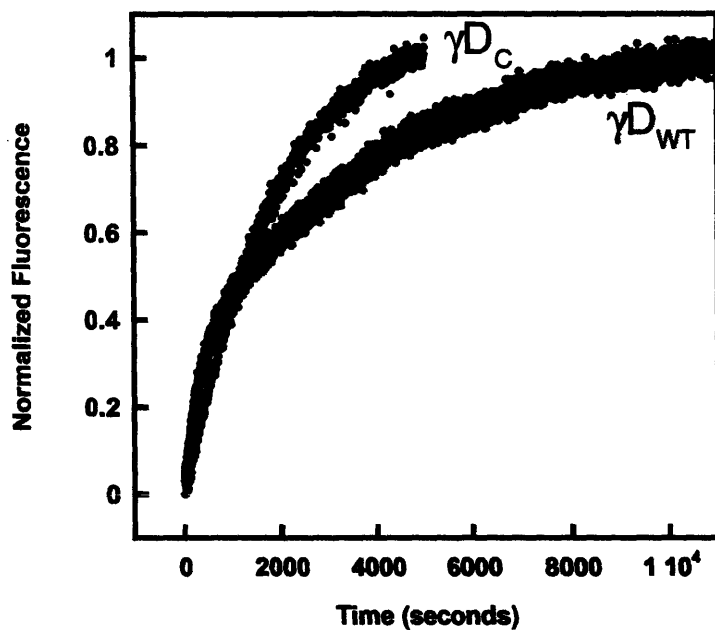
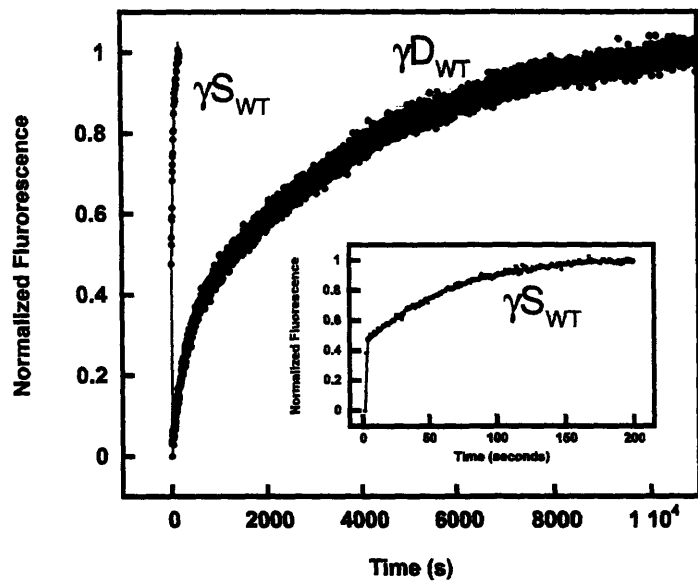


Figure 3-3. Kinetic unfolding of γD_{WT} and γS_{WT} at 5.5 M GuHCl at 18°C (Top). Inset is γS_{WT} refolding kinetics on a smaller time scale in order to observe the kinetic transitions. γS_{WT} (light blue) γD_{WT} (black)

Kinetic Unfolding of γD_{WT} and γS_{WT} at 5.5 M GuHCl at 18°C (Bottom).
 γD_{WT} (black), γD_C (red)

Protein fluorescence emission at 350 nm was recorded every second unless otherwise noted and all data normalized for comparison.

γS_{WT} kinetic data at 3.5 M GuHCl could also be fit to a two-state model; however, since at increasing concentrations of GuHCl the data exhibited a biphasic transition, this data was fit to a three state model as well (Fig. 3-3). In other words, at the higher GuHCl concentrations, the presence of an intermediate in γS_{WT} is observable by a biphasic transition, a fast first phase and a slower second phase. At the lower GuHCl concentration, the biphasic kinetic rates for the first and second phase represent only 2-fold difference, indicating that at this lower concentration, the intermediate observed in the higher GuHCl concentrations is not observable by this method. The residuals for both the two and three state model were similar.

Kinetic unfolding of γS_C , γD_N , and γS_N yielded differential kinetic stability, with the hierarchy of stability from most stable to the least stable in that order. Kinetic unfolding of γS_N , the least stable, was best fit to a three state model with an average calculated $t_{1/2}$ of 0.37 s for the first kinetic transition and 9 s for the second transition (Fig. 3-4, Table 3-2). Kinetic unfolding of γS_C , the most stable, was similar to γS_{WT} and was best fit to a two state model with a $t_{1/2}$ of 937 s (Fig. 3-4, Table 3-2). γD_N was slightly more difficult to fit, because of a consistent sharp increase then slower decrease in fluorescence within 50 seconds after monitoring the fluorescence emission (Fig 3-4). Because of the consistence of this phenomenon, we suspect that it is a result of a short-lived structured intermediate in which the structure surrounding Trp 42 and/or Trp 68 is partially relaxed to produce a slightly fluorescent intermediate. However qualitatively, we can predict that this intermediate has a $t_{1/2}$ of ~25 s. In order to fit the rest of the kinetic data, ~50 s of the data was deleted and was fit to a two-state model (Fig. 3-4). The results of this analysis yielded an average calculated $t_{1/2}$ of 385 s (Table 3-2).

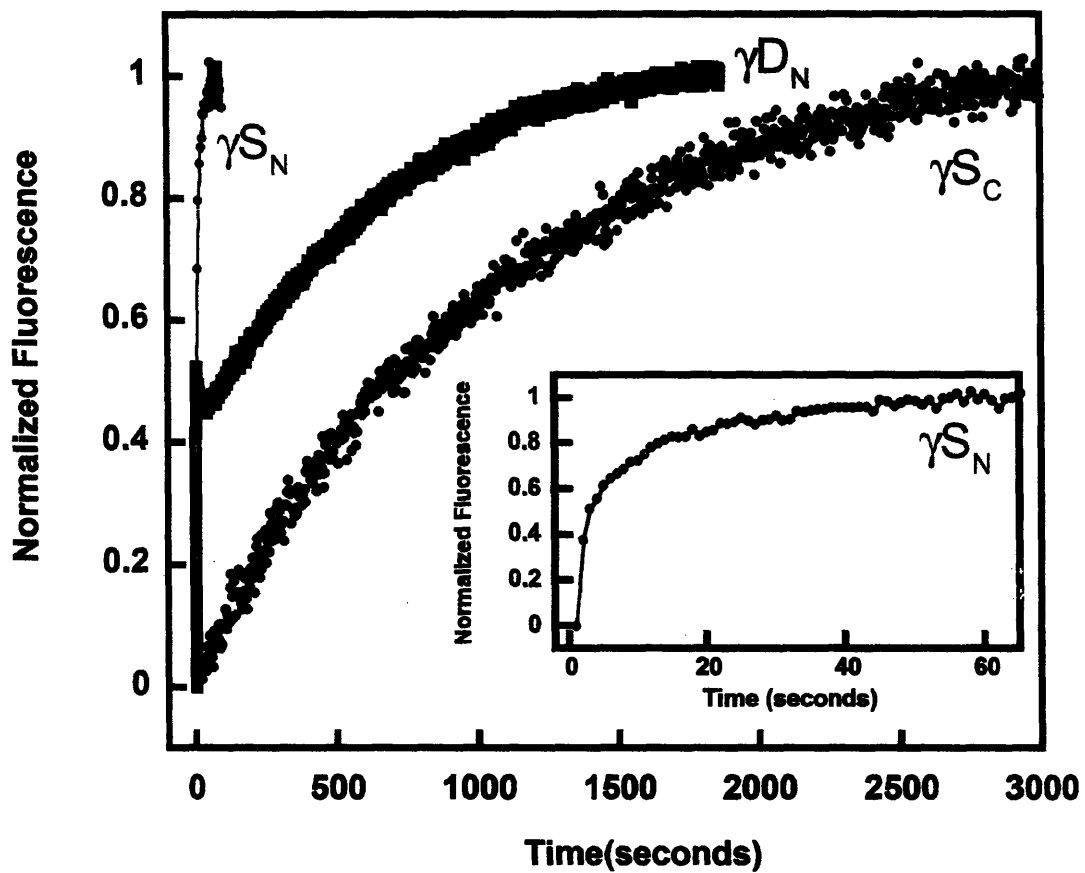


Figure 3-4. Kinetic unfolding at 3.5 M GuHCl at 18°C. (A) γS_N (orange) data taken in milliseconds (B) γS_C (green), γD_N (blue) (inset is completion of γS_C unfolding kinetics reaction)

Table 3-2. Kinetic unfolding parameters of γ D and γ S wild type and isolated domain proteins at 18°C.

Protein	Kinetic Unfolding (5.5M GuHCl)					Kinetic Unfolding (3.5M GuHCl)				
	k_{u1}^a	k_{u2}^a	$t_{1/2u1}^b$	$t_{1/2u2}^b$	-	k_{u1}^a	k_{u2}^a	$t_{1/2u1}^b$	$t_{1/2u2}^b$	-
γ D _{WT}	0.003 ± 0.00095	$0.0002 \pm 5.2E-05$	251 ± 111	3244 ± 944	-	-	-	-	-	-
γ D _N	-	-	-	-	-	-	$0.0018 \pm 6.8E-05$	-	385 ± 14	-
γ D _C	$0.00061 \pm 6.3E-05$	-	1143 ± 117	-	-	-	-	-	-	-
γ S _{WT}	1.39 ± 0.12	$0.018 \pm 9.6E-05$	0.5 ± 0.04	37 ± 0.2	0.0023 ± 0.0003	0.00087 ± 0.0001	308 ± 34	804 ± 98	-	-
γ S _N	-	-	-	-	2.0 ± 0.6	0.08 ± 0.003	0.37 ± 0.1	9.1 ± 0.4	-	-
γ S _C	0.02 ± 0.002	-	33 ± 3	-	$0.00074 \pm 7.0E-05$	-	937 ± 81	-	-	-

^a Kinetic unfolding rates in units of s^{-1} .

^b Half-life in units of seconds.

4. Half-Chevron Plot Analysis Predicts Longer Lived Life Times for γD_{WT} than for γS_{WT}

We determined extrapolated unfolded kinetic rates and half-lives for each protein in the absence of denaturant by performing unfolding kinetics at various GuHCl concentrations. The logarithmic unfolding kinetic rates versus denaturant concentration are graphed linearly extrapolating the kinetic rate at no denaturant (Schellman 1987). This “half-chevron” plot analysis allows for comparison of all proteins to determine if in the absence of denaturant environment, the proteins maintain their differential kinetic stability (Figs. 3-5 and 3-6). All of the experiments were performed at 18°C in order to obtain unfolding kinetic comparison data on all of the proteins. γD_{WT} extrapolated values to buffer was the most stable with a $t_{1/2}$ of 19 years for the first kinetic transition (k_{u1}) and $t_{1/2}$ of 129 days for the second kinetic transition (k_{u2}). Comparatively, the γS_{WT} extrapolated values were 1.6 years for the first kinetic unfolding transition and ~ 2 days for the second kinetic unfolding transition under these conditions (Table 3-3).

The stability hierarchy that was seen in the unfolding kinetic experiments among the single domains agrees with the extrapolated values with $\gamma D_C > \gamma S_C > \gamma D_N > \gamma S_N$. γD_C has an extrapolated value of ~15 years, slightly less stable than the full-length protein in buffer (Table 3-3). While γD_N extrapolated second kinetic transition $t_{1/2}$ is approximately ~3 hours. Again, it was difficult to calculate the first transition because of the fluorescent intermediate that occurs in the first ~ 50 seconds of unfolding; therefore, only the second kinetic transition was extrapolated. For the γS_{WT} isolated domains, γS_C was more stable than γS_N with a $t_{1/2}$ of 3 days. The γS_N $t_{1/2}$ for the first kinetic transition was 1.3 minutes and a $t_{1/2}$ of 26.6 minutes for the second kinetic transition (Table 3-3).

In γD_{WT} and γS_{WT} , the slopes vary for both transitions, with a steeper slope for the first transition versus the second. For γS_{WT} , by comparing the first transitions versus the second transitions in decreasing amounts of GuHCl, the rates of the first transitions decrease faster than the second transition rates. There is a ~600 fold decrease in the first transition rate from high to the low GuHCl. Comparatively, there is a 12 fold decrease in

the second transition rate from high to the low GuHCl (Table 3-2). This indicates that the GuHCl has a stronger effect on the first transition versus the second transition. The same is true for γD_{WT} . By comparing the kinetic rates of higher concentrations of GuHCl to 5.5 M GuHCl kinetic rates, γD_{WT} has a 13 fold decrease for the first transition rate and a 4 fold decrease for the second transition rate.

All of the extrapolated values are approximations and are not meant to portray any precise values for these proteins in their native environment especially since the experiments are performed at a lower temperature and not 37°C. In addition, traditionally chevron plots are not analyzed for multi-state kinetic proteins because the presence of an intermediate is not guaranteed to be 100%. However, although the fluorescence emission at 350 nm is monitoring a sum of all species, it should reflect the most abundant species. Using fluorescence spectroscopy, other studies have extrapolated multi-state kinetic values to no denaturant with apomyoglobin (Baryshnikova et al. 2005). Furthermore, the differential stability observed between two homologous proteins as well as their individual domains are qualitatively comparable between kinetic experiments utilizing GuHCl denaturant and the linearly extrapolated kinetic values.

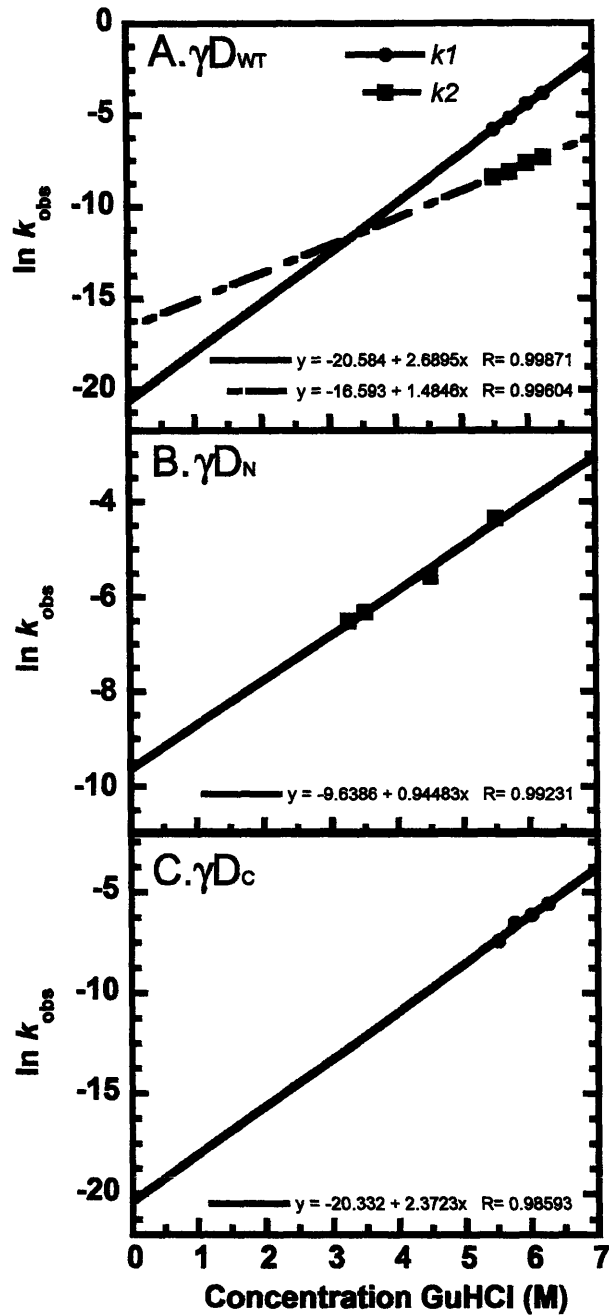


Figure 3-5. Linear extrapolations of kinetic unfolding rate constants versus GuHCl concentration for γD_{WT} and its individual domains. All experiments performed using syringe port injection apparatus and consisted of 10 μ g/mL protein concentration, 100mM sodium phosphate, 1mM EDTA, 5mM DTT, pH 7.0 buffer, various concentrations of GuHCl at 18°C. (A) γD_{WT} , (B) γD_N , and (C) γD_C

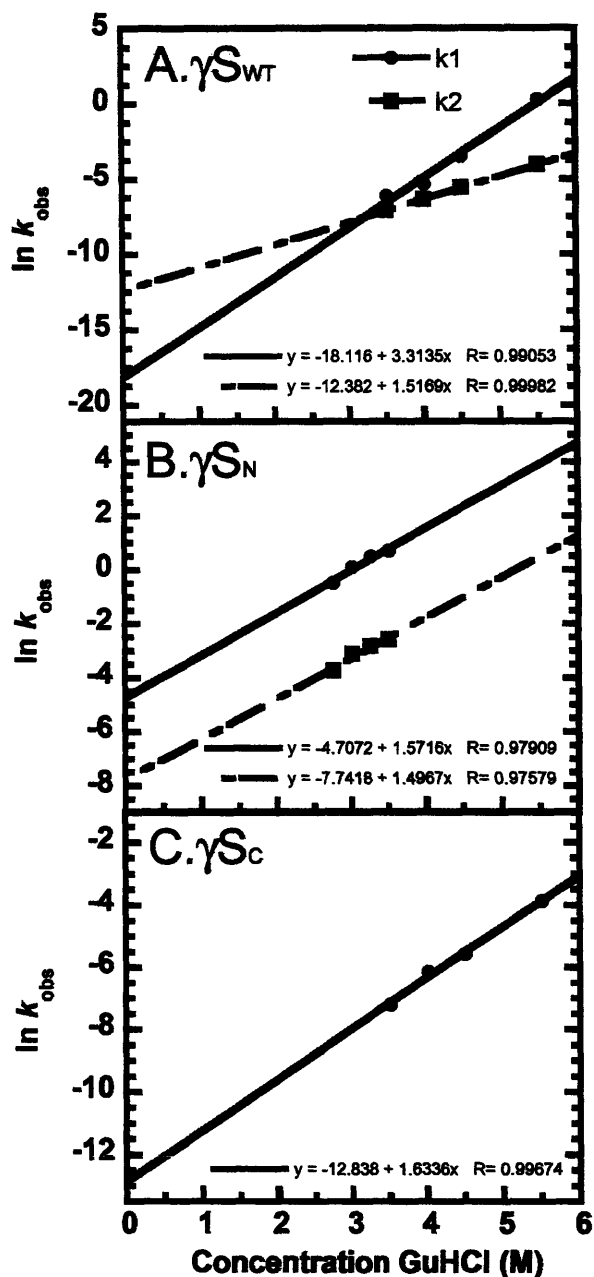


Figure 3-6. Linear extrapolation of kinetic unfolding rate constants versus GuHCl concentration for γS_{WT} and its individual domains. All experiments performed using syringe port injection apparatus and consisted of 10 μ g/mL protein concentration, 100mM sodium phosphate, 1mM EDTA, 5mM DTT, pH 7.0 buffer, various concentrations of GuHCl at 18°C. (A) γS_{WT} , (B) γS_N , and (C) γS_C

Table 3-3. Linear extrapolated unfolding rate constants and half-lives of γ D and γ S wild type and isolated domains at 18°C.

Protein	Kinetic Unfolding Transition 1		Kinetic Unfolding Transition 2	
	$k_{u1}^{H2O a}$	$t_{1/2}^{H2O b}$	$k_{u2}^{H2O a}$	$t_{1/2}^{H2O b}$
γ D _{WT}	1.15 E-09	19.1 yrs	6.22 E-08	129 days
γ D _N	6.52 E-05	2.95 hrs	-	-
γ D _C	1.48 E-09	14.9 yrs	-	-
γ S _{WT}	2.36 E-08	1.62 yrs	3.93 E-06	2.04 days
γ S _N	0.009	1.3 min	0.0004	26.6 min
γ S _C	2.66 E-06	3 days	-	-

^a Kinetic rates in units of s⁻¹.

^b Half-life in various units, min = minutes, hrs = hours, days = days, yrs = years.

DISCUSSION:

In this chapter, I have described in detail the kinetic differences in stability among γD_{WT} , γS_{WT} and their respective isolated domain proteins. Thus far only one isolated γ -crystallin domain has been found in nature in the sea squirt, *Ciona intestinalis*. This discovery raises questions as to why gene duplication was selected for in the $\beta\gamma$ -crystallin family within all extant lineages, and if this gene duplication is related to stability or longevity of the protein. In addition to comparing the isolated domains to their respective full-length two-domain protein, we compared two related but divergent γ -crystallins, γD and γS crystallins. The diverse kinetic stabilities of these γ -crystallins suggests that γD crystallin adapted to its environment by addition of a stronger interdomain interface as discussed in Chapter two. These improvements may be related to the location of these crystallins in the lens: γD_{WT} has greater requirement for stability because of its expression in primary lens fiber cells which are unable to synthesize new proteins. γS_{WT} exhibited reduced stability perhaps because of relaxed selection for kinetic stability given its continued expression in each layer of differentiating secondary lens fiber cells.

1. γD_{WT} but not γS_{WT} Demonstrate a Kinetically Controlled Hysteresis

The experiments showed that the distinct hysteresis observed in the equilibrium unfolding/refolding of γD reflected a high barrier to unfolding. The transition that was most effected was the unfolding of the N-terminal domain of γD_{WT} . This supports the view that the interdomain interface was also a source of kinetic stability as well as conformational stability (Chapter 2). A hysteresis related to kinetically controlled steps along the folding pathway has been observed in other proteins such as P22 tailspike, creatine kinase, collagen, Triosephosphate Isomerase (TIM), and transthyretin. The kinetic control of these transitions may also reflect the aspects of the protein's function. For example, in the mouse prion protein, there is a high kinetic barrier between the native α -helical form and the β -isoform preventing accumulation of the amyloidogenic β -isoform (Baskakov et al. 2001).

2. Kinetic Unfolding of the Full-length γ -Crystallins and Their Isolated Domains

Other members of the $\beta\gamma$ -crystallin super family of proteins, exhibit a high barrier to unfolding as well as their aforementioned thermodynamic stability. Microbial crystallins, Protein S from *Myxococcus xanthus* and Spherulin 3a from *Physarum polycephalum*, demonstrated kinetic stability. The different isolated domains of Protein S had similar $t_{1/2}$ values when analyzed while the full-length protein exhibited larger kinetic unfolding barriers (Wenk et al. 1998). These results suggested that the interface between the two domains contributed to the kinetic stability (Wenk et al. 1998). Also, Spherulin 3a showed a steeper kinetic barrier when bound to calcium. Linear extrapolation in the absence of denaturant yielded a $t_{1/2}$ that was 1.3 days versus 12 days without or with Ca^{2+} , respectively (Kretschmar et al. 1999a). In addition, bovine γF crystallin also has a high unfolding kinetic barrier, though no quantitative kinetic measurements have been performed (Das and Liang 1998).

In the β -crystallin family, MacDonald et al. showed that a mutation near the interface of rat βB2 N-terminal domain (C50F) caused an increase in the kinetic barrier to unfold (2005). The Cys 50 residue in the N-terminal of βB2 is implicated in the subunit exchange to form homo- and hetero-dimers with itself and other β -crystallins, contributing to the polydispersity of the crystallins. Therefore in this case, it is thought that the large barrier is lessened in order to maintain the functionality of βB2 oligomer formation.

$\gamma\text{D}_{\text{WT}}$ unfolding yielded the lowest kinetic unfolding rates, requiring ~ 3 hours to unfold completely in high concentration of GuHCl. Similarly, the equilibrium unfolding and refolding experiments of $\gamma\text{D}_{\text{WT}}$ in this study has shown that at lower temperatures the barrier to unfolding was increased, requiring eight days in order to reach equilibrium at 18°C . Under these conditions, the kinetic transitions were best fit to a three state model. Previous results have demonstrated sequential unfolding of the two domains in $\gamma\text{D}_{\text{WT}}$ (Kosinski-Collins et al. 2004). Thus, we predict that the first transition is monitoring the

unfolding of the N-terminal domain and the second transition monitors the unfolding of the C-terminal domain. If the first transition which has a half-life of over 19 years is monitoring the unfolding of the γD_N , we suspect that the interface residue interactions are the primary source of such high kinetic stability (Fig. 1-7).

Upon refolding out of denaturant, the γD_{WT} chains partition between productive refolding and a competing off-pathway aggregation. The partially folded intermediate may be an aggregation-prone species which associates via a domain swapping mechanism, a possible model for the aggregation of γD_{WT} *in vitro*. The high kinetic stability of the protein of this transition could prevent the protein from populating this aggregation-prone intermediate leading to insolubility. Kinetic analysis of Protein S, a microbial crystallin, also revealed a high kinetic barrier attributed to its domain interface (Wenk et al. 1998).

γS_{WT} unfolds substantially faster than γD_{WT} , with a decrease in the overall $t_{1/2}$ of ~90 fold compared to γD_{WT} , at 5.5 M GuHCl, 18°C. As discussed in Chapter two, the unfolding pathway of γS_{WT} has not been elucidated so it is difficult to compare its transition states with γD_{WT} . In addition, the kinetic unfolding transitions of γS_{WT} are difficult to compare to the transitions seen in the isolated domain kinetics due to the fact that these are smaller proteins and may not fold in the same manner as in the context of the full-length protein. However, one possible model for the unfolding of γS_{WT} is since the γS_N isolate domain unfolds considerably faster than γS_C isolated domain, the first transition may monitor the unfolding of the γS_N while the second transition monitors the unfolding of the γS_C . If our modeling of unfolding is correct, the first transition barrier would prevent the thermodynamically less stable N-terminal domain from unfolding. It also suggests that the interface residues though not as kinetically stable as the γD , do contribute to overall kinetic stability of the protein (Figure 3-7).

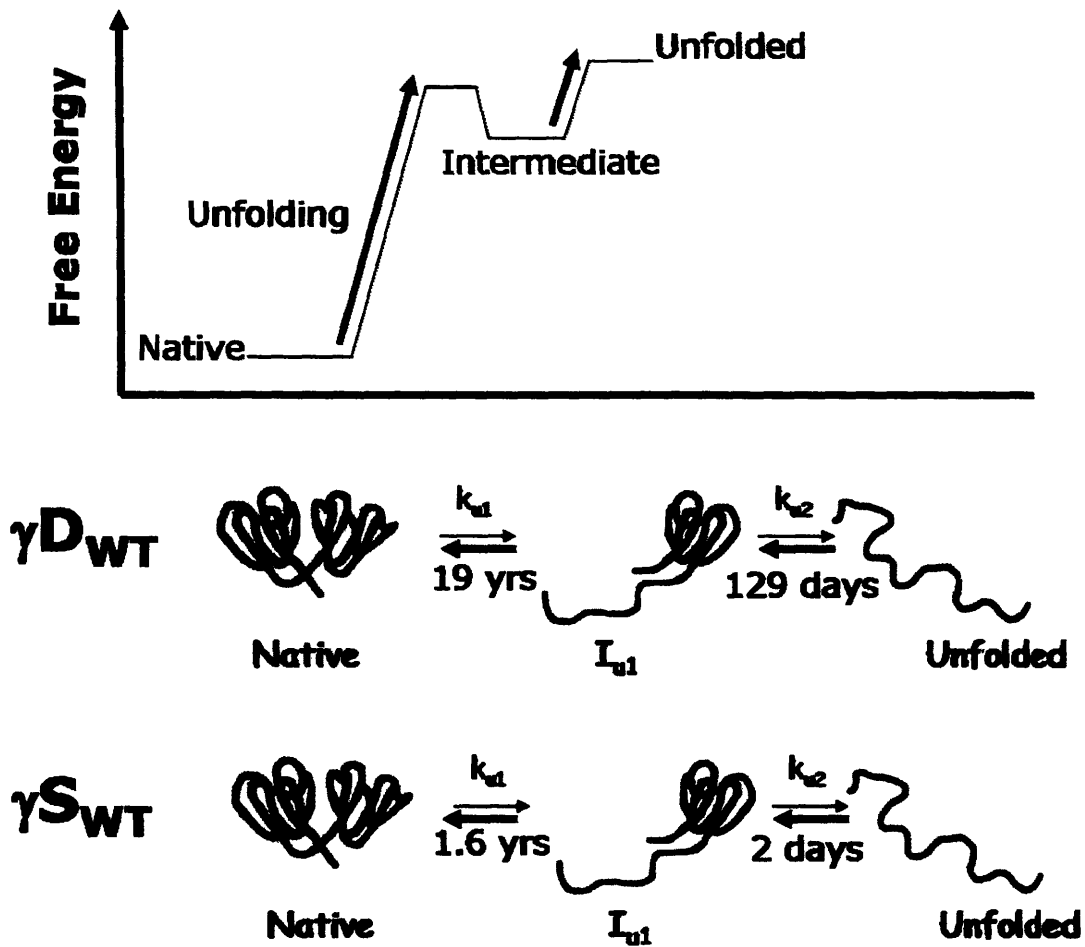


Figure 3-7. Model of γD_{WT} and γS_{WT} unfolding in **no denaturant** conditions. Linear extrapolation data suggests that the unfolding of the N-terminal domain has a high kinetic barrier to unfolding in both wild type proteins. Predicted half-time for each transition is noted.

The unfolding kinetics of the isolated domains represent a hierarchy in kinetic stability, with $\gamma D_C > \gamma S_C > \gamma D_N > \gamma S_N$. γD_C was the only isolated domain protein that could be monitored at high GuHCl (5.5 M) concentration; it was largely more stable than full-length γS_{WT} but significantly less stable than full-length γD_{WT} . The presence of a high fluorescence intermediate in the first kinetic phase interfered with its analysis. Additional biophysical and structural analysis of the γD_N should be performed to determine the conformation of this fluorescent intermediate. An interesting observation is that the conformational stabilities of the two C-terminal domains are similar with ΔG 's of 8 kcal* mol^{-1} , but have diverse kinetic stabilities, $t_{1/2}$ of 14 years versus 3 days.

3. Basis for Kinetic Stability in Crystallins

Why does the β and γ -crystallin superfamily have such high kinetic stability? One might hypothesize that a high barrier to unfolding is important to prevent exposure to regions of the protein which are aggregation prone or prevent increased solvent accessibility to regions of the protein that can be post-translationally modified. Studies of these post-translational modifications in the interface *in vitro* have been suggested to be deleterious to the stability and solubility of the γD , $\beta B2$, and $\beta B1$ -crystallins (Flaugh et al. 2006; Harms et al. 2004; Lampi et al. 2006; Wilmarth et al. 2006). In this study, we have compared the kinetic stability of both γ -crystallins and their isolated domains to gain insight into the intrinsic stability of the domains and the overall kinetic stability of the full-length protein.

The properties important for high kinetic barriers are poorly understood and have been difficult to decipher. In this case, perhaps non-local and local interactions are involved in the kinetic stability of these proteins. However, some common properties of proteins that have exhibited high kinetic stability have included β -sheet, oligomeric, protease and SDS resistant proteins (Manning and Colon 2004). Other high kinetic stability proteins include metal binding properties such as superoxide dismutase (Lynch et al. 2004). In addition, increasing surface hydrophobic residues have been shown to increase kinetic stability in α -amylase (Machius et al. 2003). Increased relative contact

order, a measure of complex topology, has been implicated in the folding rates of the proteins (Jung et al. 2005; Plaxco et al. 1998). This property may also be correlated to unfolding rates. The complex topology of the Greek Key crystallins and high relative contact order of the $\beta\gamma$ -crystallins have been suggested to be important in their kinetic stability (MacDonald et al. 2005).

However, it is perplexing that two structurally homologous proteins $\gamma\text{D}_{\text{WT}}$ and $\gamma\text{S}_{\text{WT}}$ with similar topologies demonstrate such diverse kinetic properties. Many of the kinetic models have been proteins that displayed two state kinetics and equilibrium. The complex multi-state kinetics demonstrated by $\gamma\text{D}_{\text{WT}}$ and $\gamma\text{S}_{\text{WT}}$ may add to the complexity of this theory. In addition to relative contact order which take into account the local and non-local contacts in the native structure of the protein, local and non-local contacts in the folding or unfolding intermediates of these proteins may be important. These contacts might not be obvious in the solved native structure of the protein. The essential contacts in protein folding and unfolding intermediates would be difficult to determine and calculate, without additional data for example, two-dimensional NMR, on the conformation of the intermediates.

4. Evolutionary and Physiological Relevance for the Increased Kinetic Stability of the γD -Crystallin

Homologous crystallins, γD and γS , are evolutionarily divergent. γS is thought to have evolved before $\gamma\text{A-F}$ due to its presence in all vertebrates (Van Rens et al. 1991). Upon the divergence of the ancestral single domain crystallin, a gene duplication and fusion event is thought to have occurred (Lubsen et al. 1988). This gene duplication event provided the crystallin with an additional domain. Although one domain is less intrinsically stable than the other, the creation of the domain interface between the two symmetrical domains added a kinetic barrier. As organisms increased in size and lifespan, the creation of this kinetic barrier became increasingly important. The prevalence of the $\gamma\text{A-F}$ crystallins in mammalian organisms supports this idea. These

studies demonstrate that the interdomain interface contributes to the kinetic stability of the crystallin in addition to thermodynamic stability as studied in Chapter two.

In addition, the biophysical properties may be related to the location of the crystallins in the human lens. The embryonic crystallins localized in the lens nuclear core must remain stable from early lens development throughout life to maintain transparency and prevent insolubility. The cortical crystallins are expressed after birth and are continually expressed throughout life. Whether the crystallins were recruited due to their stability properties or if they adapted to the changing environment and lifespan of the organisms is unclear. However, it seems likely that the crystallins were originally recruited for their stress-related roles and the structural characteristics of the γ -crystallins allowed for adaptation to the organism and environment.

We observed half-times close to 20 years for a lens nuclear core crystallin in these studies. High viscosity environments have demonstrated effects on the folding and unfolding rates of protein (Bieri and Kiefhaber 1999). It has been proposed that the viscosity of the lens would affect the rates of unfolding by increasing the kinetic stability of the crystallins (Bloemendal et al. 2004). Macromolecular crowding may also contribute to the stability of the protein since at this high concentrated environment, α -chaperone binding to these intermediates may be enhanced (Bloemendal et al. 2004; Minton 2000).

Although there is a high concentration of crystallins in the lens, partially folded intermediates would be difficult to populate since the equilibrium will drive intermediate state molecules to the native state (Fig. 3-7). Thus, partially folded intermediates such as one domain folded, and one domain unfolded would not be expected to unfold unless destabilized by post-translational mutations that may be age-related. Therefore, although the half-lives of these proteins seem shorter than what would be required, it would take longer periods to produce a concentration of intermediates that may have detrimental effects.

In the following chapter, I will try to identify the regions of these intermediates that contribute to the aggregation process by creating chimeras of γ D and γ S crystallins and studying their properties.

CHAPTER FOUR:

**DOMAIN-EXCHANGED HUMAN GAMMA D AND S CRYSTALLIN
CHIMERAS DEMONSTRATE DIFFERENTIAL AGGREGATION
PROPERTIES**

A. INTRODUCTION:

As discussed in Chapter one, misfolding, aggregation, or deposition of specific proteins is a major cause of several diseases. The insoluble fractions removed by surgery from cataractous lenses are composed of multiple species of the crystallins including γ D and γ S (Hanson et al. 2000; Lampi et al. 1998; Wilmarth et al. 2006). These molecules are tightly interacting with each other and require strong denaturants for solubilization, a feature of protein aggregates and inclusion bodies. This chapter compares the aggregation properties of two γ -crystallins, γ D and γ S crystallin, which are abundant γ -crystallins in the human lens. As mentioned before, γ D is localized in the lens core region while γ S is localized primarily in the outer region of the lens. Additionally, these proteins are highly homologous structurally and in sequence. Previously, refolding studies *in vitro* identified an off-pathway polymerization or aggregation reaction for γ D upon dilution out of denaturant competing with productive refolding (Kosinski-Collins and King 2003). Surprisingly, although structurally similar, off-pathway polymerization has not been observed for γ S under the same conditions. These results suggest that there are particular amino acid interactions that may be involved in the formation of the aggregate.

As described in Chapter two, other dissimilarities between γ D and γ S are features of their unfolding and refolding pathways. γ D crystallin populates a partially folded intermediate along its folding pathway in which the C terminal domain is folded and the N terminal domain is unfolded. γ S crystallin folds more cooperatively than γ D and does not populate this partially folded intermediate during equilibrium folding/unfolding. The partially folded intermediate of γ D with C-terminal domain folded and the N-terminal domain unstructured was also populated during kinetic refolding experiments (Flaugh et al. 2006; Kosinski-Collins and King 2003). The kinetic analysis of γ S also suggested the presence of partially folded intermediates along the refolding pathways; but, whether the conformation of these intermediates resembles γ D remains to be determined. These differences may have an effect on the aggregation properties as well.

Other studies have identified particular sequences in homologous proteins that lead to aggregation in one protein but not in the other. In the case of the bovine and human growth hormones, the refolding of the bovine hormone *in vitro* is limited by association of a partially folded intermediate. On the contrary, human growth hormone did not exhibit an off-pathway aggregation reaction during its refolding (Brems and Havel 1989). By engineering chimera proteins in which one of the α -helices - the aggregation prone sequence - in the bovine growth hormone was replaced with the corresponding human growth hormone sequence, the aggregation sharply decreased (Brems et al. 1988; Lehrman et al. 1991).

The experiments reported in this chapter involved constructing domain-exchanged chimeras of these two proteins in order to identify the particular sequences of γ D promoting aggregation or the sequences in γ S which may inhibit aggregation. The γ D aggregation reaction may be a useful model *in vitro* in understanding cataract formation in the lens. Further characterization of the aggregation pathway may lead to the identification of small molecules that inhibit this reaction (Blanchard et al. 2004).

B. MATERIALS AND METHODS:

1. Expression and Purification of Proteins

At the time of creating the chimeras, no full length structures of γ S were available. Therefore, it was difficult to predict what region corresponded to the N-terminal domain and the linker of γ S. The chimeras were created based off of the alignment with γ D and γ S C-terminal domain structural information that was available. The γ D_N- γ S_C chimera consisted of γ D_N G1-I81; γ S_C L87-E177. The γ S_N- γ D_C chimera consisted of γ S_N S1-H86; γ D_C P82-S174 (Fig. 4-1). The γ D_N- γ S_C and γ S_N- γ D_C chimera γ S interface residues were mutated to restore the γ D interface. These chimeras are referred to their original names plus [γ D_{INT}] to refer to the replacement of the γ S interface residues to the corresponding γ D interface residues. All chimeras were cloned into the

pQE1 vector which has a His tag including a N-terminal Met (Qiagen). All vectors were transformed into *E. coli* M15[pRep4] cells (Qiagen), utilized for tightly regulated protein expression. The cells were purified according to the methods outlined in Chapter 2. The protein purification protocol produced proteins with a purity of >90% confirmed by SDS-PAGE.

Protein concentrations were determined by unfolding of proteins in 5.5 M GuHCl and measuring absorbance at 280 nm using their respective protein extinction coefficients; $\gamma_{D_{WT}}$, $\gamma_{S_{WT}}$, $\gamma_{D_N-\gamma_{S_C}}$, $\gamma_{S_N-\gamma_{D_C}}$, $\gamma_{D_N-\gamma_{S_C}}$, $\gamma_{D_{INT}}$, and $\gamma_{S_N-\gamma_{D_C}}$, $\gamma_{D_{INT}}$, 41,040 $\text{cm}^{-1}\text{M}^{-1}$, 41,040 $\text{cm}^{-1}\text{M}^{-1}$, 39,790 $\text{cm}^{-1}\text{M}^{-1}$, 42,320 $\text{cm}^{-1}\text{M}^{-1}$, 41,745 $\text{cm}^{-1}\text{M}^{-1}$, 44,725 $\text{cm}^{-1}\text{M}^{-1}$, respectively.

2. Analytical Size Exclusion Chromatography

All samples were prepared by diluting to a final protein concentration of 80 $\mu\text{g}/\text{ml}$ in 10 mM Ammonium Acetate buffer, pH 7.0. The column was equilibrated with 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer. The samples were loaded onto the Superdex™200 10/300 GL (Pharmacia Biotech) column using a FPLC apparatus (Pharmacia). Absorption at 280 nm was monitored to determine the relative amount of protein. Molecular weight standards were utilized to determine relative elution times for various protein sizes. The molecular weight standards were Bovine Serum Albumin (BSA) (66 kDa), Ovalbumin (45 kDa), Chymotrypsinogen A (25 kDa), Ribonuclease A (13.7 kDa). Overlays of each spectrum were made using the Unicorn and Kaleidagraph programs. Peak fractions were collected and SDS-PAGE analysis confirmed the presence of each protein.

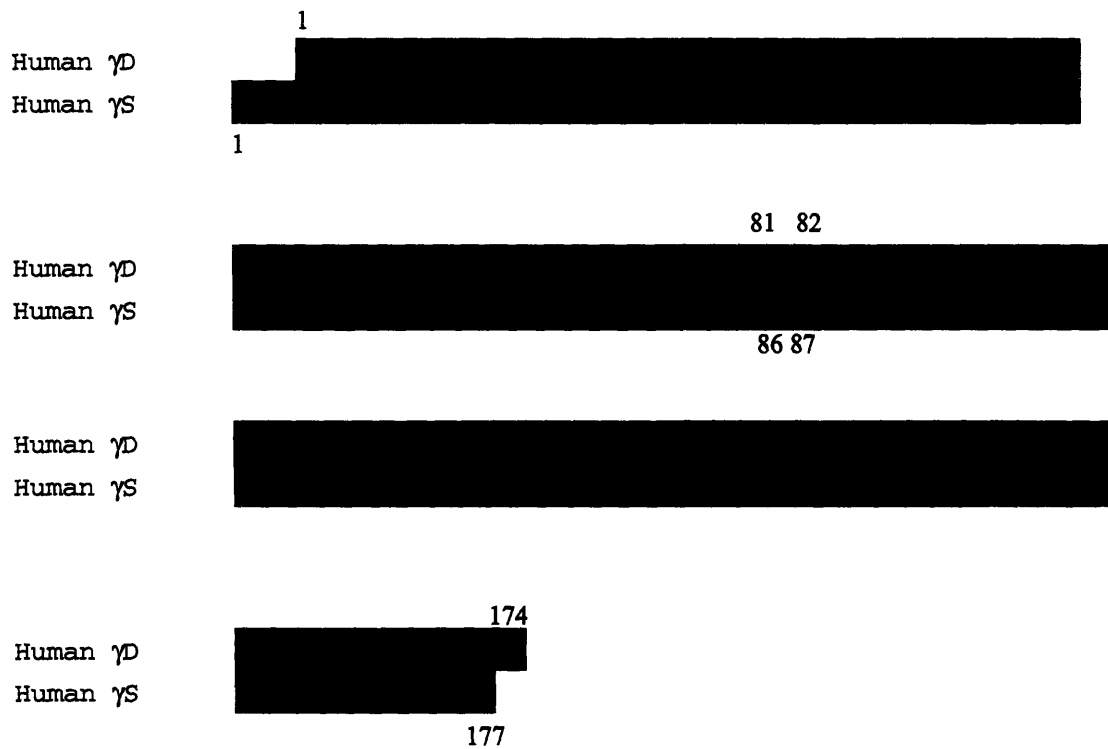


Figure 4-1. Sequence alignment of γ D_{wt} and γ S_{wt} proteins, 50% identity and 69% similarity. In green is the sequence of the γ D_N- γ S_C chimera. In gray is the sequence of the γ S_N- γ D_C chimera. The interface residues are in bold and underlined. The linker sequences are also underlined.

3. *Circular Dichroism Spectroscopy*

All experiments were performed using the same conditions as in Chapter 2. Deconvolution of the CD spectrum was performed using CDPro suite software package consisting of the CONTILL, CDSSTR, and SELCON3 programs (<http://lamar.colostate.edu/~sreeram/CDPro/>) (Sreerama and Woody 2000). As with the isolated crystallin domains, the IBasis 1 parameter (Johnson 1999) was chosen since it analyzed a larger range for the recorded data, had a large reference set (29 proteins) and had a lower RMSD and NRMSD for most of the data. The results from each program were averaged to obtain the overall secondary structure percentages.

4. *Fluorescence Spectroscopy*

Fluorescence emission spectra were taken using a Hitachi F-4500 fluorimeter equipped with a temperature control circulating water bath to maintain 37°C. A protein concentration of 10 µg/ml was added to 100 mM phosphate, 1 mM EDTA, and 5 mM DTT, buffer at pH 7.0 and 5.5 M GuHCl for the unfolded protein samples. All proteins were excited at 295 nm and the emission fluorescence was recorded from a wavelength range of 310 – 400 nm. The fluorimeter parameters were the same parameters described in Chapter 2.

5. *Thermal Denaturation*

All experiments were performed on the Aviv Model 202 CD spectrometer with an internal Peltier Thermoelectric controller (Lakewood, NJ) according to the previously described parameters (Chapter 2). In short, all proteins were at a 100 µg/ml concentration in a degassed 10 mM phosphate buffer, pH 7.0. Specifically, decrease in the β-sheet secondary structure minimum at 218 nm versus increase in temperature was monitored. Fraction native of the protein was calculated due to aggregation of the crystallins at high temperatures.

$$F_N = (y - y_U) / (y_N - y_U)$$

Where F_N = Fraction Native, y = Ellipticity at 218 nm, y_U = the unfolded/aggregation baseline, y_N = the native baseline. All experiments were repeated three times to acquire calculate averages and standard deviation.

6. *Equilibrium Unfolding and Refolding*

Equilibrium unfolding samples were performed in accordance with previous method described in detail in Chapter 2. Briefly, for equilibrium unfolding experiments, the protein was diluted to 10 $\mu\text{g/ml}$ with increasing concentrations of GuHCl (0 M – 5.5 M) in 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer (Guanidine Hydrochloride solution, 8 M (GuHCl), Sigma®). For equilibrium refolding experiments, a 10X protein solution was unfolded at 5.5 M GuHCl for 5 hours. The unfolded protein was then diluted 10-fold into various concentrations of GuHCl (0 M-5.5 M) giving a lowest GuHCl concentration of .55 M GuHCl. All samples had a 24 hour equilibration. Emission spectrum was recorded from wavelength range 310 – 400 nm and corrected for buffer. 360 nm data was used for calculating m and ΔG_{H_2O} values. 360/320 nm is shown for visual clarity of equilibrium transitions. Each experiment was repeated three times to determine the average and standard deviation parameters.

7. *Solution Turbidity Measurements*

Each chimera and wild type crystallins were unfolded at a 100 $\mu\text{g/ml}$ (10X) concentration in 5.5 M GuHCl, 100 mM Sodium Phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 at 37°C overnight. Each 10X protein solution was subsequently diluted into the 100 mM Sodium Phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer and incubated at 37°C for 24 hours as in the equilibrium refolding experiments. The Cary UV spectrophotometer was used to monitor the accumulation of high molecular weight species by light scattering at 350 nm. A fluorescence emission scan under the same parameters as the equilibrium experiments was performed to monitor right angle scattering of the sample. Each sample was subsequently centrifuged at high speed (12,000 rev/min) for 30 minutes to pellet the aggregate. Solution turbidity readings and

fluorescence emission scans were performed after centrifugation for comparative purposes.

8. Aggregation Kinetics

All proteins were unfolded overnight at 37°C according to the same method previously described for the equilibrium refolding experiments. The 10X samples were then quickly added into 100 mM Sodium Phosphate, 1 mM EDTA, 5 mM DTT buffer, pH 7.0. Solution turbidity was monitored over 30 minutes using the Cary UV spectrophotometer at 350 nm immediately after addition. Temperature at 37°C was maintained by a Cary single cell peltier. Every experiment was repeated at least three times to confirm the data. UV absorption data at 350 nm was corrected for absorption of each native protein.

9. Infrared (IR) Spectroscopy

All proteins were unfolded as previously described. The proteins were subsequently refolded in the D₂O 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer but at higher protein concentrations in order to obtain a signal by IR. The protein was in a final protein concentration of ~10 mg/ml. At these high concentrations, the aggregate was visible and was centrifuged at high speed for 30 minutes. The buffer was removed and the aggregated washed with D₂O buffer. The aggregate was pelleted again and resuspended in 20 µl of deuterated buffer. The sample was carefully added to an ATR (Attenuated total reflection) cell used for solid state samples and a Vector 22 Fourier transform Infrared (FTIR) spectrometer was used to obtain measurements. Temperature at 37°C was maintained by using a Peltier controller. 512 scans were taken and averaged to obtain the final protein interferogram.

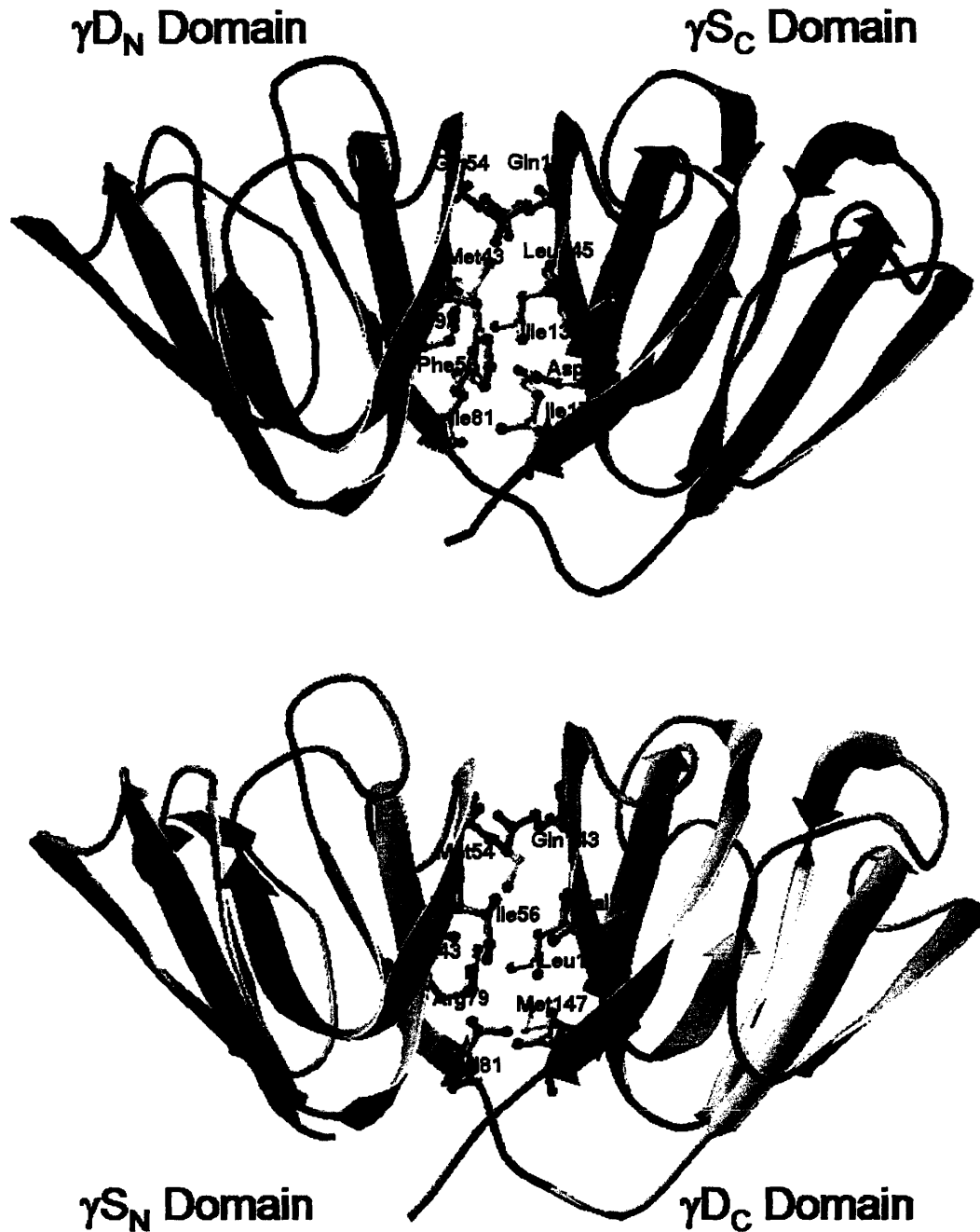


Figure 4-2. Homology modeling of the γ_{D_N} - γ_{S_C} and γ_{S_N} - γ_{D_C} chimeras based on the $\gamma_{D_{WT}}$ structure (PDB ID: 1HK0). The γ_S domains of the chimeras were modeled from the γ_D structure; the γ_D domains of the chimeras are the same as the published structure (Basak et al. 2003). The red regions represent regions that were difficult to thread onto the $\gamma_{D_{WT}}$ structure. Interface residues are represented in ball and stick and are noted based on the alignment with $\gamma_{D_{WT}}$ structure.

C. RESULTS:

1. Preparation of Chimera Proteins

As in previous chapters, I will refer to the natural γ D and γ S sequences as γ D_{WT} and γ S_{WT} for further clarity. The polymerization reaction competing with productive refolding of γ D_{WT} has been characterized by Atomic Force Microscopy (AFM) and by spectroscopic analysis (Kosinski-Collins and King 2003). In brief, by monitoring solution turbidity spectroscopically, γ D_{WT} diluted out of denaturant formed a high molecular weight species within seconds. Additionally, the γ D_{WT} aggregate specifically bound the hydrophobic dye, Bis-ANS, indicating that the aggregate included exposed hydrophobic sites. Electron micrographs of the aggregate revealed that it had a fibrous structure and was not amorphous. Time course AFM studies demonstrated that the formation of the aggregate consisted of several phases starting from small globular structures to protofibrils and subsequently arranging into branched and unbranched fibril bundles similar to the amyloid pathway. However, the fibrous aggregate was not in the characteristic amyloid cross β -sheet conformation, as it did not bind Congo red or display birefringence. In contrast, incubation of γ D_{WT} at pH 3 at 37°C, formed amyloid fibrils confirmed by Congo Red binding, electron microscopy, X-ray diffraction and infrared spectroscopy (Papanikolopoulou et al., unpublished results).

Given the aggregation results *in vitro*, a particular sequence in γ D_{WT} may direct aggregation or a particular sequence of γ S_{WT} may inhibit aggregation. The sequence identity and similarity between γ D_{WT} and γ S_{WT} is 50% and 69%, respectively. The sequence alignment of the two proteins gave regions of disparity dispersed throughout the proteins (Fig 4-1). Therefore, no particular sequences emerged as crucial in promoting or inhibiting the off-pathway aggregation reaction. One way to approach identifying key regions with minimum protein disruption is to create chimeras with one domain from γ D_{WT} and the other domain from γ S_{WT}, such as the N-terminal domain of γ D_{WT} and the

C-terminal domain of γS_{WT} (noted as γD_N - γS_C) and vice versa, N-terminal domain γS_{WT} and C-terminal domain of γD_{WT} (noted as γS_N - γD_C).

In constructing the chimera proteins, the protein alignments between γD_{WT} and γS_{WT} were taken into consideration (Fig. 4-1). Concerns over interfering with protein folding and loss of stability were addressed by including the linker sequence belonging to the protein whose C-terminal domain was used. Since previous results determined that the γD C-terminal domain (γD_C) was important in the folding of the N-terminal domain, the entire γD_C and linker was maintained in the γS_N - γD_C chimera. The opposite chimera, γD_N - γS_C was created by γD_{WT} N-terminal domain (γD_N) connected by the γS_{WT} linker to the C-terminal domain (γS_C).

The Swiss PDB threading program was used to predict the structure of the two chimeras based on the structural homology with the γD_{WT} structure (PDB ID: 1HK0). The areas in red indicated the regions difficult to thread onto the γD_{WT} structure (Fig. 4-2). These areas were primarily in the linker due to the 2 amino acid longer linker (His-Leu) in γS_{WT} . The murine γS structure indicates that the linker residues do not interact directly with the domains but the linker has been proposed to lower the entropic factor during interdomain interface interactions (Wu et al. 2005). The N-terminal extension of γS_{WT} , consisting of four amino acids (Ser-Lys-Thr-Gly) was not modeled to the structure since it is absent in γD_{WT} . According to the murine γS structure, the N-terminal extension of γS_{WT} is unstructured and has little interaction with the rest of the N-terminal domain (Wu et al. 2005).

Since previous evidence indicated the importance of the domain interface in stability, interface interactions were retained in the structure (Fig.4-2 and Table 4-1). There were no obvious clashes in the potential interface residue interactions of the chimeras. The γD_{WT} interface consists of six hydrophobic residues, three in each domain (M43, F5, I81:V132, L145, V170) and two pairs of peripheral residues (Q54 and Q143, R79 and M147) on each side of the hydrophobic interface. In γD_{WT} , the hydrophobic

residues form tight interactions between the two domains. The peripheral residues are thought to act as barriers to prevent exposure of the hydrophobic region of the interface to the solvent (Flaugh et al. 2005b). In $\gamma_{\text{S}_{\text{WT}}}$, the hydrophobic pairs consist of homologous residues according to the alignment of $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$, along with another hydrophobic pair not seen in the $\gamma_{\text{D}_{\text{WT}}}$ structure. In murine $\gamma_{\text{S}_{\text{WT}}}$, these hydrophobic residues are T45, A47, I60, and V85 in the $\gamma_{\text{S}_{\text{N}}}$ domain and T135, I137, L150, and V175 in the $\gamma_{\text{S}_{\text{C}}}$ domain. In human $\gamma_{\text{S}_{\text{WT}}}$, T135 is replaced by a valine and V175 is replaced by an isoleucine. The peripheral residues are M58 and Q148 and R83 and D152 (Wu et al. 2005). The peripheral Arg and Asp may form an ionic or hydrogen bond contributing to stability.

Table 4-1 indicates the composition of the interface in all of the chimera proteins. For simplicity, the numbering of the positions refers to the $\gamma_{\text{D}_{\text{WT}}}$ crystal structure (PDB ID: 1HK0) and does not include the other hydrophobic interaction in $\gamma_{\text{S}_{\text{WT}}}$. Major differences between the $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$ interfaces include the Phe at position 56 which is Ile in $\gamma_{\text{S}_{\text{WT}}}$, substituting a less bulky hydrophobic residue. At positions 81 and 170, the Ile and Val interface pairs are switched to Val and Ile in $\gamma_{\text{S}_{\text{WT}}}$. These interface differences in the chimeras for $\gamma_{\text{S}_{\text{N}}}$ - $\gamma_{\text{D}_{\text{C}}}$ resulted in a Val-Val interaction. In $\gamma_{\text{D}_{\text{N}}}$ - $\gamma_{\text{S}_{\text{C}}}$, the domain-exchange resulted in Ile-Ile interactions in this region.

In addition to creating the $\gamma_{\text{D}_{\text{N}}}$ - $\gamma_{\text{S}_{\text{C}}}$ and $\gamma_{\text{S}_{\text{N}}}$ - $\gamma_{\text{D}_{\text{C}}}$ chimeras, the γ_{D} interface was restored in both chimeras by mutating the γ_{S} interface residues to the structural homologous γ_{D} interface residues. These chimeras are designated as $\gamma_{\text{D}_{\text{N}}}$ - $\gamma_{\text{S}_{\text{C}}}$ [$\gamma_{\text{D}_{\text{INT}}}$] and $\gamma_{\text{S}_{\text{N}}}$ - $\gamma_{\text{D}_{\text{C}}}$ [$\gamma_{\text{D}_{\text{INT}}}$] (Table 4-1). Since the structures of the chimeras are unknown, it is difficult to assess if the interface associations completely mimic the γ_{D} interface. However, the interface residues are aligned to the same positions in $\gamma_{\text{S}_{\text{WT}}}$ supporting that these substitutions may interact in a similar manner. Biophysical comparisons between the mismatched interface chimeras and the $\gamma_{\text{D}_{\text{INT}}}$ chimeras may give insight into the differences between the interfaces of the chimeras.

Table 4-1. Composition of domain interface in wild type and chimera proteins.

		WT Proteins		Chimera Proteins		
Residues	AA# ^a	γ D _{WT}	γ S _{WT}	γ D _N - γ S _C	γ S _N - γ D _C	γ D _N - γ S _C [γ D _{INT}] γ S _N - γ D _C [γ D _{INT}]
Hydrophobic	43 – 145	Met – Leu	Ala – Leu	Met – Leu	Ala – Leu	Met – Leu
	56 – 132	Phe – Val	Ile – Ile	Phe – Ile	Ile – Val	Phe – Val
	81 – 170	Ile – Val	Val – Ile	Ile – Ile	Val – Val	Ile – Val
Peripheral	54 – 143	Gln – Gln	Met – Gln	Gln – Gln	Met – Gln	Gln – Gln
	79 – 147	Arg – Met	Arg – Asp	Arg – Asp	Arg – Met	Arg – Met

^a Residues are numbered based on the crystal structure of γ D_{WT} (PDB# 1HK0) and sequence alignment between γ D_{WT} and γ S_{WT}.

2. Purification of the Chimeras Proteins

All of the proteins were expressed in *E. coli* and accumulated to comparable levels. All chimeras were soluble and could be purified by standard lysis protocols. In addition, all crystallins were able to fold into native-like proteins and were stable through dialysis and storage. Small losses of protein due to precipitation may indicate a slight loss in stability. Each recombinant chimera was purified using Ni-NTA affinity chromatography. There were some slight differences in the purification of the chimeras, with $\gamma_{\text{SN}}\text{-}\gamma_{\text{DC}}$ and $\gamma_{\text{SN}}\text{-}\gamma_{\text{DC}}[\gamma_{\text{D}_{\text{INT}}}]$ requiring higher imidazole concentration to elute from the Ni-NTA column. The biochemical basis for this observation is unknown.

3. Analytical Size Exclusion Chromatography

Analytical SEC profiles were obtained on each chimera to determine if any formed stable dimers. All of the recombinant wild type and chimera proteins eluted after the 25 kDa protein standard (Chymotrypsinogen A) as expected, and near the 13.7 kDa protein standard (Ribonuclease A) (Fig. 4-3). This result was surprising; however, the isolated domain proteins eluted at ~1 ml later than the full length proteins indicating in comparison all full length proteins are eluting at the appropriate volume (Chapter 2). This discrepancy may be due to the overall shape of the crystallins since all of them elute at similar volumes.

All chimeras were loaded onto a column at a concentration of 80 $\mu\text{g/ml}$, 8-fold higher than the experimental conditions. At this concentration, all of the chimeras eluted off of the SuperdexTM200 column at similar volumes as $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$ (Fig. 4-3). The elution peak volumes for $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$ were 17.86 ml and 17.4 ml, respectively. The $\gamma_{\text{D}_{\text{N}}}\text{-}\gamma_{\text{S}_{\text{C}}}$ chimera eluted at 17.66 ml and $\gamma_{\text{D}_{\text{N}}}\text{-}\gamma_{\text{S}_{\text{C}}}[\gamma_{\text{D}_{\text{INT}}}]$ eluted at a similar volume of 17.67 ml. Also, both $\gamma_{\text{S}_{\text{N}}}\text{-}\gamma_{\text{D}_{\text{C}}}$ and $\gamma_{\text{S}_{\text{N}}}\text{-}\gamma_{\text{D}_{\text{C}}}[\gamma_{\text{D}_{\text{INT}}}]$ eluted at similar volumes of 17.5 ml and 17.59 ml respectively. The retention volumes of the chimeras were between $\gamma_{\text{D}_{\text{WT}}}$ and $\gamma_{\text{S}_{\text{WT}}}$ retention volumes as would be expected. Restoration of the $\gamma_{\text{D}_{\text{INT}}}$ did not change the elution volume of the chimeras. Taking into account the isolated domain

elution data along with the wild type data, these results suggests that all of the chimeras were in the monomeric form and did not appear to form stable dimers. However, oligomerization was not tested at higher concentrations of the chimeras which may demonstrate different properties.

4. Circular Dichroism Spectroscopy

Analysis of secondary and tertiary structure of these chimeric proteins was performed to determine if they exhibited significant differences from the two parent molecules, γD_{WT} and γS_{WT} . We utilized Far UV CD spectroscopy to access the secondary structure of the chimera crystallins. All CD spectra had a minimum at 218 nm characteristic of the β -sheet γ -crystallins. The chimeras which contained the N-terminal domain of the parent molecule had CD spectra most similar to that of the parent (Fig. 4-4). For example, the γD_N - γS_C chimeras had a CD profile similar to γD_{WT} , while the γS_N - γD_C chimeras had CD profiles similar to γS_{WT} . By deconvoluting the chimera spectrum, a more sensitive comparison was made between the wild type and the chimera proteins. The deconvolution of the chimeras revealed the similarity between all of the chimeras along with slight differences in the amount of β -sheet and α -helical structure (Table 4-2).

5. Fluorescence Spectroscopy

In order to probe the tertiary conformation of the protein, fluorescence spectroscopy was utilized. Both γD_{WT} and γS_{WT} have four buried tryptophans at different regions of the protein, Trp 68 and Trp 156 (Trp 72 and Trp 162 in γS_{WT}) at the top region of the protein and Trp 42 and 130 (Trps 46 and 136 in γS_{WT}) at the bottom region of the protein. Fluorescence spectroscopy showed that the γD_N - γS_C chimeras had a fluorescence maximum at 327 nm similar to the γD_{WT} in the native state (Fig. 4-5, Table 4-2). γS_N - γD_C chimeras had a similar fluorescence profile to γS_{WT} with the maximum intensity slightly red shifted to 330 nm (Fig. 4-5, Table 4-2). All chimeras unfolded with a maximum intensity of ~ 350 nm in the unfolded state indicative of solvent exposed tryptophans (Fig. 4-5).

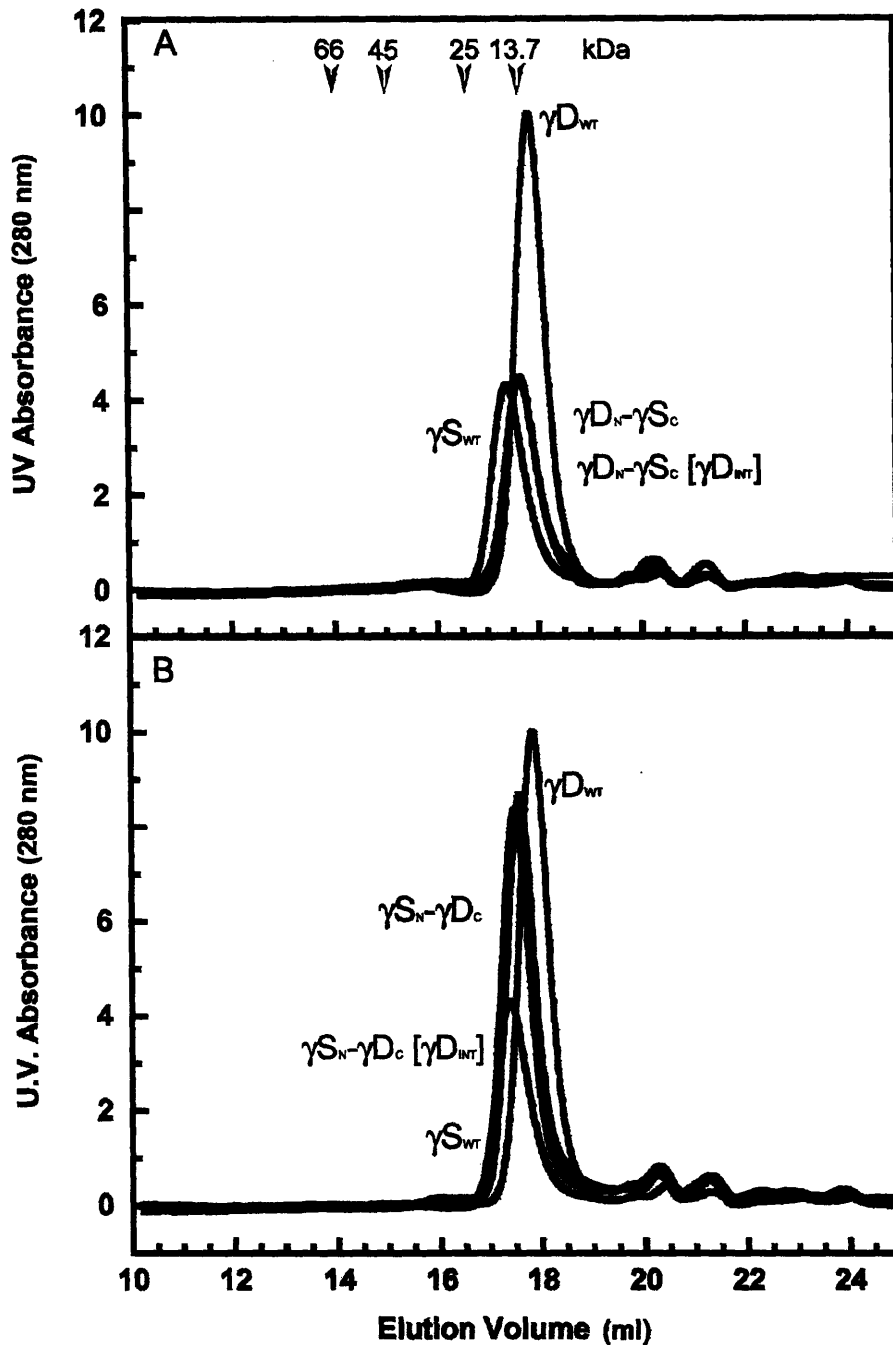


Figure 4-3. Analytical Size Exclusion Chromatography profiles of the chimeras and wild type proteins. All samples were loaded onto a Superdex™200 10/300 GL column at a protein concentration of 80 $\mu\text{g/ml}$. The molecular weight standards were BSA (66 kDa), Ovalbumin (45 kDa), Chymotrypsinogen A (25 kDa), Ribonuclease A (13.7 kDa) and are noted by arrow. (A) γD_{WT} (black), γS_{WT} (blue), $\gamma D_N-\gamma S_C$ (green), and $\gamma D_N-\gamma S_C [\gamma D_{WT}]$ (purple). (B) γD_{WT} (black), γS_{WT} (blue), $\gamma S_N-\gamma D_C$ (gray), and $\gamma S_N-\gamma D_C [\gamma D_{INT}]$ (red).

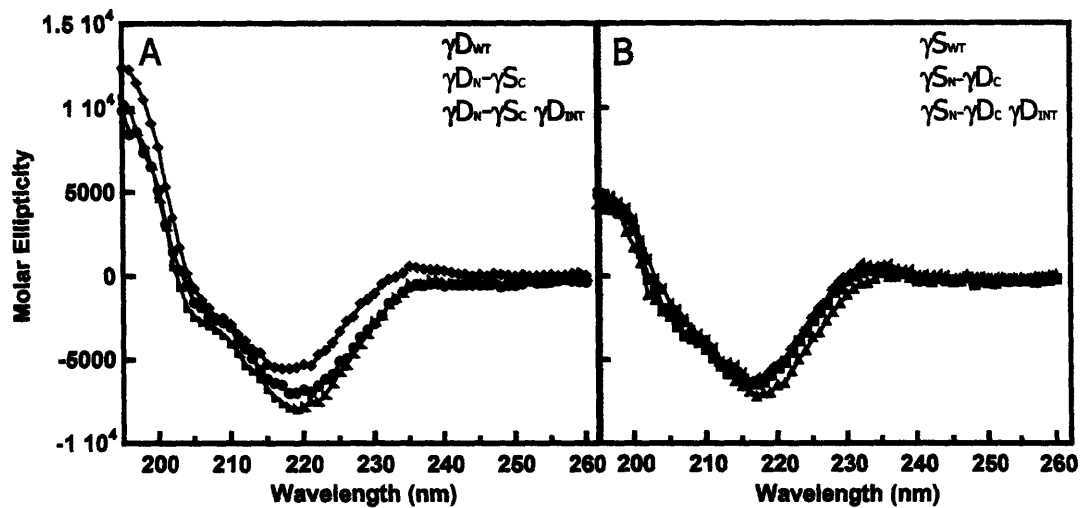


Figure 4-4. Far-UV CD spectroscopy of wild type and chimera proteins. All samples were at a protein concentration of 100 $\mu\text{g/ml}$ in 10 mM sodium phosphate buffer, pH 7.0 at 37°C.

(A) CD spectra recorded from 195-260 nm wavelengths for $\gamma\text{D}_{\text{WT}}$ (black \blacklozenge), $\gamma\text{D}_{\text{N}}\text{-}\gamma\text{S}_{\text{C}}$ (green \blacktriangleright), and $\gamma\text{D}_{\text{N}}\text{-}\gamma\text{S}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] (purple \bullet).

(B) CD spectra recorded from 195-260 nm wavelengths for $\gamma\text{S}_{\text{WT}}$ (blue \blacktriangle), $\gamma\text{S}_{\text{N}}\text{-}\gamma\text{D}_{\text{C}}$ (gray \blacktriangleleft), and $\gamma\text{S}_{\text{N}}\text{-}\gamma\text{D}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] (red \blacksquare).

Table 4-2. Deconvoluted CD spectra and fluorescence emission spectra maximums for γ D and γ S wild type and chimera proteins.

PROTEIN	CD Spectra				Fluorescence Emission	
	% β -Sheet	% α -Helix	% Turns	% Unordered	Native (nm)	Unfolded (nm)
γ D _{WT}	40	6	21	31	326	350
γ D _N - γ S _C	34	12	23	32	327	350
γ D _N - γ S _C [γ D _{INT}]	36	11	22	31	327	350
γ S _{WT}	33	6	24	36	330	350
γ S _N - γ D _C	34	5	24.5	35	330	350
γ S _N - γ D _C [γ D _{INT}]	34	6	25	35	330	350

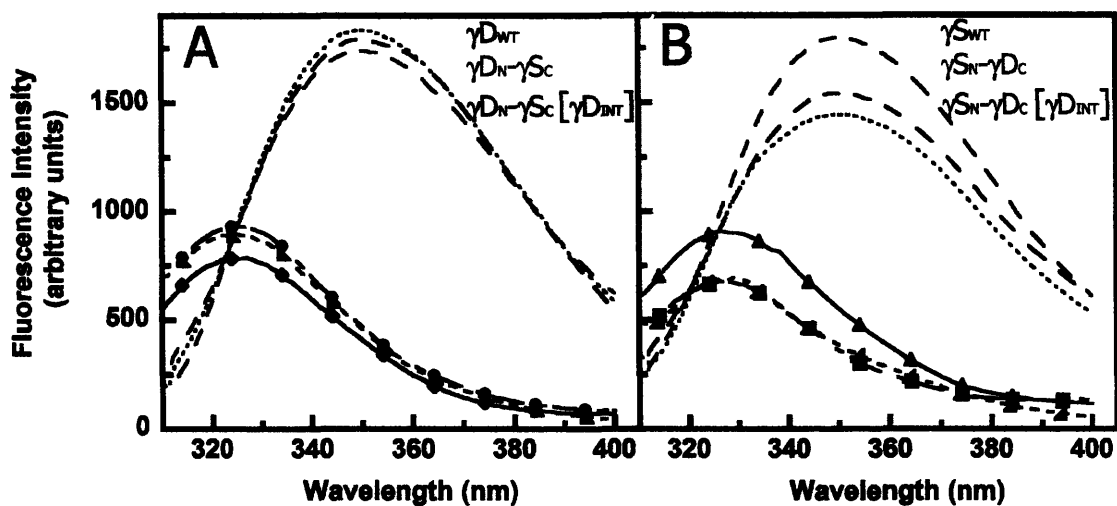


Figure 4-5. Native and unfolded fluorescence emission spectra of wild type and chimera proteins. All proteins excited at 295nm and emission recorded from 310-400 nm wavelengths. Samples consisted of 10 $\mu\text{g/ml}$ protein in 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and additionally 5.5 M GuHCl for unfolded samples equilibrated at 37°C.

(A) $\gamma\text{D}_{\text{WT}}$ Native (black \blacklozenge) and Unfolded spectra (black dotted line); $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ Native (green \blacktriangleright) and $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ Unfolded spectra (green dotted line); $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] Native (purple \bullet) and $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] Unfolded spectra (purple dotted line).

(B) $\gamma\text{S}_{\text{WT}}$ Native (blue \blacktriangle), Unfolded spectra (blue dotted line); $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ Native (gray \blacktriangle) and $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ Unfolded (gray dotted line); $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] Native (red \blacksquare), and $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] Unfolded (red dotted line).

As previously reported, γD_{WT} and γS_{WT} exhibited an increase in fluorescence intensity in the unfolded state compared to the native state of the protein. All chimeras maintained this quenching phenomenon further demonstrating the similarity in the native structure of the chimeras compared to the wild type proteins. Upper Trps 68 (in γS 72) and 156 (in γS 162) are quenched through charge transfer to the protein backbone (Chen et al. 2006). In γD_{WT} , the bottom two Trps 42 and 130 are involved in energy charge transfer to their intradomain upper Trps (Chen et al. 2006). In γS_{WT} , the γS N-terminal domain does not exhibit energy transfer from the bottom Trp 46 to the top Trp 72 while the γS_C does exhibit energy transfer (J. Chen, personal communication).

6. Thermal Denaturation of the Chimeras Compared to Wild Type Proteins

The stabilities of the chimeras were assessed by thermal denaturation, a qualitative measure of stability in the context of a more physiological relevant stress. The stabilities of the chimeras were initially monitored by thermal unfolding in low ionic strength conditions. Far-UV CD spectroscopy was utilized to monitor the transitions by measuring the loss of β -sheet structure at 218 nm with increase in temperature. ΔG and m values could not be extrapolated from the reaction data because of the irreversibility caused by thermally induced aggregation. Instead, the fraction of native protein was calculated since signal interference from the aggregate would possibly affect the amount of unfolded protein calculated. Temperature midpoints (T_M) of the reaction were determined by fitting the thermal curve to a two state model indicating no populated intermediates (Fig. 4-6).

The thermal stabilities of the γD_N - γS_C chimeras, containing the γD N-terminal domain, were similar to the γD_{WT} protein. The T_M of γD_N - γS_C chimera was 81.5°C. The replaced γD interface chimera, γD_N - γS_C [γD_{INT}], had a T_M of 79.5°C. Comparatively, γD_{WT} had a temperature midpoint of 83.8°C (Table 4-3). The restoration of the γD interface in γD_N - γS_C did not increase the thermal stability of the chimera. On the contrary, the γD_N - γS_C [γD_{INT}] exhibited slight destabilization compared to the chimera

that consists of a mismatched interface, γD_N - γS_C . The chimera T_M s were significantly higher compared to the T_M of isolated γD N-terminal domain (64.5°C) and γS C-terminal domain (75.1°C). This result demonstrates the importance of the domain interface, whether it was the mismatched interface or the restored γD interface, in the overall stability of the γD_N - γS_C chimera.

Neither γD_N - γS_C chimera had similar temperature midpoints comparable to γD_{WT} . It was hypothesized that mutating the interface back to the γD interface in the γD_N - γS_C [γD_{INT}] would increase the T_M of the γD_N - γS_C chimera to match the γD_{WT} values. However, the restoration of the γD interface did not improve the thermal stability of the protein. The T_M of the γD_N - γS_C chimeras was ~ 2 - 4°C less than the temperature midpoint of γD_{WT} . This result confirms that the intrinsic stability of the γD C-terminal domain is greater than the γS C-terminal domain in context of a full length protein.

The analyses of the γD and γS isolated domains presented in Chapter two revealed that the C-terminal domain of γD (γD_C) had higher temperature midpoints than all other isolated domains. Additionally, both γD and γS N-terminal domains in isolation exhibited lower T_M s than their C-terminal domain counterparts. Therefore, it was predicted that the γS_N - γD_C chimera, containing the γD C-terminal domain, would likely have a higher midpoint compared to the other chimera. In contrast to our predictions, both γS_N - γD_C chimeras had lower T_M s similar to the γS_{WT} protein (Fig. 4-6, Table 4-3). The γS_N - γD_C chimera had a temperature midpoint of 73.1°C while the chimera with the γD interface, γS_N - γD_C [γD_{INT}], had a T_M of 73.5°C. The γS_{WT} T_M was 74.1°C, within the standard deviations of both chimera temperature midpoints.

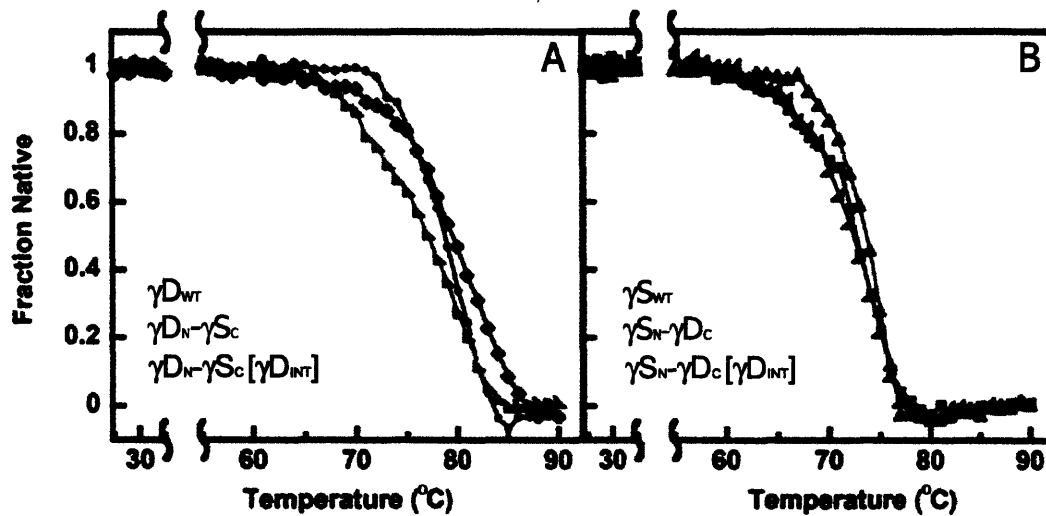


Figure 4-6. Thermal Denaturation of wild type proteins compared to chimera proteins. Samples were prepared at 100 $\mu\text{g/ml}$ protein concentration in 10 mM sodium phosphate buffer, pH 7.0 and CD wavelength 218nm was monitored as the temperature increased from 25-90°C. Data normalized and calculated as Fraction Native.
 (A) γD_{WT} (black \blacklozenge), $\gamma D_N-\gamma S_C$ (green \blacktriangle), and $\gamma D_N-\gamma S_C [\gamma D_{INT}]$ (purple \bullet).
 (B) γS_{WT} (blue \blacktriangle), $\gamma S_N-\gamma D_C$ (gray \blacktriangle), and $\gamma S_N-\gamma D_C [\gamma D_{INT}]$ (red \blacksquare).

Interestingly, in the case of the γS_N - γD_C chimeras, both were less thermally stable compared to γD C-terminal domain in isolation, with temperature midpoints $\sim 3^\circ\text{C}$ less than the isolated γD C-terminus (Table 4-3). The mismatched or γD interface did not affect the overall stability of the protein and the temperature midpoint values of the proteins were similar to the γS C-terminal domain in isolation. The reason for the destabilization from the addition of the γS N-terminus linked to the γD C-terminus is unknown and requires further investigation.

7. Equilibrium Unfolding and Refolding of the Chimeras

As mentioned above, all of the chimeric crystallins were able to refold to native-like states *in vitro*, upon dilution out of denaturant. This was also observed in the overlay of the transitions observed in the chimera equilibrium unfolding/refolding curves at lower GuHCl concentrations (Fig. 4-7C, D). However, it was possible that the mismatched interface of the chimeras affect the chimera's stability. Since the thermal denaturation experiments could not give quantitative stability values, equilibrium unfolding/refolding experiments using GuHCl denaturant were utilized to evaluate the stabilities of the chimeras. The ratio of tryptophan fluorescence emission at 360 and 320 nm was plotted to clearly represent the transitions of the equilibrium unfolding/refolding curves. The concentration midpoint (C_M) of each transition, the concentration of GuHCl where 50% of the protein is unfolded, was determined by fitting the curves to two or three state models. C_M was used as a reproducible measurement of stability, although extrapolated ΔG and m -values were calculated using single wavelength tryptophan fluorescence emission values at 360 nm.

Previously, equilibrium unfolding/refolding data of γD_WT was fit best to a three state model (Flaugh et al. 2005b, Fig. 4-7A). In contrast to γD_WT , the chimera with the N-terminal domain of γD , γD_N - γS_C , had a C_M of 2.3 M GuHCl, and was best fit to a two state transition. Calculated m and ΔG values for this chimera were $3.15 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ and $6.9 \text{ kcal}\cdot\text{mol}^{-1}$ respectively. Restoration of the γD crystallin interface in the γD_N - γS_C [$\gamma\text{D}_\text{INT}$] chimera did not improve the stability of the γD_N - γS_C chimera. This was

consistent with the thermal denaturation experiments (Fig. 4-6). The $\gamma_{D_N}\text{-}\gamma_{S_C}$ [$\gamma_{D_{INT}}$] chimera had a similar transition midpoint at 2.3 M GuHCl and a calculated m and ΔG value of $2.4 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ and $5.4 \text{ kcal}\cdot\text{mol}^{-1}$ respectively (Table 4-3). All of the $\gamma_{D_N}\text{-}\gamma_{S_C}$ [$\gamma_{D_{INT}}$] values were within standard deviation of the calculated values of $\gamma_{D_N}\text{-}\gamma_{S_C}$ (Table 4-3). Comparison between the $\gamma_{D_N}\text{-}\gamma_{S_C}$ chimera and the isolated γ_D N-terminal domain, indicated that the γ_{S_C} interface interactions were adequate to stabilize the N-terminus comparable to the $\gamma_{D_{WT}}$ first transition. This first transition monitored the unfolding of the γ_D N-terminal domain. The isolated C-terminal domain of γ_S had a C_M of 2.3 M GuHCl similar to the C_M of 2.2 M GuHCl of the γ_D first transition. The midpoints of the different transitions monitoring γ_D N-terminus and γ_S C-terminus in context of the chimera overlap depicting a more cooperative transition compared to $\gamma_{D_{WT}}$.

In contrast, the equilibrium transition of the $\gamma_{S_N}\text{-}\gamma_{D_C}$ chimera was clearly best fit to a three state transition, with concentration midpoints of 1.8 M GuHCl and 2.9 M GuHCl (Table 4-3, Fig. 4-7D). These concentration midpoints are similar to the isolated N-terminal domain of γ_S (1.7 M GuHCl) and the isolated C-terminal domain of γ_D (2.8 M GuHCl) (Table 4-3). Therefore, the $\gamma_{S_N}\text{-}\gamma_{D_C}$ chimera likely monitored the intrinsic stability of both individual domains and not contributions from the interface. The $\gamma_{S_N}\text{-}\gamma_{D_C}$ chimera m and ΔG values were $3.4 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ and $6.0 \text{ kcal}\cdot\text{mol}^{-1}$ for the first transition; and $2.8 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ and $8.24 \text{ kcal}\cdot\text{mol}^{-1}$ for the second transition (Table 4-3).

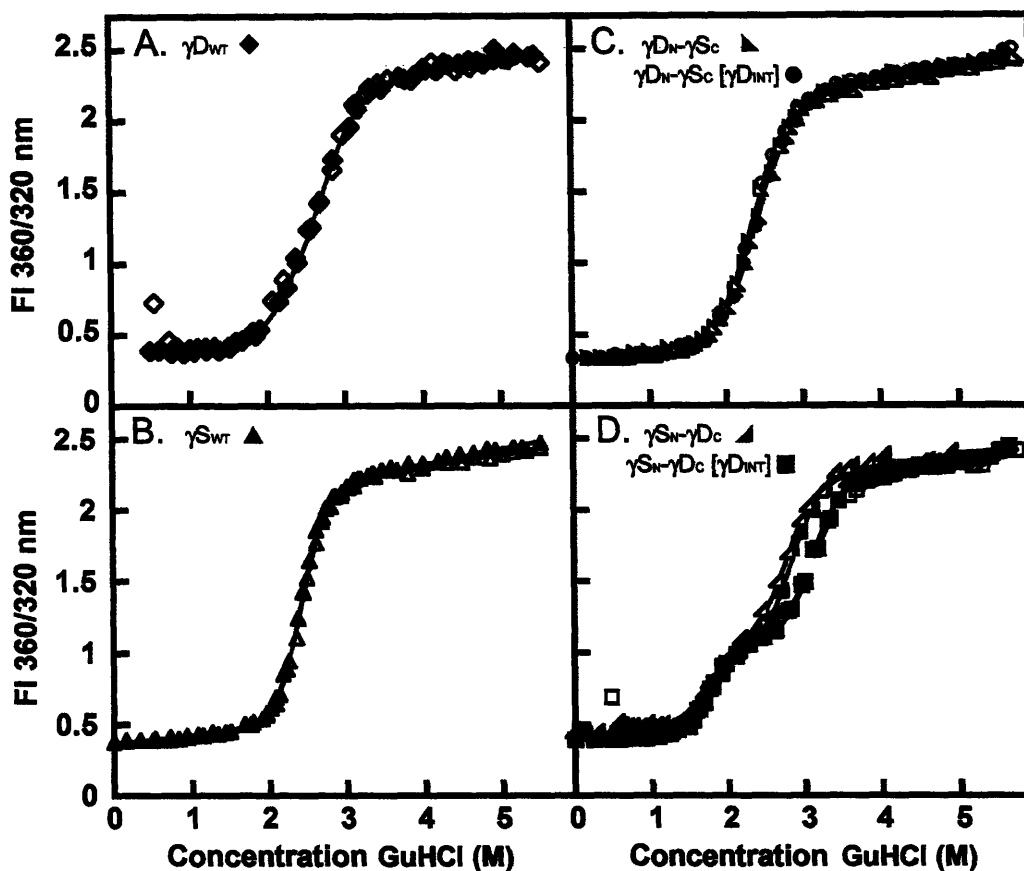


Figure 4-7. Equilibrium unfolding (closed symbols) and refolding (open symbols) for A. γ_{DWT} (\blacklozenge), B. γ_{SWT} (\blacktriangle), and C. $\gamma_{DN-\gamma_{SC}}$ (\blacktriangle) and $\gamma_{DN-\gamma_{SC}}$ [γ_{DN-INT}] (\bullet), D. $\gamma_{SN-\gamma_{DC}}$ (\blacktriangle) and $\gamma_{SN-\gamma_{DC}}$ [γ_{SN-INT}] (\blacksquare).

Samples consisted of 10 $\mu\text{g/ml}$ protein concentration, 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and various concentrations of GuHCl at 37°C. Fluorescence emission at 360 nm and 320 nm were calculated as a ratio. Equilibrium data fit indicated by solid black line.

Table 4-3. Equilibrium unfolding/refolding at 37°C and thermal unfolding parameters of γ D and γ S wild type and chimera proteins.

Protein	Thermal Unfolding Transition				Equilibrium Transition 1			Equilibrium Transition 2		
	T_M^a	ΔT_M^b	C_M^c	Apparent m value ^d	Apparent ΔG_{N-U}^e	C_M^c	Apparent m value ^d	Apparent m value ^d	Apparent ΔG_{I-U}^e	
γD_{WT}^f	83.8 ± 1.3	-	2.2 ± 0.1	3.6 ± 0.1	7.7 ± 0.2	2.8 ± 0.1	3.1 ± 0.4	3.1 ± 0.4	8.9 ± 1.3	
$\gamma D_N\text{-}\gamma S_C$	81.5 ± 1.1	2.3	2.2 ± 0.09	3.15 ± 0.6	6.9 ± 1.1	-	-	-	-	
$\gamma D_N\text{-}\gamma S_C$ [γD_{INT}]	79.5 ± 0.3	4.3	2.3 ± 0.08	2.4 ± 0.4	5.4 ± 0.8	-	-	-	-	
γD_N^g	64.5 ± 0.6	19.3	1.2 ± 0.1	3.3 ± 0.9	3.7 ± 0.7	-	-	-	-	
γS_C^g	75.1 ± 0.75	8.7	2.3 ± 0.07	3.5 ± 0.4	8.2 ± 0.6	-	-	-	-	
γS_{WT}^g	74.1 ± 0.2	-	2.3 ± 0.02	4.6 ± 0.4	10.5 ± 0.9	-	-	-	-	
$\gamma S_N\text{-}\gamma D_C$	73.1 ± 0.2	1	1.8 ± 0.1	3.4 ± 0.6	6.01 ± 1.2	2.9 ± 0.07	2.8 ± 0.3	2.8 ± 0.3	8.24 ± 0.9	
$\gamma S_N\text{-}\gamma D_C$ [γD_{INT}]	73.5 ± 0.03	0.6	1.7 ± 0.02	3.5 ± 0.3	6.1 ± 0.4	3.0 ± 0.05	3.0 ± 0.25	3.0 ± 0.25	9.1 ± 0.8	
γS_N^g	69.1 ± 2.4	5	1.7 ± 0.08	2.9 ± 0.4	4.9 ± 0.8	-	-	-	-	
γD_C^g	76.2 ± 0.2	-2.1	2.7 ± 0.09	3.2 ± 0.1	8.7 ± 0.5	-	-	-	-	

^a Midpoints of melting transitions monitored by 218 nm FarUV CD in units of °C.

^b $\Delta T_M = T_{M_{WT}} - T_{M_{Isolated\ Domain/Chimera}}$ in units of °C.

^c Transition midpoints in units of M GuHCl.

^d Apparent m values in units of kcal* mol^{-1} * M^{-1} .

^e Free energy of unfolding in the absence of GuHCl in units of kcal* mol^{-1} .

^f Thermal denaturation parameters are from Flaugh et al. 2006, Equilibrium parameters of γD_{WT} are from Flaugh et al. 2005.

^g Thermal denaturation parameters are from Chapter 2, Equilibrium parameters of γS_{WT} are from Chapter 2.

Restoration of the γ D interface in γ S_N- γ D_C obtained similar C_{MS} to the C_{MS} observed for mismatched chimera interface (Table 4-3). There was a slight increase in the stability of the γ D C-terminal domain, but this affect was minimal with ~ 0.1 M increase in the second transition C_M . The lack of increased stability in the γ S_N- γ D_C [γ D_{INT}] chimera suggests that the interface contacts do not stabilize this chimera. These results differed from the thermal denaturation data which showed a decrease in γ D C-terminus stability in context of the γ S_N- γ D_C chimeras. On the other hand, the equilibrium results do not reveal a stability decrease in γ D C-terminus in the context of the chimeras.

As mentioned above, upon refolding of γ D_{WT} from 5.5 M GuHCl into buffer, the refolding chains enter an off-pathway aggregation reaction that competes with productive refolding of the protein (Kosinski-Collins and King 2003). This off-pathway aggregation reaction results in an increase in light scattering as observed by fluorescence spectroscopy (Flaugh et al. 2005a; Flaugh et al. 2005b; Kosinski-Collins and King 2003). The apparent increase in the fluorescence emission at 320 and 360 nm is recorded in the equilibrium refolding multiphasic curve (Fig 4-7A, γ D_{WT} (open \diamond symbol)). Only the γ S_N- γ D_C [γ D_{INT}] chimera exhibited the increased light scattering similar to γ D_{WT} (Fig. 4-7D).

8. *Equilibrium Refolding Aggregation Properties of the Chimera Proteins*

As mentioned above, previous equilibrium refolding studies of γ D_{WT} revealed an off-pathway aggregation reaction creating ordered fibrils that were not amyloid in nature (Kosinski-Collins and King 2003). These aggregates formed upon dilution out of high concentration of chaotropic denaturant even at a low protein concentration of 10 μ g/ml and at 37°C. Under these conditions, $\sim 50\%$ of the protein partitioned into the off-pathway polymerization reaction and $\sim 50\%$ refolded into native-like protein. The highly ordered character of these aggregates suggested that other γ -crystallins, which have homologous structure, may exhibit a similar off-pathway aggregation reaction. Consequently, γ S_{WT} was tested for its ability to aggregate under the same conditions.

In order to test for the presence of higher molecular weight species, solution turbidity measurements at 350 nm were utilized to determine the ability of the sample to scatter light. In addition, fluorescence emission scans of the refolded sample were taken to assess the conformation of the refolded proteins. However, when probing γD_{WT} refolded samples, they primarily scattered light yielding a red shifted emission maximum at ~335 nm and increased fluorescence intensity spectra compared to the native-like refolded spectra (Kosinski-Collins and King 2003, Fig. 4-8). The fluorescence emission maximums of the native-like refolded protein were comparable to the native fluorescence emission spectrum (compare Fig. 4-5 and Fig. 4-8). This difference in the fluorescence emission of the native-like refolded protein and light scattering aggregated state was used as a qualitative measurement for the presence of aggregates. Accordingly, fluorescence emission scans were taken before and after a high speed centrifugation in order to estimate the amount of partitioning of the protein into the off-pathway aggregation reaction.

In contrast to γD_{WT} results, upon refolding out of denaturant there was not an increase in scattering indicating the absence of high molecular weight species in γS_{WT} (Fig. 4-7 and Fig. 4-8). The γS_{WT} refolded sample was spun at high speed and a fluorescence spectrum was taken for comparative purposes. Both spectra were similar in the native-like refolded protein maximum and fluorescence emission intensities demonstrating no loss of protein during centrifugation. These results agreed with solution turbidity data that γS_{WT} does not aggregate under these conditions regardless of the structural similarities with γD_{WT} (Fig. 4-9).

To determine what domain of γD_{WT} or γS_{WT} may be responsible for the aggregation properties, we assessed if the chimeras would follow the off-pathway aggregation reaction as seen in γD_{WT} . Both $\gamma D_N\text{-}\gamma S_C$ and $\gamma S_N\text{-}\gamma D_C$ chimeras did not demonstrate an appreciable increase in solution turbidity as seen in γD_{WT} (Fig. 4-9). The $\gamma D_N\text{-}\gamma S_C$ and $\gamma S_N\text{-}\gamma D_C$ chimeras yielded spectra similar to their native spectrum (Fig. 4-8).

However, restoration of the γ D interface in the both chimeras, γ S_N- γ D_C [γ D_{INT}] and γ D_N- γ S_C [γ D_{INT}] demonstrated an increase in solution turbidity (Fig. 4-9). Moreover, γ S_N- γ D_C [γ D_{INT}] fluorescence emission spectra recapitulated the distinct spectra observed in the γ D_{WT} refolded samples (Fig. 4-8). The amount of protein partitioned into the off-pathway aggregate of γ S_N- γ D_C [γ D_{INT}] chimera was ~50% similar to γ D_{WT}. In γ D_N- γ S_C [γ D_{INT}], the fluorescence emission scan revealed a slight shift in the refolded sample and the solution turbidity was slightly increased (Fig. 4-8 and 4-9). The red emission shift and fluorescence intensity was not as dramatic as γ D_{WT} and the fluorescence intensity did not decrease in significant amounts after the centrifugation of the aggregate.

These results suggest that the aggregation process requires both domains, consistent with domain swap models.

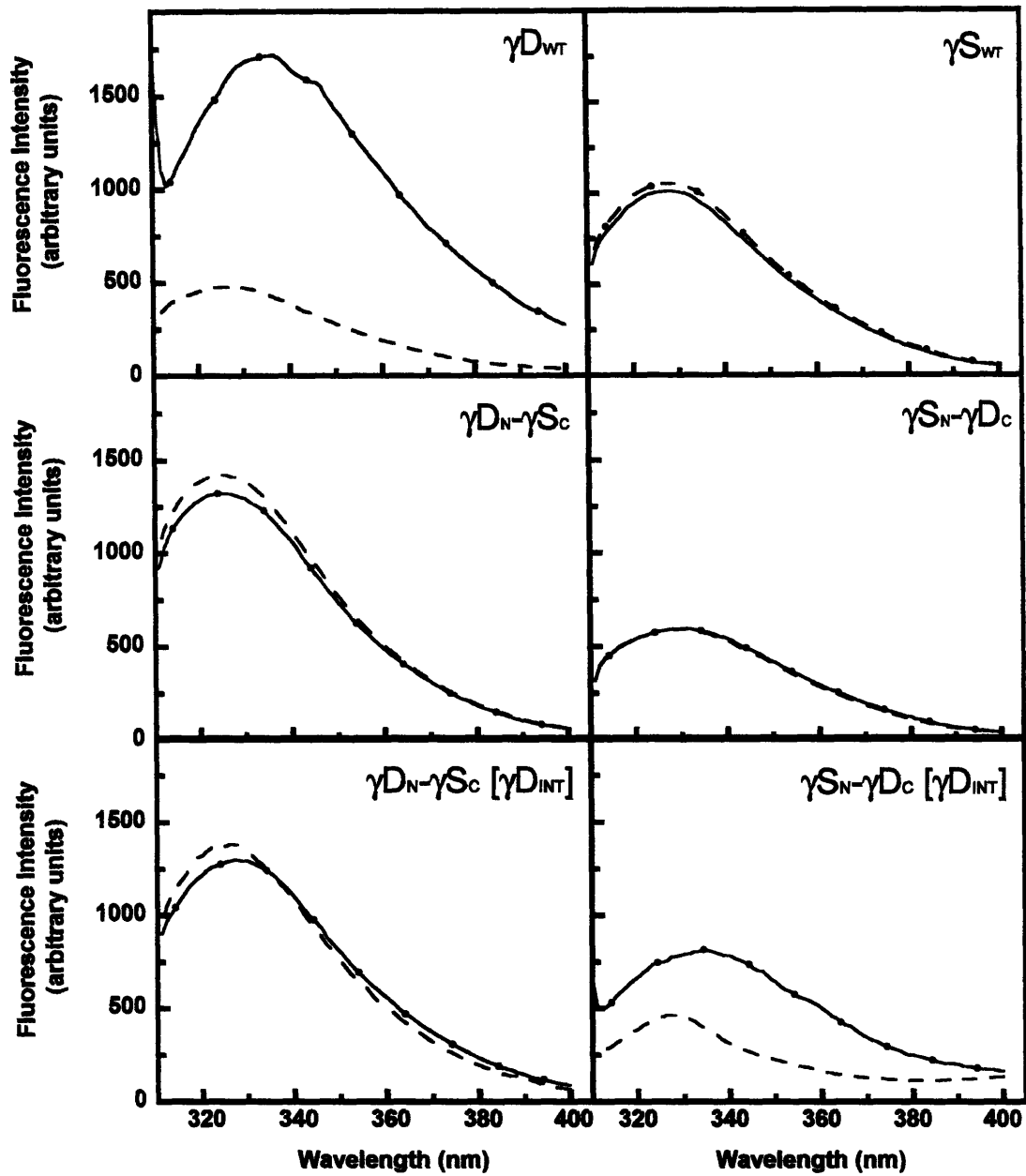


Figure 4-8. Equilibrium Refolding of wild type and chimera proteins before and after centrifugation. All samples were unfolded in 5.5 M GuHCl and refolded by dilution into 100 mM Sodium Phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0. Refold samples before centrifugation (—●—), refold sample after centrifugation (dotted lines).

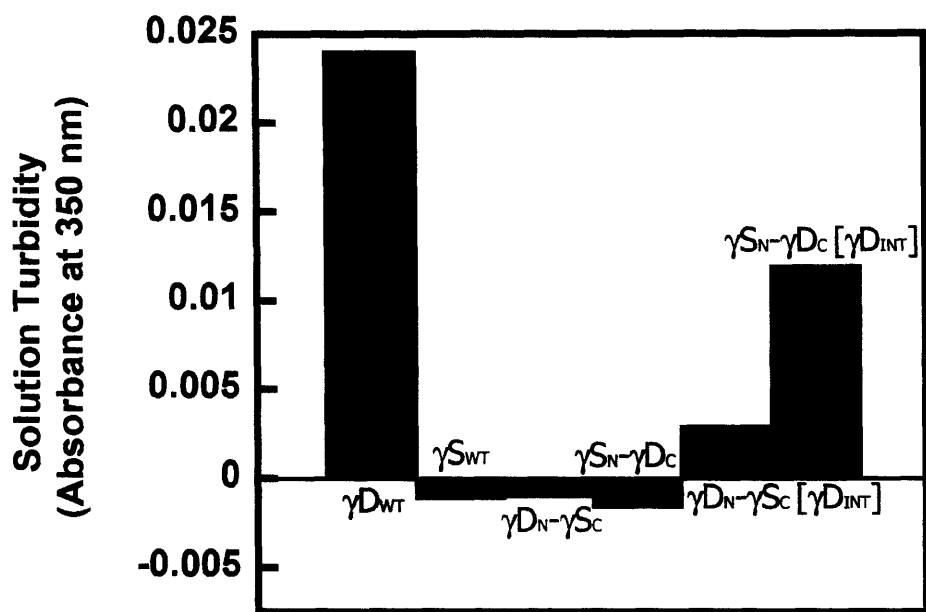


Figure 4-9. Solution turbidity of the refolding sample of wild type and chimera proteins before centrifugation. Protein concentration was 10 $\mu\text{g/ml}$. Solution turbidity measurements absorbance at 350 nm.

9. Aggregation Kinetics Demonstrate Differential Aggregation Properties

Previous experiments demonstrated that the γD_{WT} off-pathway aggregation formed within minutes and did not appear to increase with time after 40 hours incubation (Kosinski-Collins and King 2003). In order to determine if the rate of formation was altered in the γD interface chimeras and to confirm the equilibrium refolding aggregate results, we followed the solution turbidity over time for all of the chimeras and wild type proteins. We also tested the aggregation kinetics of the isolated domains of γD_{WT} , γD_N and γD_C . γS_N and γS_C were not tested since it has been established that the γS_{WT} does not exhibit aggregation under these conditions.

As observed in Fig. 4-10, the aggregation of H γD -Crys polymerized within minutes of diluting the protein out of denaturant. It was difficult to monitor the initial steps of the aggregation which was formed within the dead time of the experiments (~15 s). The reaction is best fit to a two state transition model indicating the aggregation reaction associates rapidly. The calculated $t_{1/2}$ of the absorbance curve was ~73 seconds (1.2 minutes) indicative of a short or absent lag time. The isolated γD_N and γD_C did not exhibit any increase in absorbance at 350 nm confirming that both domains are necessary for the formation of the γD_{WT} aggregate.

The γD_N - γS_C and γS_N - γD_C chimeras did not accumulate appreciable amounts of aggregated protein. γD_N - γS_C chimera absorbance at 350 nm revealed a ~ 5 fold increase above γS_{WT} absorbance but a 4 fold decrease compared to γD_{WT} . This result was reproducible and was consistently higher than the native control, suggesting that this chimera either forms smaller aggregates than γD_{WT} or forms a small amount of aggregates. On the other hand, γS_N - γD_C had a slightly higher absorbance than γS_{WT} , which likely is not significant and is within the error of the experiment.

In contrast, the γD_{WT} interface chimeras reconstituted the off-pathway aggregation observed in γD_{WT} . Two-state kinetics were fit to both γD_N - γS_C and γS_N - γD_C chimeras. It

was difficult to fit $\gamma D_N\text{-}\gamma S_C$ because of the decrease in absorbance provided a slanted baseline for the fit. However, by careful estimation, the rate of aggregation appeared similar to γD_{WT} which had a $t_{1/2}$ of ~ 73 seconds. The nature of this decrease in absorbance over time is still under investigation, but perhaps there is an equilibrium between the aggregate and the native refolded protein. In contrast to $\gamma D_N\text{-}\gamma S_C$ [D_{INT}], the opposite chimera, $\gamma S_N\text{-}\gamma D_C$ [γD_{INT}], led to the formation of the off-pathway aggregation reaction that persisted for the length of the experiment. This chimera was also fit to a two state model and it had a calculated $t_{1/2}$ of 105 seconds, slightly longer than the $t_{1/2}$ of the γD_{WT} protein.

The aggregation kinetic results suggested that the interface was the key region of the protein that was important for the aggregation reaction. Both domains are required for association in addition to the necessity of restoring the γD interface in order to promote aggregation. These results support a domain swapped model of the aggregate in which associations between the domains are maintained like the monomeric protein, but the associations are intermolecular and not intramolecular.

10. Aggregation Properties at Higher Protein Concentrations

The aggregation kinetics demonstrated that neither the N-terminal nor the C-terminal isolated domains exhibited competitive off-pathway aggregation at 10 $\mu\text{g/ml}$ concentration. Also the γS_{WT} isolated domains did not exhibit any aggregation reaction under these conditions (Chapter 2). In addition, refolding γD_{WT} isolated domains together under the same conditions did not reconstitute the γD_{WT} aggregation reaction (data not shown). These results confirmed that a two domain protein was necessary for the off-pathway aggregation reaction. Although, off-pathway aggregation was not observed at 10 $\mu\text{g/ml}$ protein concentration, we wanted to determine if increasing the protein concentration would affect the isolated domains and the chimera's ability to aggregate.

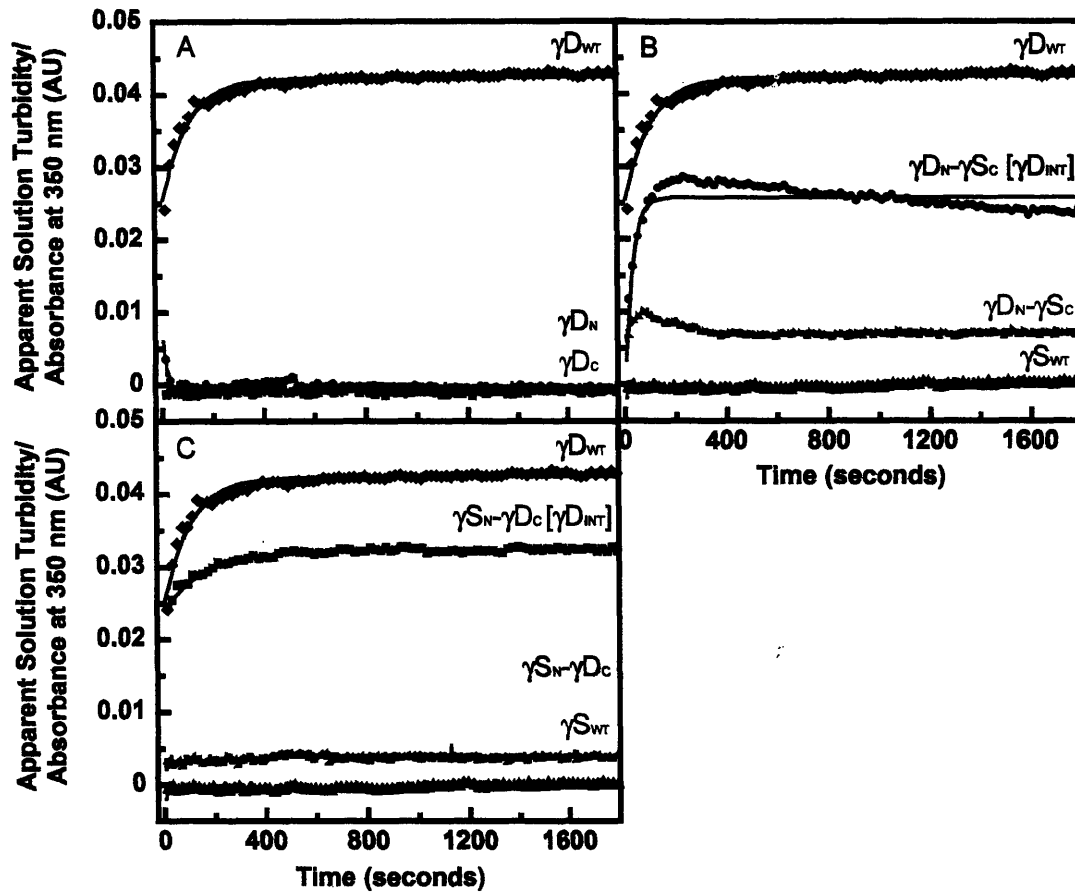


Figure 4-10. Aggregation kinetics of the wild type and chimera proteins. All proteins were rapidly refolded into 100 mM Sodium Phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 and monitored by absorbance at 350 nm over 30 minutes. All samples have a protein concentration of 10 $\mu\text{g/ml}$. $\gamma\text{D}_{\text{WT}}$, $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$], and $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] curves which were fit to two state kinetic models; the matching lines represents the fit.

(A) $\gamma\text{D}_{\text{WT}}$ (black \blacklozenge), $\gamma\text{D}_{\text{N}}$ (dark blue \bullet), and $\gamma\text{D}_{\text{C}}$ (orange \blacksquare).

(B) $\gamma\text{D}_{\text{WT}}$ (black \blacklozenge), $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ (green \blacktriangleright), $\gamma\text{D}_{\text{N}}-\gamma\text{S}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] (purple \bullet), and $\gamma\text{S}_{\text{WT}}$ (light blue \blacktriangle).

(C) $\gamma\text{D}_{\text{WT}}$ (black \blacklozenge), $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ (gray \blacktriangleleft), $\gamma\text{S}_{\text{N}}-\gamma\text{D}_{\text{C}}$ [$\gamma\text{D}_{\text{INT}}$] (red \blacksquare), $\gamma\text{S}_{\text{WT}}$ (light blue \blacktriangle).

The presence of the equilibrium refolded aggregate was assessed again by solution turbidity measurements. The ability of each protein to aggregate at higher concentrations was tested using five different protein concentrations, 10, 25, 50, 75, and 100 $\mu\text{g/ml}$. As the protein concentration increased, the $\gamma\text{D}_{\text{WT}}$ solution turbidity increased in a linear pattern (Fig. 4-11; lines are used as guides). During rapid refolding of the proteins at higher concentrations, $\gamma\text{S}_{\text{WT}}$ and isolated $\gamma\text{D}_{\text{C}}$ exhibited off-pathway aggregation competing with productive refolding (Fig. 4-11). Aggregation in $\gamma\text{S}_{\text{WT}}$ was observable spectroscopically at 75 $\mu\text{g/ml}$; although the turbidity was 5 fold less than $\gamma\text{D}_{\text{WT}}$ at this concentration. All isolated domains except for $\gamma\text{D}_{\text{C}}$ failed to show evidence of aggregation for concentrations up to 100 $\mu\text{g/ml}$. For $\gamma\text{D}_{\text{C}}$, concentrations at 50, 75, and 100 $\mu\text{g/ml}$ revealed high molecular weight species. However, the observed $\gamma\text{D}_{\text{C}}$ aggregation was ~ 3.5 fold lower than $\gamma\text{D}_{\text{WT}}$ at the respective concentrations. The $\gamma\text{D}_{\text{N}}$ - $\gamma\text{S}_{\text{C}}$, and $\gamma\text{S}_{\text{N}}$ - $\gamma\text{D}_{\text{C}}$ chimeras aggregated at higher concentrations also. The solution turbidity measurements for these chimeras were in the middle of the measurements obtained for $\gamma\text{D}_{\text{WT}}$ and $\gamma\text{S}_{\text{WT}}$ as would be expected for these chimera proteins. Further structural analysis of these higher concentration aggregates will need to be performed in order to determine if the $\gamma\text{S}_{\text{WT}}$, $\gamma\text{D}_{\text{C}}$, $\gamma\text{D}_{\text{N}}$ - $\gamma\text{S}_{\text{C}}$, and $\gamma\text{S}_{\text{N}}$ - $\gamma\text{D}_{\text{C}}$ aggregates have similar properties as the $\gamma\text{D}_{\text{WT}}$ refolding aggregate or have different morphologies.

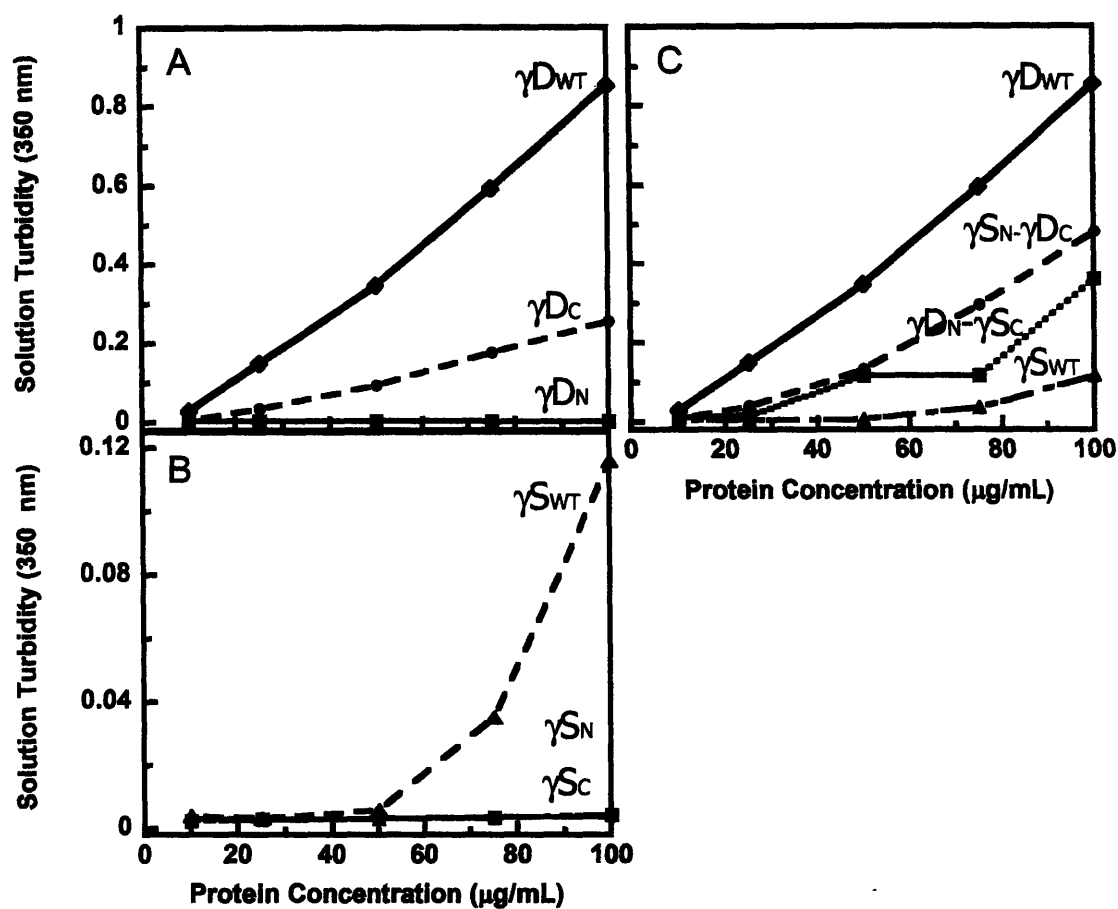


Figure 4-11. Equilibrium Refold Aggregation of wild type, isolated domains, and chimeras at different protein concentrations. Proteins were unfolded in 5.5 M GuHCl and diluted into 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer. Proteins were diluted to 10, 25, 50, 75, and 100 µg/ml final protein concentration. All samples were equilibrated for 24 hours. Lines through the data points are present for visual clarity.

(A) γ_{DWT} (black ◆), γ_{DN} (blue ■), and γ_{DC} (red ●).

(B) γ_{SWT} (green ▲), γ_{SN} (red ●), and γ_{SC} (blue ■); (Note the y axis is a different scale for panels A and B).

(C) γ_{DWT} (black ◆), γ_{SN-γDC} (red ●), γ_{DN-γSC} (blue ■), γ_{SWT} (green ▲).

11. IR spectroscopy of the Chimeras

In order to assess the conformation of the different chimera aggregates, we utilized ATR-IR spectroscopy. IR analysis of the amide I and amide II bands can give information about the overall secondary structure of the protein. Recent studies have utilized IR spectroscopy to characterize amide I (and II) bands of amyloid structures due to unique cross β -sheet conformation (Petty et al. 2005). We refolded γD_{WT} and chimera proteins and placed each individually on an ATR cell used for solid samples, to determine the refolding aggregate conformation. Comparisons were made between the IR profile of the γD_{WT} aggregate and the aggregates of γD_N - γS_C and γS_N - γD_C chimeras. All protein samples had similar profiles with a wavenumber maximum at 1630 cm^{-1} for the amide I band characteristic of a globular β -sheet protein (Fig. 4-12). A similar amide I band maximum is seen in native γD protein in solution (data not shown). These results suggest that the γD_{WT} and chimera aggregates retain a similar conformation to the native γD protein. Also deconvolution of the IR interferograms demonstrated a band at 1680 cm^{-1} indicative of β -turn structures. These results confirmed that the γD_{WT} aggregate characterized by AFM is structured and not amorphous. The IR interferograms of the chimera aggregates suggest that they may have similar characteristics to the γD_{WT} aggregate.

There were slight differences in the amide II band maximum which shifted from 1580 cm^{-1} to 1570 cm^{-1} in the chimeras. This may be attributed to the different side chain contributions in the protein. Other characteristics of the different crystallin aggregates may give insight into the conformation of the aggregate formation *in vitro*.

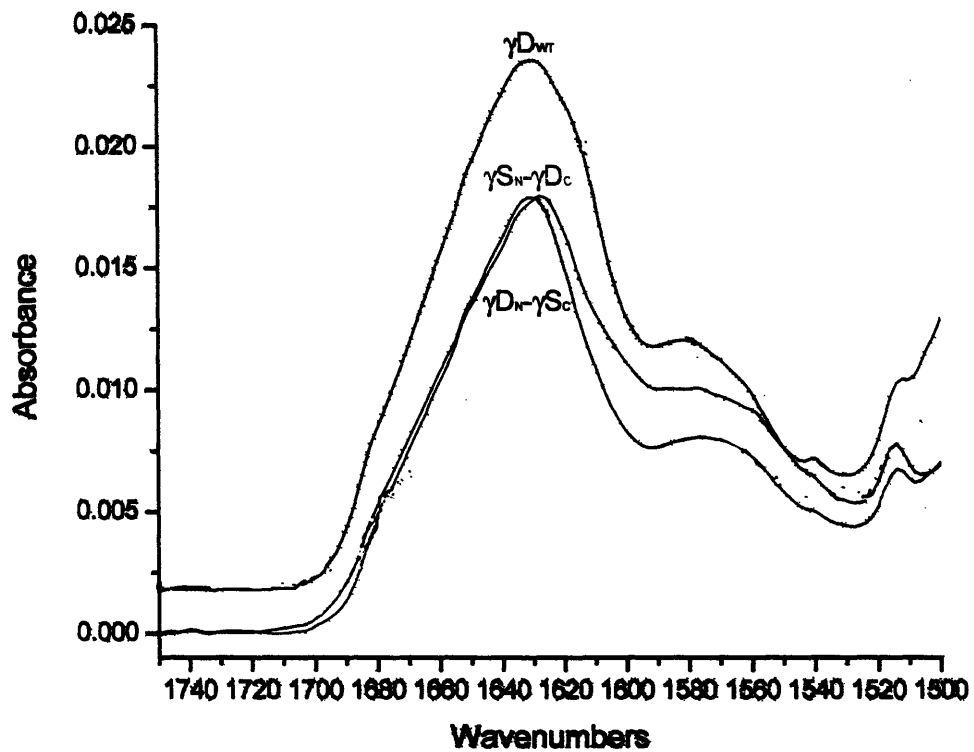


Figure 4-12. Infrared Spectroscopy of the wild type and the $\gamma D_N-\gamma S_C$ and $\gamma S_N-\gamma D_C$ chimeras. Proteins were refolded at high concentrations 10 mg/ml, in deuterated 100 mM sodium phosphate, 1 mM EDTA, 5 mM DTT, pH 7.0 buffer. 512 scans averaged to achieve final interferogram.

D. DISCUSSION:

In this chapter, the competing off-pathway polymerization reaction upon refolding of $\gamma\text{D}_{\text{WT}}$ out of denaturant was investigated. The previously characterized polymerization reaction revealed highly ordered fibrous aggregates that were not amyloid in nature, but were ordered. This aggregation reaction of $\gamma\text{D}_{\text{WT}}$ polypeptide chains may be an model for cataractogenesis *in vitro*. The structurally homologous $\gamma\text{S}_{\text{WT}}$ crystallin did not exhibit an off-pathway aggregation under the same conditions as $\gamma\text{D}_{\text{WT}}$. This suggested that the pathway of aggregation involved specific amino acids or sequences essential for association and was not a general feature of the γ -crystallins. To investigate this disparity between two structurally similar crystallins, chimera proteins were created in an attempt to narrow down regions of the protein that promoted aggregation in $\gamma\text{D}_{\text{WT}}$ or regions in $\gamma\text{S}_{\text{WT}}$ that inhibited aggregation.

Chimeras consisting of one domain of $\gamma\text{D}_{\text{WT}}$ and one domain of $\gamma\text{S}_{\text{WT}}$ were created along with chimeras that restored the natural γD interface. All of these proteins were able to fold and CD analysis demonstrated primarily β -sheet character similar to $\gamma\text{D}_{\text{WT}}$ or $\gamma\text{S}_{\text{WT}}$ secondary structure. The chimeras with the N-terminal domain of the wild type proteins had similar characteristics to their respective protein. For example, $\gamma\text{D}_{\text{N}}\text{-}\gamma\text{S}_{\text{C}}$ chimeras had characteristics similar to $\gamma\text{D}_{\text{WT}}$. This was unexpected since the structure and stability of the isolated domains demonstrated that the C-terminal domains were most similar to the wild type proteins. This surprising result may be due to the presence and effectiveness of the interdomain interface. The contribution of the interface may support structural aspects of the protein as well as stability.

1. Structural Aspects of the Chimera Proteins

The topology of the β - and γ -crystallins is identical, consisting of two domains with each domain equipped with two intercalated anti-parallel Greek Key motifs. In spite of being structurally similar, the β -crystallins form oligomers and the γ -crystallins are

monomeric. Previously, comparisons between β B2 and Bovine γ B crystallins have been used as a model to study the differences in associations observed in the $\beta\gamma$ -crystallin proteins (Mayr et al. 1994; Trinkl et al. 1994). In the case of β B2 crystallin, it can form a domain swapped dimer and can also form tetramers and possibly hetero-tetramers with other β -crystallins (Norledge et al. 1997).

The γ -crystallins have a conserved proline and glycines in the linker, thought to be important for their bent linker character. β -crystallins also have a proline in its linker albeit not in the same structural position (Lapatto et al. 1991). It was of concern that the Pro 83 of both γ D_N- γ S_C and γ S_N- γ D_C chimeras were not in same topological position as Pro 82 of γ B possibly causing these the domains to dimerize or be destabilized (Lapatto et al. 1991). However, analytical size exclusion analysis demonstrated that at concentrations above experimental conditions, all of the chimeras were monomeric. Therefore, the presence of a proline in the linker was sufficient to prevent dimerization. In addition, thermal stability of the chimeras only revealed a slight decrease in stability most likely a consequence of the different domain contributions. However, without the solved structures of the chimeras, it is difficult to assess what effect the linker might have on the structure of the chimeras.

2. Stability of the Chimera Proteins

The stability of the chimeras was determined both qualitatively by thermal denaturation and quantitatively by equilibrium unfolding/refolding experiments. Thermally, the behavior of the chimeras was similar to the wild type proteins, γ D_{WT} and γ S_{WT}. The loss of secondary structure monitored by CD was not observed until temperatures above 65°C. All thermal unfolding curves were fit to a two state model differing from the equilibrium unfolding/refolding curves of γ D_{WT} and γ S_N- γ D_C chimeras which are better fit to three state models. However, since the CD signal may be a mixture of aggregate and unfolded species, any intermediates along the thermal unfolding pathway would be difficult to decipher.

The T_{MS} of the γD_N - γS_C and γD_N - γS_C [γD_{INT}] chimeras, containing the γD N-terminal domain, were the closest to γD_{WT} 's T_M of 83.8°C. On the contrary, the γS N-terminus containing chimeras, γS_N - γD_C and γS_N - γD_C [γD_{INT}], had T_{MS} similar to γS_{WT} T_M of 74°C. These results were contrary to what one might expect since the N-terminal domains of both γD and γS are the least stable in isolation. However, the stability of the interdomain interface in the γD_N - γS_C chimeras may contribute to the overall stability of the proteins. While, the domain interface may not be a contributing factor to the overall stability of the γS_N - γD_C chimeras. The equilibrium unfolding/refolding experiments confirmed the thermal denaturation results and yielded comparable qualitative stability differences in the chimeras.

Contrary to our hypothesis, restoration of the γD interface does not significantly affect the stability of the chimeras. Both γD interface chimeras had similar thermodynamic and thermal stability values compared to their parent chimeras. These results may be a result of the lack of maximum interface contacts due to structural differences in the interface, which varies from the wild type protein. Alternatively, the intrinsic domains may be the primary factor in the thermodynamic stability of the overall protein. Thus, “fixing” the interface will not necessarily restore stability observed in the wild type thermodynamic values.

3. Aggregation of the Chimeras

Although previous studies of the γD_{WT} demonstrated an off-pathway aggregation reaction, the aggregation reaction was not observed for γS_{WT} under the same conditions as γD_{WT} . Additionally, none of the isolated domains demonstrated the aggregation. Moreover, refolding γD isolated domains together out of denaturant at equal concentrations of 10 $\mu\text{g/ml}$ did not reconstitute the aggregation reaction. Therefore, both domains of γD_{WT} were necessary to reconstitute the aggregation observed in γD_{WT} .

Analysis of the chimeras demonstrated that the γD_{WT} interdomain interface present in both domains was the key factor in the aggregation reaction. Neither γD_N - γS_C nor γS_N - γD_C chimera recapitulated the solution turbidity observed in the refolding aggregate of γD_{WT} under similar conditions. Incorporation of the γD interface in chimeras increased aggregation significantly. These results suggest that the interface residues in both domains of γD_{WT} contribute to aggregation observed at 10 $\mu\text{g/ml}$ protein concentration. Under these conditions, the γD_N - γS_C chimera exhibited solution turbidity slightly above its native protein and the refolded γS_{WT} solution turbidity. In contrast, the opposite chimera, γS_N - γD_C did not have a significant increase above background solution turbidity. These results suggest that the N-terminus of γD has a higher propensity to aggregate when paired with a partner domain.

This higher propensity could be due to γD_N domain interface or an intrinsic residue that is exposed during the refolding reaction. One noticeable difference in the interface between γD_{WT} and γS_{WT} is the Phe at position 56 in γD substituted for Ile in γS . Although the Phe56 might be necessary for the domain swapping mechanism, it is not sufficient since substitutions of Phe to Ala did not inhibit the off-pathway aggregation reaction (Flaugh et al. 2005b). Aggregation kinetics performed on mutations in γD may give more insight into the specific interface residues that contribute to aggregation. The chimera aggregation results support the idea that the interface of γD_{WT} promotes aggregation and in contrast to the hypothesis that regions in γS_{WT} inhibit aggregation.

In contrast to the results at lower protein concentration, the chimera aggregation analysis at higher concentrations indicated that the chimera proteins had a higher propensity to aggregate than the γS_{WT} protein. Thus, the inclusion of either domain from γD_{WT} is sufficient to initiate aggregation of γS regions at higher concentrations. This is likely due the presence of more of the aggregation-prone intermediates at higher concentration promoting the nucleation event for aggregation. Again, these results suggest that a region in γD_{WT} is promoting aggregation instead of the hypothesis that a region in γS_{WT} inhibits aggregation.

4. Summary of the Chimera Structural, Stability, and Aggregation Properties

The results of all chimera experiments are summarized in Table 4-4. Briefly, the chimeras containing the N-terminal domain of the wild type proteins had similar properties. For example, $\gamma_{D_N}\text{-}\gamma_{S_C}$ chimeras had similar secondary and tertiary structure to $\gamma_{D_{WT}}$. Also, their thermal stability was similar to $\gamma_{D_{WT}}$. Thermodynamically, the $\gamma_{D_N}\text{-}\gamma_{S_C}$ chimeras were not similar because the transitions were more cooperative compared to $\gamma_{D_{WT}}$ and individual domain transitions could not be resolved. However, the contribution of the domain interface in the stabilization of the γ_D N-terminal was reminiscent of $\gamma_{D_{WT}}$. The opposite chimera containing the γ_S N-terminal domain, $\gamma_{S_N}\text{-}\gamma_{D_C}$ and $\gamma_{S_N}\text{-}\gamma_{D_C}$ [$\gamma_{D_{INT}}$] had properties similar to $\gamma_{S_{WT}}$. In addition, the γ_D interface containing chimeras aggregated while the other chimeras did not.

Table 4-4. Summary of γ D and γ S wild type and chimera structural, stability, and refolding aggregation properties.

PROTEIN	Secondary/Tertiary Structure	Thermal Stability	Refolding Aggregation (10 μg/ml protein)
γ D _{WT}	-	-	Yes
γ D _N - γ S _C	Similar to γ D _{WT}	Similar to γ D _{WT}	Slightly
γ D _N - γ S _C [γ D _{INT}]	Similar to γ D _{WT}	Similar to γ D _{WT}	Yes
γ S _{WT}	-	-	No
γ S _N - γ D _C	Similar to γ S _{WT}	Similar to γ S _{WT}	No
γ S _N - γ D _C [γ D _{INT}]	Similar to γ S _{WT}	Similar to γ S _{WT}	Yes

5. A Domain-Swapping Mechanism for γD_{WT} Aggregation Model

In general, aggregation-prone species may consist of intermediates along the normal folding pathway of the protein or non-native conformers of the protein. All of these results suggest a domain swapping mechanism as a plausible model of the γD_{WT} polymerization reaction (Fig. 4-13). For example, the N-terminal of one partially folded conformer interacts intermolecularly with the C-terminal of another molecule forming open-ended multimers and leading to a highly ordered distinct aggregated state. In this model, the domain swapped interface would have the same molecular contacts as the native interface. The reconstitution of aggregation by the γD interface chimeras supports the domain swapping model for aggregation suggesting that the γD interdomain interface is the primary driving force in the aggregation reaction. Domain swapping has been proposed for other aggregation pathways. For example, open-ended extended polymers in solution have been observed for RNase A, tryptophanase, and tryptophan synthetase (Schlunegger et al. 1997). The open-ended domain-swapped polymerization has been confirmed by the filament crystal structure of the T7 helicase (Sawaya et al. 1999), RecA (Story et al. 1992; Xing and Bell 2004) and carbonic anhydrase proteins (Strop et al. 2001). Mechanisms of how these proteins form domain swapped polymers are still being studied.

Additionally, the domain swapping mechanism is an attractive model since the $\beta B2$ crystallin has been shown to form a close-ended domain swapped dimer naturally (Fig. 4-14). The $\beta B2$ crystallin associates via a domain swapping mechanism in which the N-terminal domain associates intermolecularly with its dimeric partner's C-terminal domain. Equilibrium studies showed that the N-terminal domain dimerized at higher concentrations increasing its stability while at lower concentrations (5 $\mu\text{g/ml}$) it was in the monomeric form and had a higher propensity to unfold. In context of the full-length protein, the equilibrium unfolding/refolding was more cooperative at higher concentrations, observed by a two state model. In contrast, at lower concentration the population of an intermediate was observed and the equilibrium data fit best to a three state model. These results suggested that the unfolding of the N-terminal domain

promotes dissociation of the N-terminus from the C-terminus of its partner molecule, forming a monomeric intermediate in which the N-terminus is unstructured and the C-terminus is folded. Thus, the second equilibrium transition monitors the unfolding of the C-terminal domain (Fu and Liang 2002b; Wieligmann et al. 1999).

However, we are not ruling out other possibilities such as the loop-sheet insertion mechanism involving one or more β -strands in the interface that inserts within a domain or another mechanism involving intermolecular contacts that include the interface (Carrell and Lomas 1997).

A lag phase is not observed in the aggregation pathway using this method; which is not consistent with a nucleation dependent aggregation reaction. The extent of aggregation was concentration dependent. In the domain swapped model, the initiating species could be an open ended domain swapped dimer which could lead to linear unbranched fibrils as seen in the AFM. The unbranched fibrils observed in the AFM could be created by a double domain swapped mechanism in which two structural elements of the protein would be involved. An example of a double domain swap creating branched and unbranched fibrils is the *E. coli* trp repressor structure (Lawson et al. 2004). Alternatively, the unbranched and branched fibrils can be formed through a closed-ended mechanism. This mechanism has been proposed for proteins such as cystatin C, which forms amyloid fibrils observed in its crystal structure (Janowski et al. 2005). To date, none of these structures have been observed *in vivo*.

There are a few cases reported in which proteins with similar topologies have different aggregation rates. In the case of Ig-binding proteins, protein L and G, each have different kinetic refolding pathways which have been studied extensively (McCallister et al. 2000; Park et al. 1997; Scalley et al. 1997). The rate of aggregation in these two proteins is different. Comparatively, protein G is a slow aggregator while protein L aggregates quickly (Fawzi et al. 2005). These diverse aggregation properties have been attributed to the folding intermediates populated along the folding pathway (Brown and Head-Gordon 2004). It has been shown that Protein G has a rapid folded intermediate

state that in theoretical studies prevents “native-like” associations with other molecules. On the other hand, theoretical studies on protein L have shown that protein L aggregates more rapidly, since its denatured state has more stabilizing interactions thought to be aggregation-prone (Fawzi et al. 2005).

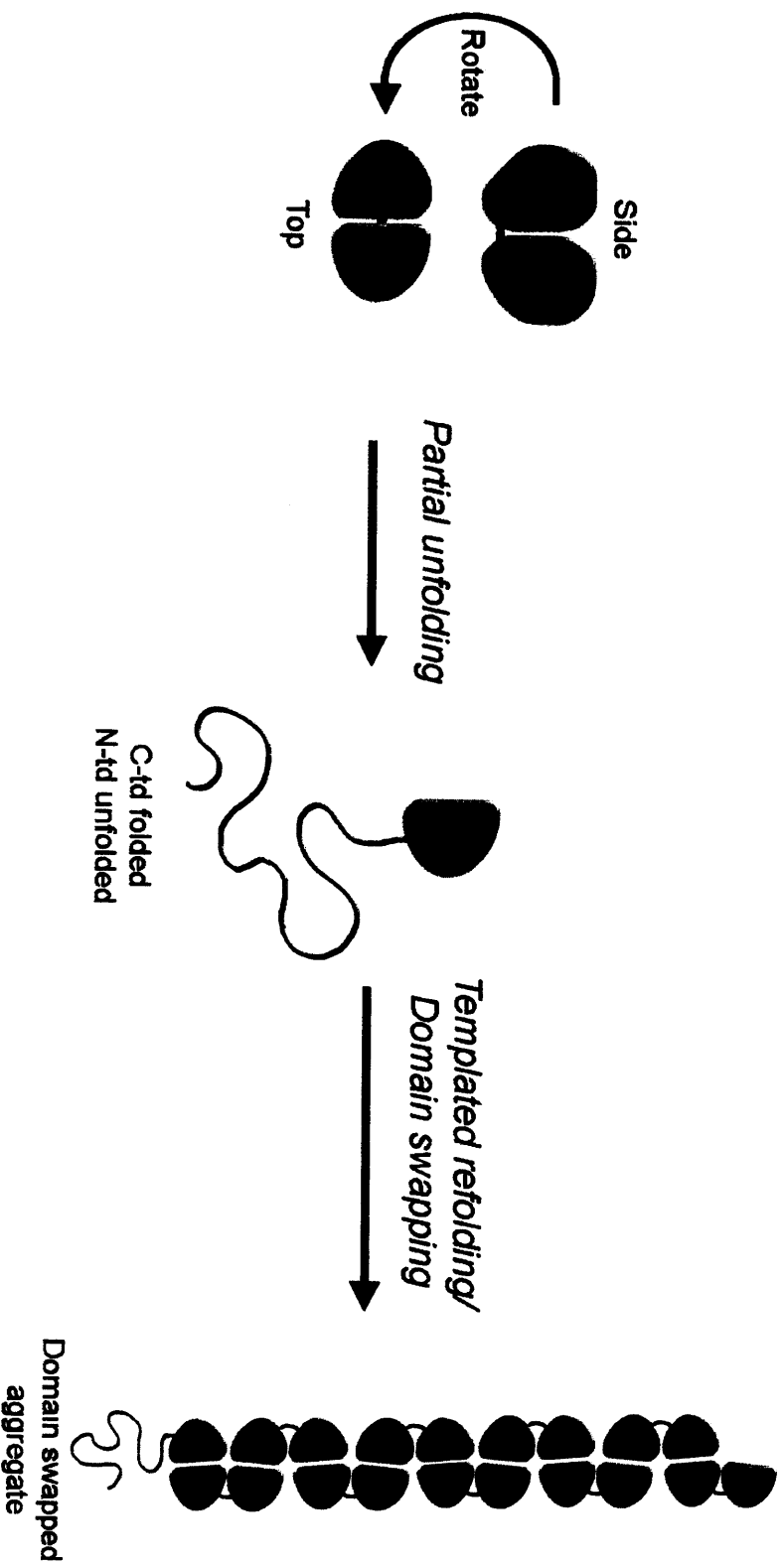


Figure 4-13. Mechanism of γ DWT Aggregation - Domain Swapping Model.
 A partially folded conformer with the interface exposed interacts with another molecule of the same conformation.



Figure 4-14. Crystal structure of domain swapped dimer in bovine β B2 crystallin. Example of the crystallin's ability to oligomerize by domain swapping. (Bax et al. Nature, 1990, 347:6295 776-80).

CHAPTER FIVE:

FINAL DISCUSSION

A. SUMMARY OF CHAPTERS

In the last section of the Chapter one, I noted that the motivating questions for me for this project as a cell biologist focused on why the crystallin proteins were duplicated and why so many different crystallins were required and located in different regions of the lens were of interest. In Chapter two, the thermal studies of the isolated domains showed that duplication increases stability in γ D crystallin. Although, qualitative, the thermal studies are a more physiologically relevant measure of stability, since heat is a common stress. However, the isolated domains were extremely stable as well. This is consistent with the notion that the crystallins were recruited to the lens possibly for their biophysical properties as well as their proposed stress related roles.

The results in Chapter two, high stability of the γ D crystallin involves the intrinsic stability of the domains as well as the interdomain interface. Mutants of the interface residues suggested that the stability of the interdomain interface was important for the folding of the N-terminal domain (Flaugh et al. 2006). The chapter two results confirmed this previous result and allowed for a calculation of the free energy contribution of the interface. Another important result from this chapter was in absence of its partner interface, the N-terminal domain was still able to fold. The stability of the domains in isolation also suggests that the ancestral single domain crystallins were already quite stable. In Figure 5-1, the interface of the N-terminal domain and C-terminal domain is observed.

Another important result from these experiments was the detection of a kinetic refolding intermediate in the formation of the intercalated double Greek Key domain. The intermediate detected could correlate to the folding of one individual Greek Key motif. It has been proposed that the innermost Greek Key motif act as a nucleus for folding the entire domain. By studying the isolated domain refolding kinetics, we propose that this observed intermediate may be the innermost Greek Key folded and the outermost Greek Key unfolded.

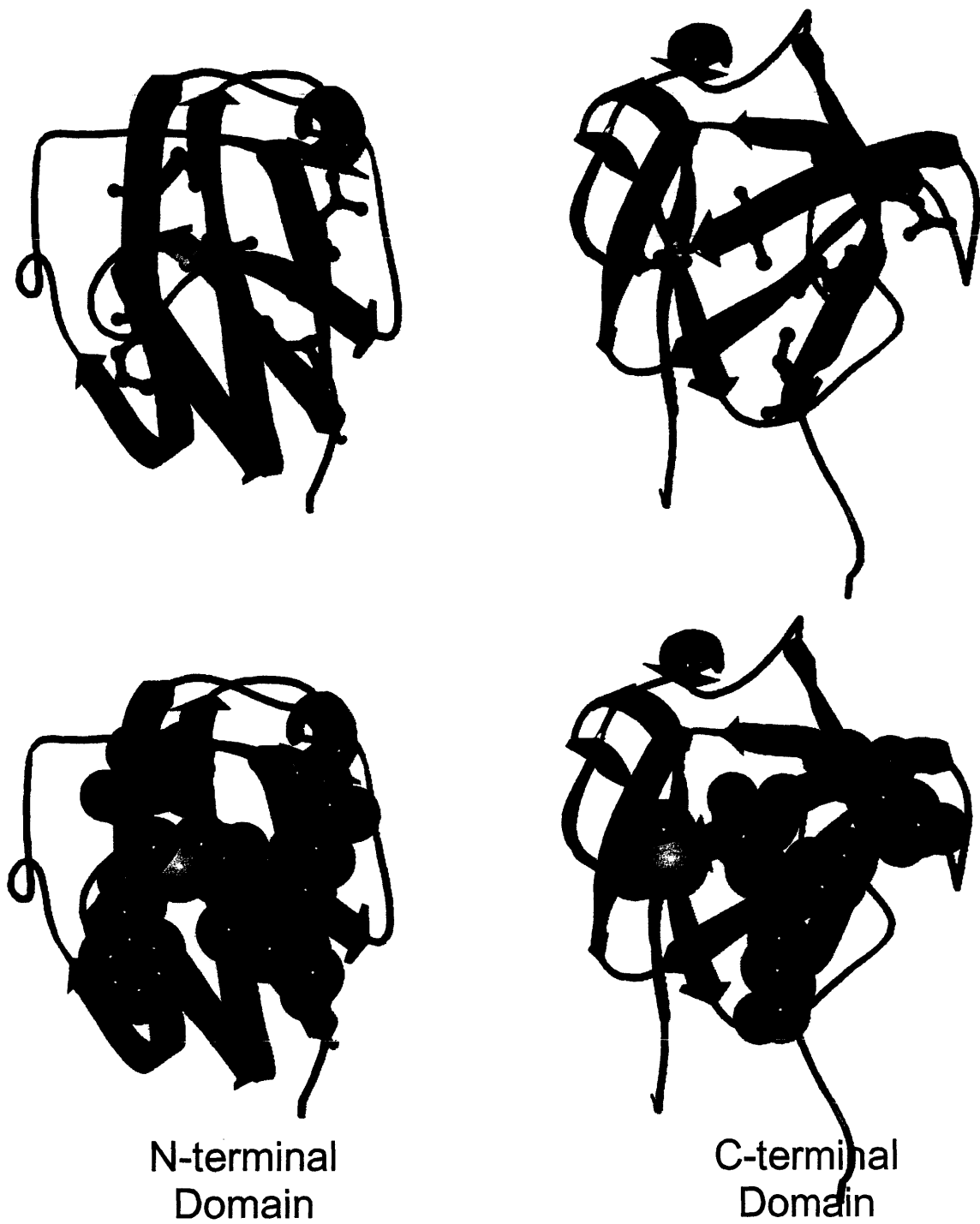


Figure 5-1. Interface of the γ D N- and C-terminal domain. Interface residues are represented in ball and stick (top) and space filled (bottom).

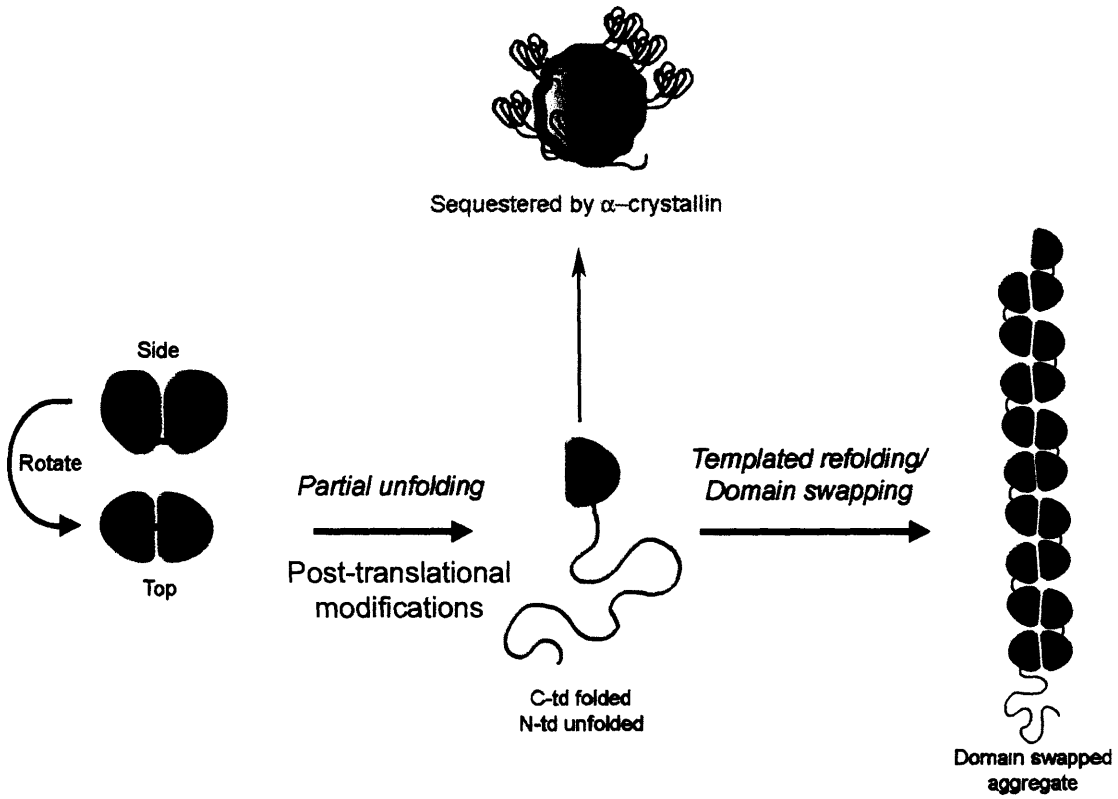


Figure 5-2. Model of Cataract with Domain Swapping Interface of γ D-crystallin.

In Chapter three, I discovered that the γ -crystallins that would require the longest life-time exhibited a distinct kinetic stability consistent with its biological function. γ D crystallin exhibited high kinetic stability in addition to its high thermodynamic stability. On the other hand, γ S crystallin did not exhibit such high stability kinetically, thermally or thermodynamically. This data supported the hypothesis that the crystallins expressed in different regions of the lens have adapted to their surrounding environments. The γ S crystallin is present in the outer regions of the lens where there is synthesis and degradation. The nuclear lens core γ D crystallin localization requires the high kinetic stability.

Our results here observed half-lives close to 20 years for the lens nuclear core γ D crystallin. These numbers may increase in the environment of the lens due to macromolecular crowding. It has been suggested that high concentrations in the cell improves kinetic stability through. Also excluded volume improves stability by macromolecular crowding. The molecular crowding could assist crystallin binding to its partially folded substrates. In addition, the presence of intermediates would not accumulate in appreciable amounts since the equilibrium between the unfolded or intermediate state and the native state would drive the partially unfolded intermediates back to the native state. It is predicted that the post-translational modifications would shift this equilibrium by destabilizing the native-state.

Although these experiments were performed at low concentrations compared to the high concentration of the lens proteins, these results may still be physiologically relevant. In the lens, the amount of partially folded intermediate that would be populated upon a destabilizing event would be at relatively low concentrations. The presence of α -crystallin would also have an effect on the concentrations of the partially folded intermediates. However, late in life the α -crystallin may be saturated and no longer available to prevent aggregation of the other crystallins.

In Chapter four, I have observed that there comes a price with increased stability. In γ D, the stabilizing interdomain interface residues are associated with increased aggregation. This result supports the necessity for the high stability of γ D crystallin, preventing the accumulation of partially folded intermediates with an exposed interface. The results of Chapter four supports a domain swapping model for the formation of the aggregation state of γ D-crystallin and perhaps other crystallins as well (Figure 5-2).

An interesting observation of this thesis work is the differential thermodynamic and kinetic stability between two homologous crystallins. While this difference may be explained by the adaptation of these crystallins in two different regions of the lens, the biochemical differences that contribute to the stability are still unknown. By comparing these two crystallins by sequence or structural alignment, it is difficult to assess the differences in the stability. By studying mutations between both crystallins, the amino acids important for the overall stability may be determined in the future. Thus, the crystallins are model proteins for understanding β -sheet folding, amino acid interactions contributions to stability and protein aggregation. In addition to these general important unanswered questions in biology, the comparison of the crystallins promotes understanding of protein evolution and how it relates to its function.

CHAPTER SIX: REFERENCES

- Aarts, H.J., Lubsen, N.H. and Schoenmakers, J.G., 1989. Crystallin gene expression during rat lens development. *Eur J Biochem* **183**: 31-36.
- Akke, M., and Forsen, S., 1990. Protein stability and electrostatic interactions between solvent exposed charged side chains. *Proteins* **8**: 23-29.
- Anderson, D.E., Peters, R.J., Wilk, B. and Agard, D.A., 1999. Alpha-lytic protease precursor: Characterization of a structured folding intermediate. *Biochemistry* **38**: 4728-4735.
- Aravind, P., Rajini, B., Sharma, Y. and Sankaranarayanan, R., 2006. Crystallization and preliminary X-ray crystallographic investigations on a betagamma-crystallin domain of absent in melanoma 1 (AIM1), a protein from homo sapiens. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **62**: 282-284.
- Arendt, D., 2003. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol* **47**: 563-571.
- Bagby, S., Go, S., Inouye, S., Ikura, M. and Chakrabarty, A., 1998. Equilibrium folding intermediates of a greek key beta-barrel protein. *J Mol Biol* **276**: 669-681.
- Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. and Ikura, M., 1994a. Structural similarity of a developmentally regulated bacterial spore coat protein to beta gamma-crystallins of the vertebrate eye lens. *Proc Natl Acad Sci U S A* **91**: 4308-4312.
- Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. and Ikura, M., 1994b. NMR-derived three-dimensional solution structure of protein S complexed with calcium. *Structure* **2**: 107-122.
- Bagby, S., Harvey, T.S., Kay, L.E., Eagle, S.G., Inouye, S. and Ikura, M., 1994c. Unusual helix-containing greek keys in development-specific ca(2+)-binding protein S. ¹H, ¹⁵N, and ¹³C assignments and secondary structure determined with the use of multidimensional double and triple resonance heteronuclear NMR spectroscopy. *Biochemistry* **33**: 2409-2421.
- Baryshnikova, E.N., Melnik, B.S., Finkelstein, A.V., Semisotnov, G.V. and Bychkova, V.E., 2005. Three-state protein folding: Experimental determination of free-energy profile. *Protein Sci* **14**: 2658-2667.
- Basak, A., Bateman, O., Slingsby, C., Pande, A., Asherie, N., Ogun, O., Benedek, G.B. and Pande, J., 2003. High-resolution X-ray crystal structures of human gammaD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J Mol Biol* **328**: 1137-1147.

- Baskakov, I.V., Legname, G., Prusiner, S.B. and Cohen, F.E., 2001. Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J Biol Chem* **276**: 19687-19690.
- Bateman, O.A., Lubsen, N.H. and Slingsby, C., 2001. Association behaviour of human betaB1-crystallin and its truncated forms. *Exp Eye Res* **73**: 321-331.
- Bateman, O.A., Sarra, R., van Genesen, S.T., Kappe, G., Lubsen, N.H. and Slingsby, C., 2003. The stability of human acidic beta-crystallin oligomers and hetero-oligomers. *Exp Eye Res* **77**: 409-422.
- Bax, B., Lapatto, R., Nalini, V., Driessen, H., Lindley, P.F., Mahadevan, D., Blundell, T.L. and Slingsby, C., 1990a. X-ray analysis of beta B2-crystallin and evolution of oligomeric lens proteins. *Nature* **347**: 776-780.
- Benedek, G.B., 1997. Cataract as a protein condensation disease: The proctor lecture. *Invest Ophthalmol Vis Sci* **38**: 1911-1921.
- Bennett, M.J., Sawaya, M.R. and Eisenberg, D., 2006. Deposition diseases and 3D domain swapping. *Structure* **14**: 811-824.
- Bennett, M.J., Choe, S. and Eisenberg, D., 1994. Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci U S A* **91**: 3127-3131.
- Bernier, F., Lemieux, G. and Pallotta, D., 1987. Gene families encode the major encystment-specific proteins of physarum polycephalum plasmodia. *Gene* **59**: 265-277.
- Bernier, F., Pallotta, D. and Lemieux, G., 1986. Molecular cloning of mRNAs expressed specifically during spherulation of physarum polycephalum. *Biochim Biophys Acta* **867**: 234-243.
- Bettelheim, F.A., and Siew, E.L., 1983. Effect of change in concentration upon lens turbidity as predicted by the random fluctuation theory. *Biophys J* **41**: 29-33.
- Betts, S.D., and King, J., 1998. Cold rescue of the thermolabile tailspike intermediate at the junction between productive folding and off-pathway aggregation. *Protein Sci* **7**: 1516-1523.
- Bieri, O., and Kiefhaber, T., 1999. Elementary steps in protein folding. *Biol Chem* **380**: 923-929.
- Biswas, A., and Das, K.P., 2004. Role of ATP on the interaction of alpha-crystallin with its substrates and its implications for the molecular chaperone function. *J Biol Chem* **279**: 42648-42657.

- Blanchard, B.J., Chen, A., Rozeboom, L.M., Stafford, K.A., Weigele, P. and Ingram, V.M., 2004. Efficient reversal of alzheimer's disease fibril formation and elimination of neurotoxicity by a small molecule. *Proc Natl Acad Sci U S A* **101**: 14326-14332.
- Bloemendal, H., de Jong, W., Jaenicke, R., Lubsen, N.H., Slingsby, C. and Tardieu, A., 2004. Ageing and vision: Structure, stability and function of lens crystallins. *Prog Biophys Mol Biol* **86**: 407-485.
- Blundell, T., Lindley, P., Miller, L., Moss, D., Slingsby, C., Tickle, I., Turnell, B. and Wistow, G., 1981. The molecular structure and stability of the eye lens: X-ray analysis of gamma-crystallin II. *Nature* **289**: 771-777.
- Brakenhoff, R.H., Aarts, H.J., Reek, F.H., Lubsen, N.H. and Schoenmakers, J.G., 1990. Human gamma-crystallin genes. A gene family on its way to extinction. *J Mol Biol* **216**: 519-532.
- Brems, D.N., and Havel, H.A., 1989. Folding of bovine growth hormone is consistent with the molten globule hypothesis. *Proteins* **5**: 93-95.
- Brems, D.N., Plaisted, S.M., Havel, H.A. and Tomich, C.S., 1988. Stabilization of an associated folding intermediate of bovine growth hormone by site-directed mutagenesis. *Proc Natl Acad Sci U S A* **85**: 3367-3371.
- Bron, A.J., Vrensen, G.F., Koretz, J., Maraini, G. and Harding, J.J., 2000. The ageing lens. *Ophthalmologica* **214**: 86-104.
- Brown, B.M., and Sauer, R.T., 1999. Tolerance of arc repressor to multiple-alanine substitutions. *Proc Natl Acad Sci U S A* **96**: 1983-1988.
- Brown, S., and Head-Gordon, T., 2004. Intermediates and the folding of proteins L and G. *Protein Sci* **13**: 958-970.
- Burley, S.K., and Petsko, G.A., 1985. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science* **229**: 23-28.
- Carrell, R.W., and Huntington, J.A., 2003. How serpins change their fold for better and for worse. *Biochem Soc Symp* (70): 163-178.
- Carrell, R.W., and Lomas, D.A., 2002. Alpha1-antitrypsin deficiency--a model for conformational diseases. *N Engl J Med* **346**: 45-53.
- Carrell, R.W., and Lomas, D.A., 1997. Conformational disease. *Lancet* **350**: 134-138.
- Caughey, B., and Lansbury, P.T., 2003. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26**: 267-298.

- Chambers, C., and Russell, P., 1991. Deletion mutation in an eye lens beta-crystallin. an animal model for inherited cataracts. *J Biol Chem* **266**: 6742-6746.
- Chen, J., Flaugh, S.L., Callis, P.R. and King, J., 2006. Mechanism of the highly efficient quenching of tryptophan fluorescence in human gammaD-crystallin. *Biochemistry* **45**: 11552-11563.
- Clark, A.C., Sinclair, J.F. and Baldwin, T.O., 1993. Folding of bacterial luciferase involves a non-native heterodimeric intermediate in equilibrium with the native enzyme and the unfolded subunits. *J Biol Chem* **268**: 10773-10779.
- Clark, J.I., and Muchowski, P.J., 2000. Small heat-shock proteins and their potential role in human disease. *Curr Opin Struct Biol* **10**: 52-59.
- Clarke, J., Cota, E., Fowler, S.B. and Hamill, S.J., 1999. Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* **7**: 1145-1153.
- Clout, N.J., Slingsby, C. and Wistow, G.J., 1997. Picture story. an eye on crystallins. *Nat Struct Biol* **4**: 685.
- Clout, N.J., Kretschmar, M., Jaenicke, R. and Slingsby, C., 2001. Crystal structure of the calcium-loaded spherulin 3a dimer sheds light on the evolution of the eye lens betagamma-crystallin domain fold. *Structure* **9**: 115-124.
- Colon, W., and Kelly, J.W., 1992. Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31**: 8654-8660.
- Cook, C.A., Koretz, J.F., Pfahnl, A., Hyun, J. and Kaufman, P.L., 1994. Aging of the human crystalline lens and anterior segment. *Vision Res* **34**: 2945-2954.
- Costello, M.J., Oliver, T.N. and Cobo, L.M., 1992. Cellular architecture in age-related human nuclear cataracts. *Invest Ophthalmol Vis Sci* **33**: 3209-3227.
- Costello, M.J., Lane, C.W., Hatchell, D.L., Saloupis, P. and Cobo, L.M., 1993. Ultrastructure of fiber cells and multilamellar inclusions in experimental diabetes. *Invest Ophthalmol Vis Sci* **34**: 2174-2185.
- Cunningham, E.L., and Agard, D.A., 2003. Interdependent folding of the N- and C-terminal domains defines the cooperative folding of alpha-lytic protease. *Biochemistry* **42**: 13212-13219.
- Cunningham, E.L., Jaswal, S.S., Sohl, J.L. and Agard, D.A., 1999. Kinetic stability as a mechanism for protease longevity. *Proc Natl Acad Sci U S A* **96**: 11008-11014.
- Dao-pin, S., Nicholson, H., Baase, W.A., Zhang, X.J., Wozniak, J.A. and Matthews, B.W., 1991. Structural and genetic analysis of electrostatic and other interactions in bacteriophage T4 lysozyme. *Ciba Found Symp* **161**: 52-62.

- Das, B.K., and Liang, J.J., 1998. Thermodynamic and kinetic characterization of calf lens gammaF-crystallin. *Int J Biol Macromol* **23**: 191-197.
- Das, B.K., and Liang, J.J., 1997. Detection and characterization of alpha-crystallin intermediate with maximal chaperone-like activity. *Biochem Biophys Res Commun* **236**: 370-374.
- Das, B.K., Liang, J.J. and Chakrabarti, B., 1997. Heat-induced conformational change and increased chaperone activity of lens alpha-crystallin. *Curr Eye Res* **16**: 303-309.
- de Jong, W.W., Caspers, G.J. and Leunissen, J.A., 1998. Genealogy of the alpha-crystallin--small heat-shock protein superfamily. *Int J Biol Macromol* **22**: 151-162.
- Delaye, M., and Tardieu, A., 1983. Short-range order of crystallin proteins accounts for eye lens transparency. *Nature* **302**: 415-417.
- D'onofrio, V., and Mosci, L., 1960. Cataract caused by heat: Observations in metallurgic workers and glass blowers. *Rass Med Ind Ig Lav* **29**: 253-267.
- Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P. and Matthews, B.W., 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**: 178-183.
- Erwin, C.R., Barnett, B.L., Oliver, J.D. and Sullivan, J.F., 1990. Effects of engineered salt bridges on the stability of subtilisin BPN'. *Protein Eng* **4**: 87-97.
- Evans, P., Wyatt, K., Wistow, G.J., Bateman, O.A., Wallace, B.A. and Slingsby, C., 2004. The P23T cataract mutation causes loss of solubility of folded gammaD-crystallin. *J Mol Biol* **343**: 435-444.
- Fagerholm, P., Philipson, B. and Carlstrom, D., 1981a. Calcification in the human lens. *Curr Eye Res* **1**: 629-633.
- Fagerholm, P.P., Philipson, B.T. and Lindstrom, B., 1981b. Normal human lens - the distribution of protein. *Exp Eye Res* **33**: 615-620.
- Fawzi, N.L., Chubukov, V., Clark, L.A., Brown, S. and Head-Gordon, T., 2005. Influence of denatured and intermediate states of folding on protein aggregation. *Protein Sci* **14**: 993-1003.
- Fernald, R.D., 2000. Evolution of eyes. *Curr Opin Neurobiol* **10**: 444-450.
- Fernald, R.D., and Wright, S.E., 1983. Maintenance of optical quality during crystalline lens growth. *Nature* **301**: 618-620.
- Fersht, A., 1999. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W.H. Freeman and Company, New York.

- Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M. and Winter, G., 1985. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**: 235-238.
- Flaugh, S.L., Mills, I.A. and King, J., 2006. Glutamine deamidation destabilizes human gammaD-crystallin and lowers the kinetic barrier to unfolding. *J Biol Chem* **281**: 30782-30793.
- Flaugh, S.L., Kosinski-Collins, M.S. and King, J., 2005a. Interdomain side-chain interactions in human gammaD crystallin influencing folding and stability. *Protein Sci* **14**: 2030-2043.
- Flaugh, S.L., Kosinski-Collins, M.S. and King, J., 2005b. Contributions of hydrophobic domain interface interactions to the folding and stability of human gammaD-crystallin. *Protein Sci* **14**: 569-581.
- Freel, C.D., Gilliland, K.O., Mekeel, H.E., Giblin, F.J. and Costello, M.J., 2003. Ultrastructural characterization and fourier analysis of fiber cell cytoplasm in the hyperbaric oxygen treated guinea pig lens opacification model. *Exp Eye Res* **76**: 405-415.
- Fu, L., and Liang, J.J., 2003. Alteration of protein-protein interactions of congenital cataract crystallin mutants. *Invest Ophthalmol Vis Sci* **44**: 1155-1159.
- Fu, L., and Liang, J.J., 2002a. Conformational change and destabilization of cataract gammaC-crystallin T5P mutant. *FEBS Lett* **513**: 213-216.
- Fu, L., and Liang, J.J., 2002b. Unfolding of human lens recombinant betaB2- and gammaC-crystallins. *J Struct Biol* **139**: 191-198.
- Ganea, E., and Harding, J.J., 2000. Alpha-crystallin protects glucose 6-phosphate dehydrogenase against inactivation by malondialdehyde. *Biochim Biophys Acta* **1500**: 49-58.
- Gehring, W.J., 2002. The genetic control of eye development and its implications for the evolution of the various eye-types. *Int J Dev Biol* **46**: 65-73.
- Gettins, P.G., 2000. Keeping the serpin machine running smoothly. *Genome Res* **10**: 1833-1835.
- Ghosh, S., and Mandal, D.K., 2006. Kinetic stability plays a dominant role in the denaturant-induced unfolding of erythrina indica lectin. *Biochim Biophys Acta* **1764**: 1021-1028.
- Gilliland, K.O., Freel, C.D., Lane, C.W., Fowler, W.C. and Costello, M.J., 2001. Multilamellar bodies as potential scattering particles in human age-related nuclear cataracts. *Mol Vis* **7**: 120-130.

- Graw, J., 2004. Congenital hereditary cataracts. *Int J Dev Biol* **48**: 1031-1044.
- Greene, R.F., Jr, and Pace, C.N., 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* **249**: 5388-5393.
- Gunasekaran, K., Hagler, A.T. and Gierasch, L.M., 2004. Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins* **54**: 179-194.
- Halder, G., Callaerts, P. and Gehring, W.J., 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in drosophila. *Science* **267**: 1788-1792.
- Haley, D.A., Horwitz, J. and Stewart, P.L., 1998. The small heat-shock protein, alphaB-crystallin, has a variable quaternary structure. *J Mol Biol* **277**: 27-35.
- Haley, D.A., Bova, M.P., Huang, Q.L., Mchaourab, H.S. and Stewart, P.L., 2000. Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies. *J Mol Biol* **298**: 261-272.
- Hanson, S.R., Hasan, A., Smith, D.L. and Smith, J.B., 2000. The major in vivo modifications of the human water-insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Exp Eye Res* **71**: 195-207.
- Harding JJ, Crabbe MJC, 1984. The lens: Development, proteins, metabolism and cataract, pp. 207-492 in *The Eye*, edited by Davson H. Academic Press, Orlando, FL.
- Harms, M.J., Wilmarth, P.A., Kapfer, D.M., Steel, E.A., David, L.L., Bachinger, H.P. and Lampi, K.J., 2004. Laser light-scattering evidence for an altered association of beta B1-crystallin deamidated in the connecting peptide. *Protein Sci* **13**: 678-686.
- Hejtmancik, J.F., Kaiser, M.I. and Piatigorsky, J., 2001. Molecular biology and inherited disorders of the eye lens, pp. 6033-6060 in *The Metabolic and Molecular Bases of Inherited Disease*, edited by C. Scriver, A. Beaudet, W. Sly and D. Valle. McGraw-Hill Medical Publishing Division, New York.
- Hemmingsen, J.M., Gernert, K.M., Richardson, J.S. and Richardson, D.C., 1994. The tyrosine corner: A feature of most greek key beta-barrel proteins. *Protein Sci* **3**: 1927-1937.
- Hendriks, W., Mulders, J.W., Bibby, M.A., Slingsby, C., Bloemendal, H. and de Jong, W.W., 1988. Duck lens epsilon-crystallin and lactate dehydrogenase B4 are identical: A single-copy gene product with two distinct functions. *Proc Natl Acad Sci USA* **85**: 7114-7118.
- Henry, J.J., and Grainger, R.M., 1990. Early tissue interactions leading to embryonic lens formation in *xenopus laevis*. *Dev Biol* **141**: 149-163.

- Heon, E., Priston, M., Schorderet, D.F., Billingsley, G.D., Girard, P.O., Lubsen, N. and Munier, F.L., 1999. The gamma-crystallins and human cataracts: A puzzle made clearer. *Am J Hum Genet* **65**: 1261-1267.
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M. and Fersht, A.R., 1990. Strength and co-operativity of contributions of surface salt bridges to protein stability. *J Mol Biol* **216**: 1031-1044.
- Horwitz, J., 2003. Alpha-crystallin. *Exp Eye Res* **76**: 145-153.
- Hughes, R.M., and Waters, M.L., 2006. Model systems for beta-hairpins and beta-sheets. *Curr Opin Struct Biol* **16**: 514-524.
- Huntington, J.A., Read, R.J. and Carrell, R.W., 2000. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**: 923-926.
- Inouye, M., Inouye, S. and Zusman, D.R., 1979. Biosynthesis and self-assembly of protein S, a development-specific protein of myxococcus xanthus. *Proc Natl Acad Sci USA* **76**: 209-213.
- Jaenicke, R., 1999. Stability and folding of domain proteins. *Prog Biophys Mol Biol* **71**: 155-241.
- Jaenicke, R., 1996. Stability and folding of ultrastable proteins: Eye lens crystallins and enzymes from thermophiles. *FASEB J* **10**: 84-92.
- Jaenicke, R., and Sterner, R., 2003. Protein design at the crossroads of biotechnology, chemistry, theory, and evolution. *Angew Chem Int Ed Engl* **42**: 140-142.
- Jaenicke, R., and Slingsby, C., 2001. Lens crystallins and their microbial homologs: Structure, stability, and function. *Crit Rev Biochem Mol Biol* **36**: 435-499.
- Jaenicke, R., and Bohm, G., 1998. The stability of proteins in extreme environments. *Curr Opin Struct Biol* **8**: 738-748.
- Jaenicke, R., and Seckler, R., 1997. Protein misassembly in vitro. *Adv Protein Chem* **50**: 1-59.
- Janciauskiene, S., Dominaitiene, R., Sternby, N.H., Piitulainen, E. and Eriksson, S., 2002. Detection of circulating and endothelial cell polymers of Z and wild type alpha 1-antitrypsin by a monoclonal antibody. *J Biol Chem* **277**: 26540-26546.
- Janowski, R., Kozak, M., Abrahamson, M., Grubb, A. and Jaskolski, M., 2005. 3D domain-swapped human cystatin C with amyloidlike intermolecular beta-sheets. *Proteins* **61**: 570-578.
- Jaswal, S.S., Sohl, J.L., Davis, J.H. and Agard, D.A., 2002. Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* **415**: 343-346.

- Johnson, S.M., Wiseman, R.L., Sekijima, Y., Green, N.S., Adamski-Werner, S.L. and Kelly, J.W., 2005. Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: A focus on the transthyretin amyloidoses. *Acc Chem Res* **38**: 911-921.
- Johnson, W.C., 1999. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**: 307-312.
- Jung, J., Lee, J. and Moon, H.T., 2005. Topological determinants of protein unfolding rates. *Proteins* **58**: 389-395.
- Kaiser, D., Manoil, C. and Dworkin, M., 1979. Myxobacteria: Cell interactions, genetics, and development. *Annu Rev Microbiol* **33**: 595-639.
- Karpusas, M., Baase, W.A., Matsumura, M. and Matthews, B.W., 1989. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc Natl Acad Sci U S A* **86**: 8237-8241.
- Khorasanizadeh, S., Peters, I.D., Butt, T.R. and Roder, H., 1993. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**: 7054-7063.
- Kim, Y.H., Kapfer, D.M., Boekhorst, J., Lubsen, N.H., Bachinger, H.P., Shearer, T.R., David, L.L., Feix, J.B. and Lampi, K.J., 2002. Deamidation, but not truncation, decreases the urea stability of a lens structural protein, betaB1-crystallin. *Biochemistry* **41**: 14076-14084.
- Kosinski-Collins, M.S., and King, J., 2003. In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation. *Protein Sci* **12**: 480-490.
- Kosinski-Collins, M.S., Flaugh, S.L. and King, J., 2004. Probing folding and fluorescence quenching in human gammaD crystallin greek key domains using triple tryptophan mutant proteins. *Protein Sci* **13**: 2223-2235.
- Kretschmar, M., and Jaenicke, R., 1999. Stability of a homo-dimeric Ca^{2+} -binding member of the beta gamma-crystallin superfamily: DSC measurements on spherulin 3a from physarum polycephalum. *J Mol Biol* **291**: 1147-1153.
- Kretschmar, M., Mayr, E.M. and Jaenicke, R., 1999a. Kinetic and thermodynamic stabilization of the betagamma-crystallin homolog spherulin 3a from physarum polycephalum by calcium binding. *J Mol Biol* **289**: 701-705.
- Kretschmar, M., Mayr, E.M. and Jaenicke, R., 1999b. Homo-dimeric spherulin 3a: A single-domain member of the beta gamma-crystallin superfamily. *Biol Chem* **380**: 89-94.

- Lai, Z., McCulloch, J., Lashuel, H.A. and Kelly, J.W., 1997. Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: Equilibria with high kinetic barriers. *Biochemistry* **36**: 10230-10239.
- Lampi, K.J., Amyx, K.K., Ahmann, P. and Steel, E.A., 2006. Deamidation in human lens betaB2-crystallin destabilizes the dimer. *Biochemistry* **45**: 3146-3153.
- Lampi, K.J., Shih, M., Ueda, Y., Shearer, T.R. and David, L.L., 2002. Lens proteomics: Analysis of rat crystallin sequences and two-dimensional electrophoresis map. *Invest Ophthalmol Vis Sci* **43**: 216-224.
- Lampi, K.J., Oxford, J.T., Bachinger, H.P., Shearer, T.R., David, L.L. and Kapfer, D.M., 2001. Deamidation of human beta B1 alters the elongated structure of the dimer. *Exp Eye Res* **72**: 279-288.
- Lampi, K.J., Ma, Z., Hanson, S.R., Azuma, M., Shih, M., Shearer, T.R., Smith, D.L., Smith, J.B. and David, L.L., 1998. Age-related changes in human lens crystallins identified by two-dimensional electrophoresis and mass spectrometry. *Exp Eye Res* **67**: 31-43.
- Land, M.F., 2005. The optical structures of animal eyes. *Curr Biol* **15**: R319-23.
- Lapatto, R., Nalini, V., Bax, B., Driessen, H., Lindley, P.F., Blundell, T.L. and Slingsby, C., 1991. High resolution structure of an oligomeric eye lens beta-crystallin. loops, arches, linkers and interfaces in beta B2 dimer compared to a monomeric gamma-crystallin. *J Mol Biol* **222**: 1067-1083.
- Lapko, V.N., Cerny, R.L., Smith, D.L. and Smith, J.B., 2005. Modifications of human betaA1/betaA3-crystallins include S-methylation, glutathiolation, and truncation. *Protein Sci* **14**: 45-54.
- Lawson, C.L., Benoff, B., Berger, T., Berman, H.M. and Carey, J., 2004. E. coli trp repressor forms a domain-swapped array in aqueous alcohol. *Structure* **12**: 1099-1108.
- Le, A., Ferrell, G.A., Dishon, D.S., Le, Q.Q. and Sifers, R.N., 1992. Soluble aggregates of the human PiZ alpha 1-antitrypsin variant are degraded within the endoplasmic reticulum by a mechanism sensitive to inhibitors of protein synthesis. *J Biol Chem* **267**: 1072-1080.
- Lee, S., and Eisenberg, D., 2003. Seeded conversion of recombinant prion protein to a disulfide-bonded oligomer by a reduction-oxidation process. *Nat Struct Biol* **10**: 725-730.
- Lehrman, S.R., Tuls, J.L., Havel, H.A., Haskell, R.J., Putnam, S.D. and Tomich, C.S., 1991. Site-directed mutagenesis to probe protein folding: Evidence that the formation and aggregation of a bovine growth hormone folding intermediate are dissociable processes. *Biochemistry* **30**: 5777-5784.

- Li, N., Yang, Y., Bu, J., Zhao, C., Lu, S., Zhao, J., Yan, L., Cui, L., Zheng, R., Li, J., Tang, J. and Zhao, K., 2006. An autosomal dominant progressive congenital zonular nuclear cataract linked to chromosome 20p12.2-p11.23. *Mol Vis* **12**: 1506-1510.
- Lim, W.A., and Sauer, R.T., 1991. The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol* **219**: 359-376.
- Liu, B.F., and Liang, J.J., 2006. Domain interaction sites of human lens betaB2-crystallin. *J Biol Chem* **281**: 2624-2630.
- Liu, Y., and Eisenberg, D., 2002. 3D domain swapping: As domains continue to swap. *Protein Sci* **11**: 1285-1299.
- Lomas, D.A., and Carrell, R.W., 2002. Serpinopathies and the conformational dementias. *Nat Rev Genet* **3**: 759-768.
- Lomas, D.A., Evans, D.L., Finch, J.T. and Carrell, R.W., 1992. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* **357**: 605-607.
- Lovicu, F., and Robinson, M., 2004. *Development of the ocular lens*. Cambridge University Press, Cambridge, UK.
- Lovicu, F.J., and McAvoy, J.W., 2005. Growth factor regulation of lens development. *Dev Biol* **280**: 1-14.
- Lubsen, N.H., Aarts, H.J. and Schoenmakers, J.G., 1988. The evolution of lenticular proteins: The beta- and gamma-crystallin super gene family. *Prog Biophys Mol Biol* **51**: 47-76.
- Lynch, S.M., Boswell, S.A. and Colon, W., 2004. Kinetic stability of Cu/Zn superoxide dismutase is dependent on its metal ligands: Implications for ALS. *Biochemistry* **43**: 16525-16531.
- Ma, Z., Hanson, S.R., Lampi, K.J., David, L.L., Smith, D.L. and Smith, J.B., 1998. Age-related changes in human lens crystallins identified by HPLC and mass spectrometry. *Exp Eye Res* **67**: 21-30.
- MacDonald, J.T., Purkiss, A.G., Smith, M.A., Evans, P., Goodfellow, J.M. and Slingsby, C., 2005. Unfolding crystallins: The destabilizing role of a beta-hairpin cysteine in betaB2-crystallin by simulation and experiment. *Protein Sci* **14**: 1282-1292.
- Machius, M., Declerck, N., Huber, R. and Wiegand, G., 2003. Kinetic stabilization of bacillus licheniformis alpha-amylase through introduction of hydrophobic residues at the surface. *J Biol Chem* **278**: 11546-11553.
- MacRae, T.H., 2000. Structure and function of small heat shock/alpha-crystallin proteins: Established concepts and emerging ideas. *Cell Mol Life Sci* **57**: 899-913.

- Manning, M., and Colon, W., 2004. Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* **43**: 11248-11254.
- Mansfeld, J., Vriend, G., Dijkstra, B.W., Veltman, O.R., Van den Burg, B., Venema, G., Ulbrich-Hofmann, R. and Eijsink, V.G., 1997. Extreme stabilization of a thermolysin-like protease by an engineered disulfide bond. *J Biol Chem* **272**: 11152-11156.
- Matouschek, A., Kellis, J.T., Jr, Serrano, L. and Fersht, A.R., 1989. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**: 122-126.
- Matsumura, M., Wozniak, J.A., Sun, D.P. and Matthews, B.W., 1989. Structural studies of mutants of T4 lysozyme that alter hydrophobic stabilization. *J Biol Chem* **264**: 16059-16066.
- Matthews, B.W., 1993. Structural and genetic analysis of protein stability. *Annu Rev Biochem* **62**: 139-160.
- Matthews, B.W., 1987. Genetic and structural analysis of the protein stability problem. *Biochemistry* **26**: 6885-6888.
- Mayr, E.M., Jaenicke, R. and Glockshuber, R., 1997. The domains in gammaB-crystallin: Identical fold-different stabilities. *J Mol Biol* **269**: 260-269.
- Mayr, E.M., Jaenicke, R. and Glockshuber, R., 1994. Domain interactions and connecting peptides in lens crystallins. *J Mol Biol* **235**: 84-88.
- McCaldon, P., and Argos, P., 1988. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**: 99-122.
- McCallister, E.L., Alm, E. and Baker, D., 2000. Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* **7**: 669-673.
- Minton, A.P., 2000. Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* **10**: 34-39.
- Mirny, L.A., Abkevich, V.I. and Shakhnovich, E.I., 1998. How evolution makes proteins fold quickly. *Proc Natl Acad Sci U S A* **95**: 4976-4981.
- Mitraki, A., and King, J., 1992. Amino acid substitutions influencing intracellular protein folding pathways. *FEBS Lett* **307**: 20-25.
- Mornon, J.P., Halaby, D., Malfois, M., Durand, P., Callebaut, I. and Tardieu, A., 1998. Alpha-crystallin C-terminal domain: On the track of an ig fold. *Int J Biol Macromol* **22**: 219-227.

- Negin, R.S., and Carbeck, J.D., 2002. Measurement of electrostatic interactions in protein folding with the use of protein charge ladders. *J Am Chem Soc* **124**: 2911-2916.
- Nilsson, D.E., 2004. Eye evolution: A question of genetic promiscuity. *Curr Opin Neurobiol* **14**: 407-414.
- Nordberg Karlsson, E., Crennell, S.J., Higgins, C., Nawaz, S., Yeoh, L., Hough, D.W. and Danson, M.J., 2003. Citrate synthase from thermus aquaticus: A thermostable bacterial enzyme with a five-membered inter-subunit ionic network. *Extremophiles* **7**: 9-16.
- Norledge, B.V., Trinkl, S., Jaenicke, R. and Slingsby, C., 1997. The X-ray structure of a mutant eye lens beta B2-crystallin with truncated sequence extensions. *Protein Sci* **6**: 1612-1620.
- Oyster, C., 1999. The lens and the vitreous, pp. 492-510 in *The Human Eye: Structure and Function*, edited by C. Oyster. Sinauer Associates, Inc., Sunderland, MA.
- Palme, S., Jaenicke, R. and Slingsby, C., 1998a. Unusual domain pairing in a mutant of bovine lens gammaB-crystallin. *J Mol Biol* **279**: 1053-1059.
- Palme, S., Jaenicke, R. and Slingsby, C., 1998b. X-ray structures of three interface mutants of gammaB-crystallin from bovine eye lens. *Protein Sci* **7**: 611-618.
- Palme, S., Slingsby, C. and Jaenicke, R., 1997. Mutational analysis of hydrophobic domain interactions in gamma B-crystallin from bovine eye lens. *Protein Sci* **6**: 1529-1536.
- Pande, A., Annunziata, O., Asherie, N., Ogun, O., Benedek, G.B. and Pande, J., 2005. Decrease in protein solubility and cataract formation caused by the Pro23 to thr mutation in human gamma D-crystallin. *Biochemistry* **44**: 2491-2500.
- Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J. and Benedek, G.B., 2001. Crystal cataracts: Human genetic cataract caused by protein crystallization. *Proc Natl Acad Sci U S A* **98**: 6116-6120.
- Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J.A., Lubsen, N.H., Walton, D. and Benedek, G.B., 2000. Molecular basis of a progressive juvenile-onset hereditary cataract. *Proc Natl Acad Sci U S A* **97**: 1993-1998.
- Park, S.H., O'Neil, K.T. and Roder, H., 1997. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry* **36**: 14277-14283.
- Peek, R., McAvoy, J.W., Lubsen, N.H. and Schoenmakers, J.G., 1992a. Rise and fall of crystallin gene messenger levels during fibroblast growth factor induced terminal differentiation of lens cells. *Dev Biol* **152**: 152-160.

- Peek, R., Kraft, H.J., Klok, E.J., Lubsen, N.H. and Schoenmakers, J.G., 1992b. Activation and repression sequences determine the lens-specific expression of the rat gamma D-crystallin gene. *Nucleic Acids Res* **20**: 4865-4871.
- Petty, S.A., and Decatur, S.M., 2005. Intersheet rearrangement of polypeptides during nucleation of {beta}-sheet aggregates. *Proc Natl Acad Sci U S A* **102**: 14272-14277.
- Petty, S.A., Adalsteinsson, T. and Decatur, S.M., 2005. Correlations among morphology, beta-sheet stability, and molecular structure in prion peptide aggregates. *Biochemistry* **44**: 4720-4726.
- Piatigorsky, J., 2003. Crystallin genes: Specialization by changes in gene regulation may precede gene duplication. *J Struct Funct Genomics* **3**: 131-137.
- Piatigorsky, J., and Wistow, G.J., 1989. Enzyme/crystallins: Gene sharing as an evolutionary strategy. *Cell* **57**: 197-199.
- Piatigorsky, J., O'Brien, W.E., Norman, B.L., Kalumuck, K., Wistow, G.J., Borrás, T., Nickerson, J.M. and Wawrousek, E.F., 1988. Gene sharing by delta-crystallin and argininosuccinate lyase. *Proc Natl Acad Sci U S A* **85**: 3479-3483.
- Plaxco, K.W., Simons, K.T. and Baker, D., 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**: 985-994.
- Ponce, A., Sorensen, C. and Takemoto, L., 2006. Role of short-range protein interactions in lens opacifications. *Mol Vis* **12**: 879-884.
- Pozdnyakova, I., Guidry, J. and Wittung-Stafshede, P., 2001. Copper stabilizes azurin by decreasing the unfolding rate. *Arch Biochem Biophys* **390**: 146-148.
- Ptitsyn, O.B., Pain, R.H., Semisotnov, G.V., Zerovnik, E. and Razgulyaev, O.I., 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett* **262**: 20-24.
- Purkiss, A.G., Bateman, O.A., Goodfellow, J.M., Lubsen, N.H. and Slingsby, C., 2002. The X-ray crystal structure of human gamma S-crystallin C-terminal domain. *J Biol Chem* **277**: 4199-4205.
- Quiring, R., Walldorf, U., Kloter, U. and Gehring, W.J., 1994. Homology of the eyeless gene of drosophila to the small eye gene in mice and aniridia in humans. *Science* **265**: 785-789.
- Rajini, B., Graham, C., Wistow, G. and Sharma, Y., 2003. Stability, homodimerization, and calcium-binding properties of a single, variant betagamma-crystallin domain of the protein absent in melanoma 1 (AIM1). *Biochemistry* **42**: 4552-4559.

- Raman, B., and Rao, C.M., 1997. Chaperone-like activity and temperature-induced structural changes of alpha-crystallin. *J Biol Chem* **272**: 23559-23564.
- Ren, Z., Li, A., Shastry, B.S., Padma, T., Ayyagari, R., Scott, M.H., Parks, M.M., Kaiser-Kupfer, M.I. and Hejtmancik, J.F., 2000. A 5-base insertion in the gammaC-crystallin gene is associated with autosomal dominant variable zonular pulverulent cataract. *Hum Genet* **106**: 531-537.
- Richards, F.M., 1977. Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* **6**: 151-176.
- Robinson, N.E., Lampi, K.J., Speir, J.P., Kruppa, G., Easterling, M. and Robinson, A.B., 2006. Quantitative measurement of young human eye lens crystallins by direct injection fourier transform ion cyclotron resonance mass spectrometry. *Mol Vis* **12**: 704-711.
- Rose, G.D., and Wolfenden, R., 1993. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu Rev Biophys Biomol Struct* **22**: 381-415.
- Rosinke, B., Renner, C., Mayr, E.M., Jaenicke, R. and Holak, T.A., 1997. Ca²⁺-loaded spherulin 3a from physarum polycephalum adopts the prototype gamma-crystallin fold in aqueous solution. *J Mol Biol* **271**: 645-655.
- Rudolph, R., Siebendritt, R., Nessler, G., Sharma, A.K. and Jaenicke, R., 1990. Folding of an all-beta protein: Independent domain folding in gamma II-crystallin from calf eye lens. *Proc Natl Acad Sci U S A* **87**: 4625-4629.
- Salim, A., and Zaidi, Z.H., 2003. Homology models of human gamma-crystallins: Structural study of the extensive charge network in gamma-crystallins. *Biochem Biophys Res Commun* **300**: 624-630.
- Santhiya, S.T., Shyam Manohar, M., Rawlley, D., Vijayalakshmi, P., Namperumalsamy, P., Gopinath, P.M., Loster, J. and Graw, J., 2002. Novel mutations in the gamma-crystallin genes cause autosomal dominant congenital cataracts. *J Med Genet* **39**: 352-358.
- Santini, S.A., Mordente, A., Meucci, E., Miggiano, G.A. and Martorana, G.E., 1992. Conformational stability of bovine alpha-crystallin. evidence for a destabilizing effect of ascorbate. *Biochem J* **287** (Pt 1): 107-112.
- Sauer, R.T., Milla, M.E., Waldburger, C.D., Brown, B.M. and Schildbach, J.F., 1996. Sequence determinants of folding and stability for the P22 arc repressor dimer. *FASEB J* **10**: 42-48.
- Sawaya, M.R., Guo, S., Tabor, S., Richardson, C.C. and Ellenberger, T., 1999. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**: 167-177.

- Scalley, M.L., Yi, Q., Gu, H., McCormack, A., Yates, J.R., 3rd and Baker, D., 1997. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* **36**: 3373-3382.
- Schellman, J.A., 1987. Selective binding and solvent denaturation. *Biopolymers* **26**: 549-559.
- Schlunegger, M.P., Bennett, M.J. and Eisenberg, D., 1997. Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly. *Adv Protein Chem* **50**: 61-122.
- Searle, B.C., Dasari, S., Wilmarth, P.A., Turner, M., Reddy, A.P., David, L.L. and Nagalla, S.R., 2005. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J Proteome Res* **4**: 546-554.
- Searle, M.S., and Ciani, B., 2004. Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding. *Curr Opin Struct Biol* **14**: 458-464.
- Sen, A.C., Walsh, M.T. and Chakrabarti, B., 1992. An insight into domain structures and thermal stability of gamma-crystallins. *J Biol Chem* **267**: 11898-11907.
- Shen, W., and Mardon, G., 1997. Ectopic eye development in drosophila induced by directed dachshund expression. *Development* **124**: 45-52.
- Shimeld, S.M., Purkiss, A.G., Dirks, R.P., Bateman, O.A., Slingsby, C. and Lubsen, N.H., 2005. Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol* **15**: 1684-1689.
- Siezen, R.J., Wu, E., Kaplan, E.D., Thomson, J.A. and Benedek, G.B., 1988. Rat lens gamma-crystallins. characterization of the six gene products and their spatial and temporal distribution resulting from differential synthesis. *J Mol Biol* **199**: 475-490.
- Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J. and Whisstock, J.C., 2001. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem* **276**: 33293-33296.
- Slingsby, C., and Clout, N.J., 1999. Structure of the crystallins. *Eye* **13 (Pt 3b)**: 395-402.
- Smith, R.S., Hawes, N.L., Chang, B., Roderick, T.H., Akeson, E.C., Heckenlively, J.R., Gong, X., Wang, X. and Davisson, M.T., 2000. Lop12, a mutation in mouse crygd causing lens opacity similar to human coppock cataract. *Genomics* **63**: 314-320.
- Sohl, J.L., Jaswal, S.S. and Agard, D.A., 1998. Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **395**: 817-819.

- Solis-Mendiola, S., Gutierrez-Gonzalez, L.H., Arroyo-Reyna, A., Padilla-Zuniga, J., Rojo-Dominguez, A. and Hernandez-Arana, A., 1998. pH dependence of the activation parameters for chymopapain unfolding: Influence of ion pairs on the kinetic stability of proteins. *Biochim Biophys Acta* **1388**: 363-372.
- Sreerama, N., and Woody, R.W., 2003. Structural composition of betaI- and betaII-proteins. *Protein Sci* **12**: 384-388.
- Sreerama, N., and Woody, R.W., 2000. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**: 252-260.
- Sreerama, N., Venyaminov, S.Y. and Woody, R.W., 2000. Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **287**: 243-251.
- Stephan, D.A., Gillanders, E., Vanderveen, D., Freas-Lutz, D., Wistow, G., Baxevanis, A.D., Robbins, C.M., VanAuken, A., Quesenberry, M.I., Bailey-Wilson, J., Juo, S.H., Trent, J.M., Smith, L. and Brownstein, M.J., 1999. Progressive juvenile-onset punctate cataracts caused by mutation of the gammaD-crystallin gene. *Proc Natl Acad Sci U S A* **96**: 1008-1012.
- Story, R.M., Weber, I.T. and Steitz, T.A., 1992. The structure of the E. coli recA protein monomer and polymer. *Nature* **355**: 318-325.
- Strop, P., Smith, K.S., Iverson, T.M., Ferry, J.G. and Rees, D.C., 2001. Crystal structure of the "cab"-type beta class carbonic anhydrase from the archaeon methanobacterium thermoautotrophicum. *J Biol Chem* **276**: 10299-10305.
- Sun, D.P., Sauer, U., Nicholson, H. and Matthews, B.W., 1991a. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30**: 7142-7153.
- Sun, D.P., Soderlind, E., Baase, W.A., Wozniak, J.A., Sauer, U. and Matthews, B.W., 1991b. Cumulative site-directed charge-change replacements in bacteriophage T4 lysozyme suggest that long-range electrostatic interactions contribute little to protein stability. *J Mol Biol* **221**: 873-887.
- Sun, T.X., Akhtar, N.J. and Liang, J.J., 1999. Thermodynamic stability of human lens recombinant alphaA- and alphaB-crystallins. *J Biol Chem* **274**: 34067-34071.
- Surewicz, W.K., and Olesen, P.R., 1995. On the thermal stability of alpha-crystallin: A new insight from infrared spectroscopy. *Biochemistry* **34**: 9655-9660.
- Tardieu, A., 1988. Eye lens proteins and transparency: From light transmission theory to solution X-ray structural analysis. *Annu Rev Biophys Biophys Chem* **17**: 47-70.

- Tardieu, A., Veretout, F., Krop, B. and Slingsby, C., 1992. Protein interactions in the calf eye lens: Interactions between beta-crystallins are repulsive whereas in gamma-crystallins they are attractive. *Eur Biophys J* **21**: 1-12.
- Thomas, G., Zelenka, P.S., Cuthbertson, R.A., Norman, B.L. and Piatigorsky, J., 1990. Differential expression of the two delta-crystallin/argininosuccinate lyase genes in lens, heart, and brain of chicken embryos. *New Biol* **2**: 903-914.
- Tootle, T.L., Silver, S.J., Davies, E.L., Newman, V., Latek, R.R., Mills, I.A., Selengut, J.D., Parlikar, B.E. and Rebay, I., 2003. The transcription factor eyes absent is a protein tyrosine phosphatase. *Nature* **426**: 299-302.
- Trinkl, S., Glockshuber, R. and Jaenicke, R., 1994. Dimerization of beta B2-crystallin: The role of the linker peptide and the N- and C-terminal extensions. *Protein Sci* **3**: 1392-1400.
- Tycko, R., 2006. Molecular structure of amyloid fibrils: Insights from solid-state NMR. *Q Rev Biophys* **39**: 1-55.
- Ueda, Y., Duncan, M.K. and David, L.L., 2002. Lens proteomics: The accumulation of crystallin modifications in the mouse lens with age. *Invest Ophthalmol Vis Sci* **43**: 205-215.
- Van Montfort, R., Slingsby, C. and Vierling, E., 2001. Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv Protein Chem* **59**: 105-156.
- Van Montfort, R.L., Bateman, O.A., Lubsen, N.H. and Slingsby, C., 2003. Crystal structure of truncated human betaB1-crystallin. *Protein Sci* **12**: 2606-2612.
- Van Montfort, R.L., Basha, E., Friedrich, K.L., Slingsby, C. and Vierling, E., 2001. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol* **8**: 1025-1030.
- Van Rens, G.L., De Jong, W.W. and Bloemendal, H., 1991. One member of the gamma-crystallin gene family, gamma s, is expressed in birds. *Exp Eye Res* **53**: 135-138.
- Vanita, V., Hennies, H.C., Singh, D., Nurnberg, P., Sperling, K. and Singh, J.R., 2006. A novel mutation in GJA8 associated with autosomal dominant congenital cataract in a family of indian origin. *Mol Vis* **12**: 1217-1222.
- Varley, P., Gronenborn, A.M., Christensen, H., Wingfield, P.T., Pain, R.H. and Clore, G.M., 1993. Kinetics of folding of the all-beta sheet protein interleukin-1 beta. *Science* **260**: 1110-1113.
- Viteri, G., Carrard, G., Birlouez-Aragon, I., Silva, E. and Friguet, B., 2004. Age-dependent protein modifications and declining proteasome activity in the human lens. *Arch Biochem Biophys* **427**: 197-203.

- Voorter, C.E., Salemink, I. and De Jong, W.W., 1993. Delta-crystallin is more thermostable than mammalian argininosuccinate lyase. *Exp Eye Res* **56**: 733-735.
- Vos, J.J., and van Norren, D., 2004. Thermal cataract, from furnaces to lasers. *Clin Exp Optom* **87**: 372-376.
- Vos, J.J., and van Norren, D., 1998. On the relevant spectral parts for glassblowers' cataract. *Ophthalmic Physiol Opt* **18**: 311-313.
- Wenk, M., and Jaenicke, R., 1999. Calorimetric analysis of the ca(2+)-binding betagamma-crystallin homolog protein S from myxococcus xanthus: Intrinsic stability and mutual stabilization of domains. *J Mol Biol* **293**: 117-124.
- Wenk, M., Jaenicke, R. and Mayr, E.M., 1998. Kinetic stabilisation of a modular protein by domain interactions. *FEBS Lett* **438**: 127-130.
- Wenk, M., Herbst, R., Hoeger, D., Kretschmar, M., Lubsen, N.H. and Jaenicke, R., 2000. Gamma S-crystallin of bovine and human eye lens: Solution structure, stability and folding of the intact two-domain protein and its separate domains. *Biophys Chem* **86**: 95-108.
- Wenk, M., Baumgartner, R., Holak, T.A., Huber, R., Jaenicke, R. and Mayr, E.M., 1999. The domains of protein S from myxococcus xanthus: Structure, stability and interactions. *J Mol Biol* **286**: 1533-1545.
- Went, H.M., and Jackson, S.E., 2005. Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**: 229-237.
- Werten, P.J., Lindner, R.A., Carver, J.A. and de Jong, W.W., 1999. Formation of betaA3/betaB2-crystallin mixed complexes: Involvement of N- and C-terminal extensions. *Biochim Biophys Acta* **1432**: 286-292.
- Wetzel, R., 1997. Domain stability in immunoglobulin light chain deposition disorders. *Adv Protein Chem* **50**: 183-242.
- Wieligmann, K., Mayr, E.M. and Jaenicke, R., 1999. Folding and self-assembly of the domains of betaB2-crystallin from rat eye lens. *J Mol Biol* **286**: 989-994.
- Wilmarth, P.A., Tanner, S., Dasari, S., Nagalla, S.R., Riviere, M.A., Bafna, V., Pevzner, P.A. and David, L.L., 2006. Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: Does deamidation contribute to crystallin insolubility? *J Proteome Res* **5**: 2554-2566.
- Wistow, G., 1990. Evolution of a protein superfamily: Relationships between vertebrate lens crystallins and microorganism dormancy proteins. *J Mol Evol* **30**: 140-145.

- Wistow, G., Summers, L. and Blundell, T., 1985. Myxococcus xanthus spore coat protein S may have a similar structure to vertebrate lens beta gamma-crystallins. *Nature* **315**: 771-773.
- Wistow, G., Sardarian, L., Gan, W. and Wyatt, M.K., 2000. The human gene for gammaS-crystallin: Alternative transcripts and expressed sequences from the first intron. *Mol Vis* **6**: 79-84.
- Wistow, G., Bernstein, S.L., Wyatt, M.K., Behal, A., Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K., 2002. Expressed sequence tag analysis of adult human lens for the NEIBank project: Over 2000 non-redundant transcripts, novel genes and splice variants. *Mol Vis* **8**: 171-184.
- Wistow, G., Wyatt, K., David, L., Gao, C., Bateman, O., Bernstein, S., Tomarev, S., Segovia, L., Slingsby, C. and Vihtelic, T., 2005. gammaN-crystallin and the evolution of the betagamma-crystallin superfamily in vertebrates. *FEBS J* **272**: 2276-2291.
- Wu, Z., Delaglio, F., Wyatt, K., Wistow, G. and Bax, A., 2005. Solution structure of (gamma)S-crystallin by molecular fragment replacement NMR. *Protein Sci* **14**: 3101-3114.
- Wunderlich, M., Martin, A. and Schmid, F.X., 2005. Stabilization of the cold shock protein CspB from bacillus subtilis by evolutionary optimization of coulombic interactions. *J Mol Biol* **347**: 1063-1076.
- Xing, X., and Bell, C.E., 2004. Crystal structures of escherichia coli RecA in a compressed helical filament. *J Mol Biol* **342**: 1471-1485.
- Zarina, S., Slingsby, C., Jaenicke, R., Zaidi, Z.H., Driessen, H. and Srinivasan, N., 1994. Three-dimensional model and quaternary structure of the human eye lens protein gamma S-crystallin based on beta- and gamma-crystallin X-ray coordinates and ultracentrifugation. *Protein Sci* **3**: 1840-1846.
- Zerovnik, E., 2002. Amyloid-fibril formation. proposed mechanisms and relevance to conformational disease. *Eur J Biochem* **269**: 3362-3371.
- Zhang, C., and Kim, S.H., 2000. A comprehensive analysis of the greek key motifs in protein beta-barrels and beta-sandwiches. *Proteins* **40**: 409-419.

CHAPTER SEVEN

APPENDICES

APPENDIX A: SUMMARY OF PREVIOUS WORK

Previously, I was a graduate student in Ilaria Rebay's laboratory studying the retinal determination network. Here, I will give a brief description of the various projects I addressed in her lab.

The retinal determination (RD) genes twin of eyeless (*toy*), eyeless (*ey*), eyes absent (*eya*), sine oculis (*so*), and dachshund (*dac*) encode transcription factors and cofactors essential for eye specification (Quiring et al. 1994). Hypomorphic alleles of these genes result in complete loss or severe defects of the eye. In addition, ectopic expression of *toy*, *ey*, *eya*, or *dac* results in the induction of ectopic eye formation (Halder et al. 1995; Shen and Mardon 1997). While overexpression of *so* cannot act alone to induce ectopic eyes, it can act synergistically with *eya* and *dac* to produce ectopic eyes. Null mutations of these genes are homozygous lethal, indicating essential additional roles in embryogenesis as well as their roles in eye development.

The overall question of one project addressed was the role of MAPK phosphorylation in regulating the activity of the retinal determination network. The question of whether MAPK phosphorylation regulates the RD gene network was analyzed through a biochemical and genetic approach. Biochemically, kinase assays were performed *in vitro* to test the ability of these MAPKs to phosphorylate *So*, *Dac*, and *Ey* protein substrates. The results from this experiment indicated that *So* was phosphorylated by JunK, ERK, and p38a MAPKs *in vitro*. It also suggests that *Ey* may be phosphorylated by ERK MAPK, *in vitro*. This biochemical evidence was followed by genetic studies to confirm the relevance of the phosphorylation of these proteins *in vivo*.

In addition to my first project, I contributed to a joint project in the lab which identified *Eya* as a phosphatase in addition to a transcription factor (Tootle et al. 2003). *Eya* was first identified as a potential phosphatase by sequence homology and structural modeling of the *Eya* domain to the Haloacid Dehalogenase (HAD) motif. HAD family members include magnesium dependent phosphatases that utilize a nucleophilic

aspartate. The Eya domain contains a catalytic quintet of HAD active site residues that is conserved in Eya homologs of plants, invertebrates, and vertebrates.

My contribution to this work was the preparation of the Drosophila and Mouse Eyes Absent protein by determining the conditions necessary to purify multiple deletions and fragment forms of soluble GST tagged drosophila Eya and mouse Eya fusion proteins and by optimizing protein purification of both EYA homologues. I also determined the conditions necessary for the phosphatase activity assays and optimized the conditions preferred by the phosphatase enzyme.

My second project in the lab involved determining substrates for the Eya phosphatase and transcription factor. I approached this question utilizing the *in vitro* expression cloning (IVEC) technique. The IVEC technique involves transcription and translation of cDNA pools into S³⁵ labeled proteins *in vitro*. The cDNA pools were ESTs from release 1 of the Drosophila genome project. The technique was modified as a binding assay screen by using tagged proteins as the “bait” and performing GST pull down experiments to determine protein-protein interactions. The outcome of the IVEC screen could yield Eya substrates, Eya cofactors, and enzymes that post-translationally modify Eya (i.e. kinases). I performed the initial screen and the substrates from the initial screen are now being confirmed and identified.

APPENDIX B: PROTEIN PARAMETERS

Extinction Coefficients were calculated using the ProtParam tool

(<http://us.expasy.org/tools/protparam.html>)

<u>Proteins</u>	<u># of AminoAcids</u>	<u>Extinction Coefficient</u>	<u>Molecular Weight</u>
$\gamma_{D_{WT}}$	182	41,040 $\text{cm}^{-1}\text{M}^{-1}$	21,817.2 Da
γ_{D_N}	92	20,580 $\text{cm}^{-1}\text{M}^{-1}$	10,972.1 Da
γ_{D_C}	97	20,460 $\text{cm}^{-1}\text{M}^{-1}$	11,833.2 Da
$\gamma_{S_{WT}}$	187	41,040 $\text{cm}^{-1}\text{M}^{-1}$	22,156.9 Da
γ_{S_N}	99	23,170 $\text{cm}^{-1}\text{M}^{-1}$	11,406.7 Da
γ_{S_C}	95	19,180 $\text{cm}^{-1}\text{M}^{-1}$	11,510.1 Da
<u>Chimeras</u>			
$\gamma_{Dn-\gamma_{Sc}}$	182	39,790 $\text{cm}^{-1}\text{M}^{-1}$	21,559.2 Da
$\gamma_{Dn-\gamma_{Sc}}$ [γ_{Dint}]	-	41,745 $\text{cm}^{-1}\text{M}^{-1}$	21,547.3 Da
$\gamma_{Sn-\gamma_{Dc}}$	188	42,320 $\text{cm}^{-1}\text{M}^{-1}$	22,486 Da
$\gamma_{Sn-\gamma_{Dc}}$ [γ_{Dint}]	-	44,725 $\text{cm}^{-1}\text{M}^{-1}$	22,591.1 Da

APPENDIX C: EQUILIBRIUM DATA FITTING EQUATIONS

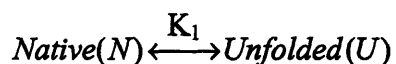
1. Calculating Guanidine Hydrochloride Concentrations

The concentration of GuHCl was determined by measuring the refractive indexes of samples and applying the following equation:

$$[\text{GuHCl}] = 57.147 \cdot \Delta N + 38.68 \cdot \Delta N^2 - 91.60 \cdot \Delta N^3$$

ΔN is the refractive index of the sample minus the refractive index of water.

2. Two-state Equilibrium Unfolding/Refolding



$$K_1 = [U]/[N]$$

K_1 is the equilibrium constant for the reaction.

The following equation was used to describe an equilibrium reaction with no intermediates:

$$Y = (Y_N^\circ + S_N \cdot [\text{den}]) \cdot (1 / (1 + \exp((m \cdot [\text{den}] - \Delta G_1^\circ) / RT))) + (Y_U^\circ + S_U \cdot [\text{den}]) \cdot (1 / (1 / \exp((m \cdot [\text{den}] - \Delta G_1^\circ) / RT) + 1))$$

Where Y is the spectroscopic signal of a mixture of native and unfolded protein. Y_N° and Y_U° are the signals of the native and unfolded proteins in the absence of denaturant, respectively and S_N and S_U are the slopes of the native and unfolded baselines, respectively. ΔG_1° and m value are the equilibrium values of the transition.

$$\Delta G_1^\circ = -RT \cdot \ln(K_1)$$

Data were fit to the two-state equilibrium model using the curve-fitting feature of KaledaGraph. The following algorithm was used for experiments performed at 37°C (310 K) or 20°C (293 K). The gas constant is in units of cal*mol⁻¹*K⁻¹.

```
x=m0;
a=m1;
b=m2;
c=m3;
d=m4;
e=m5;
f=m6;
;
K1() = exp((c*x-d)/(1.987*310))
;
twost(a0, b0, c0, d0, e0, f0) = (a+b*x)*(1/(1+K1))+(e+f*x)*(1/(1/K1+1));
a=a0\; b=b0\; c=c0\; d=d0\; e=e0\; f=f0\;
```

3. Three-state Equilibrium Unfolding/Refolding



K_1 and K_2 are the equilibrium constants for the native to intermediate and intermediate to unfolded transitions, respectively. They are related to the concentration of native, intermediate and unfolded protein in the following manner:

$$K_1 = [I]/[N]$$

$$K_2 = [U]/[I]$$

The following equation was used to describe an equilibrium reaction with one intermediate:

$$Y = (Y_N^\circ + S_N * [\text{den}])(1 / (1 + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) * \exp((m_2 * [\text{den}] - \Delta G_2) / RT))) + Y_I * (\exp((m_1 * [\text{den}] - \Delta G_1) / RT) / (1 + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) * \exp((m_2 * [\text{den}] - \Delta G_2) / RT))) + (Y_U^\circ + S_U * [\text{den}])(\exp((m_1 * [\text{den}] - \Delta G_1) / RT) * \exp((m_2 * [\text{den}] - \Delta G_2) / RT) / (1 + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) + \exp((m_1 * [\text{den}] - \Delta G_1) / RT) * \exp((m_2 * [\text{den}] - \Delta G_2) / RT)))$$

Y_N° and Y_U° are the intercepts of the native and unfolded conformations in the absence of denaturant, respectively and S_N and S_U are the slopes of the native and unfolded baselines, respectively. The signal of the intermediate was treated as a single value to reduce numbers of unknown variables during the fitting.

$$\Delta G_1^\circ = -RT * \ln(K_1)$$

$$\Delta G_2^\circ = -RT * \ln(K_2)$$

Data were fit to the three-state equilibrium model using the curve-fitting feature of KaliedaGraph. The following algorithm was used for experiments performed at 37°C (310 K) or 20°C (293 K). The gas constant is in units of kcal*mol⁻¹*K⁻¹.

```
x = m0;
a = m1;
b = m2;
c = m3;
d = m4;
e = m5;
f = m6;
g = m7;
h = m8;
i = m9;
;
```

```

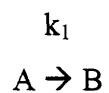
K1() = exp((c*x-d)/(1.987*310));
;
K2() = exp((e*x-f)/(1.987*310));
;
threest(a0, b0, c0, d0, e0, f0, g0, h0, i0) =
((a+b*x)+g*K1()+((h+i*x)*K1()*K2()))/(1+K1()+K1()*K2())
;a=a0; b=b0; c=c0; d=d0; e=e0; f=f0; g=g0; h=h0; i=i0;

```


APPENDIX D: KINETIC DATA FITTING EQUATIONS

1. *Two-state Kinetics*

The following equation was used to describe a kinetic reaction with no intermediates:



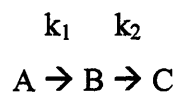
$$Y = Y_B - (Y_B - Y_A) \cdot \exp(-k_1 \cdot t)$$

Data were fit to the two-state kinetic model with the curve-fitting feature of KaliedaGraph using the following algorithm.

```
x=m0;  
a=m1;  
b=m2;  
c=m3;  
;  
Twokin(a0, b0, c0) =  
a*exp(-b*x)+c  
\; a=a0\; b=b0\; c=c0\;
```

2. Three-state Kinetics

The following equation was used to describe a kinetic reaction with one intermediate:



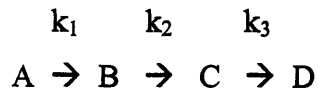
$$Y = Y_C - (Y_B - Y_A) \cdot \exp(-k_1 \cdot t) + (Y_C - Y_B) \cdot \exp(-k_2 \cdot t)$$

Data were fit to the three-state kinetic model with the curve-fitting feature of KaliedaGraph using the following algorithm.

```
x=m0;  
a=m1;  
b=m2;  
c=m3;  
d=m4;  
e=m5;  
;  
Threekin(a0, b0, c0, d0, e0) =  
a*exp(-b*x)+c*exp(-d*x)+e  
\; a=a0\; b=b0\; c=c0\; d=d0\; e=e0\;
```

3. Four-state Kinetics

The following equation was used to describe a kinetic reaction with two intermediates:



$$Y = Y_D - (Y_B - Y_A) \cdot \exp(-k_1 \cdot t) + (Y_C - (Y_C - Y_B) \cdot \exp(-k_2 \cdot t) + (Y_D - Y_C) \cdot \exp(-k_3 \cdot t)$$

Data were fit to the four-state kinetic model with the curve-fitting feature of KaledaGraph using the following algorithm.

```
x=m0;
a=m1;
b=m2;
c=m3;
d=m4;
e=m5;
f=m6;
g=m7;
;
Threekin(a0, b0, c0, d0, e0) =
a*exp(-b*x)+c*exp(-d*x)+e*exp(-f*x)+g
\; a=a0\; b=b0\; c=c0\; d=d0\; e=e0\; f=f0\; g=g0\;
```

You're back already, I didn't even know that you had left.